Comparative Analysis of the Impact of Data Augmentation Techniques on Different Encoders for Diabetic Retinopathy Segmentation

Maria Eduarda Ribeiro Donato da Silva
Centro de Informática
Universidade Federal da Paraíba
João Pessoa, Brasil
maria.donato@academico.ufpb.br

Leonardo Vidal Batista
Centro de Informática
Universidade Federal da Paraíba
João Pessoa, Brasil
leonardo@ci.ufpb.br

Túlio Emanuel Santana de Souza
Centro de Informática
Universidade Federal da Paraíba
João Pessoa, Brasil
tuliosantana@airia.com

***Abstract*—This work presents a systematic analysis of data augmentation techniques applied to diabetic retinopathy segmentation in retinal images, with a focus on hard exudates. Five encoders with the UNet++ architecture were used: MobileNet-v2, VGG11, VGG16, ResNet50, and EfficientNet-B3, evaluated on the IDRiD dataset. The data augmentation techniques analyzed include horizontal flip, shift scale rotate, RGB shift, random brightness contrast, and elastic transform. The metrics used were Intersection over Union (IoU) and Dice Coefficient. The results indicate that complex combinations of techniques do not always lead to performance improvements. The best configuration was the combination of horizontal flip, elastic transform and random brightness contrast, which achieved a Dice of 0.676 and IoU of 0.510 with the VGG11 encoder. However, horizontal flip achieved highly competitive metrics when applied in isolation. The experiments highlight the importance of evaluating each augmentation technique's impact both individually and in combination in medical segmentation tasks.**

***Index Terms*—Data Augmentation, Diabetic Retinopathy, Semantic Segmentation, Retinal Images, UNet++.**

I. Introduction

Diabetic retinopathy (DR) represents the main cause of preventable blindness in working-age individuals worldwide, affecting approximately one-third of people with diabetes mellitus. With projections indicating a 48% increase in cases by 2045 [1], early detection of this condition, characterized by progressive microvascular changes in the retina, becomes crucial to prevent irreversible visual loss. In this context, computer-aided diagnosis systems (CADs) have gained prominence, especially with the advancement of deep learning techniques, which enable automated and increasingly accurate analyses of fundus images [2] [3].

The segmentation of lesions associated with DR, such as microaneurysms, hemorrhages, and exudates, plays a crucial role in the clinical assessment of the disease, as it allows inferring its severity level. Automatic identification of these lesions can significantly contribute to early diagnosis and large-scale screening [3].

Various approaches have been proposed for this task, with emphasis on the usage of convolutional neural networks. In this regard, U-Net [4] and MobileNet [5] architectures were used for segmentation of exudates in retinal images, presenting good results in sensitivity and accuracy [1].

However, the scarcity of pixel-level annotated data remains one of the main obstacles in building effective models for medical image segmentation. The FGADR (Fine-Grained Annotated Diabetic Retinopathy Dataset), which gathers fundus images with detailed lesion annotations, was initially proposed to mitigate this limitation [6]. Complementarily, the IDRiD database [7] provides 516 images meticulously annotated by specialists, becoming an important reference for studies on segmentation and classification of retinal lesions in the Indian population. Both sets are widely used in research in the area.

In this context, data augmentation techniques stand out as a fundamental approach to artificially expand datasets and add variability to training samples. Based on this proposal, the usage of the UNet++ architecture [8] alongside various data augmentation techniques has already been explored in the field [9]. Among the approaches used were the application of geometric and photometric alterations, such as rotations, reflections, and modifications in the brightness and contrast of images.

Although data augmentation techniques are widely employed to artificially expand limited datasets, the literature lacks a systematic analysis on how each technique interacts with different segmentation architectures in retinal images. Moreover, the effectiveness of these techniques strongly depends on the task context and specific characteristics of medical images, with transformations that individually present good results potentially becoming counterproductive when indiscriminately combined [10].

This work proposes a systematic and comparative analysis of data augmentation techniques addressed by previous works, evaluating the impact of each one individually and in different combinations. The objective is to verify whether the joint application of multiple transformations effectively contributes to model performance or if certain combinations, on the contrary, impair segmentation quality. The hypothesis raised is that the cumulative use of data augmentation techniques does not always result in improvements, making it necessary

to rigorously evaluate the specific effect of each transformation in the context of semantic segmentation of diabetic retinopathy lesions.

II. Related Work

Several studies have focused on developing specialized architectures to improve the segmentation and classification of lesions associated with DR. One example is the introduction of HCCNN, a hybrid architecture based on centered CNNs, which performs exudate segmentation and binary classification into grade 0 (healthy) or grade 1 (with diabetic retinopathy) of retinal images [11]. In this work, data augmentation techniques such as rotation and displacement were applied separately to each R, G, and B band, generating 12 augmented images per sample. The model achieved impressive results: accuracy of 98.96%, recall of 98.65%, and specificity of 98.76% in the SYSU dataset [12]. This approach reinforces the positive impact of anatomical segmentation and specific data augmentation on the robustness of DR classifiers.

Another important work in the field was the development of an automated classification system for diabetic eye diseases (DED) based on a combination of traditional preprocessing techniques and deep learning models [13]. The processing flow included contrast enhancement through CLAHE [14], illumination correction, and segmentation of regions of interest, followed by data augmentation with geometric transformations. The application of these techniques significantly improved the performance of models in diabetic retinopathy classification. By adding the aforementioned preprocessing step to the original images, the authors observed gains in both accuracy and recall using the VGG16 architecture [15].

GreenBen [16], a data augmentation technique aimed at the classification of diabetic retinopathy (DR) and diabetic macular edema (DME), was later introduced focusing on preprocessing enhancement. This technique consists of extracting the green channel from the fundoscopic image, followed by a Ben Graham-type enhancement, with the goal of emphasizing relevant structures, such as blood vessels and lesion edges. Using the IDRiD [7] and DeepDRiD [17] datasets, the authors compared the performance of ResNet50 [18] and Swin Transformer [19] networks in different preprocessing scenarios, such as CLAHE, Ben, green channel, and GreenBen. In the case of the IDRiD dataset, for example, the accuracy of ResNet50 increased from 58.3% (without preprocessing) to 68.9% with the application of GreenBen.

Subsequently, three variations of the U-Net architecture [4] were evaluated for the segmentation of blood vessels in retinal images with the aim of assisting in the early detection of diabetic retinopathy [20]. The models were trained on the DRIVE [21], STARE [22], and HRF [23] datasets with the help of image transformations, such as horizontal and vertical mirroring, elastic transformation, grid distortion, and optical distortion. However, one of the limitations of the study is that data augmentation techniques were used together, without analyzing their individual impacts separately. Thus, it is not possible to identify whether any specific transformation negatively influences model performance.

Another relevant architecture in the field is RTC-Net [24], based on residual connections and an encoder-decoder structure for automatic segmentation of exudates in retinal images. The model was trained on the E-Ophtha dataset [25] and subsequently evaluated on HEI-MED [26] and DiaRetDB1 [27]. To expand the training set, the authors applied data augmentation techniques such as resizing, cropping, zooming-in, translation, and mirroring, which led to the generation of more than 2,000 synthetic images from fewer than 100 original images. However, this study also does not individually assess the impacts of data augmentation techniques.

Finally, a complete preprocessing and data augmentation workflow was proposed to improve diabetic retinopathy (DR) classification with deep neural networks [28]. After eliminating blurred images with the Laplacian operator, the authors applied data augmentations based on affine transformations to balance the EyePACS dataset [29]. The approach was tested on seven pre-trained architectures, separately evaluating performance with and without data augmentation. The EfficientNetV2-M model [30] showed the best performance after data augmentation, largely outperforming the results obtained with the original images. Other models, such as VGG16 [15], also benefited from the strategy. These results reinforce the positive impact of data augmentation techniques on class balance and improving the predictive capacity of models.

III. Method

A. Datasets

The analysis of the DiaRetDB1 [27] and IDRiD [7] datasets revealed critical limitations in the former that compromise its usefulness for segmentation of diabetic retinopathy lesions. The dataset shows significant precision problems in the label masks for segmentation, with excessively broad demarcations that frequently encompass areas without lesions. This imprecision in annotations is particularly problematic for the segmentation of small and detailed structures, such as microaneurysms and hard exudates, resulting in a decrease of approximately 3.5 percentage points (p.p.) in the Intersection over Union (IoU) for these lesions when compared to the IDRiD database [9]. Furthermore, the inconsistency in the dataset annotations tends to induce the model to identify false positives, impairing its ability to generalize to real clinical contexts, where precision in delimiting lesions is crucial for correct diagnosis.

Given these limitations, the present work proposes to exclusively use the IDRiD database. It presents significant advantages for segmentation studies: it contains 81 color fundus images, with 54 for the training set and 27 for the test set; it offers specific masks for microaneurysms, hemorrhages, hard and soft exudates, allowing detailed analyses by lesion category; and it adequately represents the characteristics of the Indian population, one of the most affected by diabetic retinopathy globally. Additionally, by focusing on a

single high-quality database, this work seeks to establish a more consistent training and evaluation protocol, eliminating problematic variables related to methodological differences between databases and providing a more robust analysis of the impact of data augmentation techniques on the segmentation of diabetic retinopathy lesions.

B. Metrics

For evaluation and comparison of results with other works in the field, two widely used metrics in semantic segmentation tasks were selected: the Dice Coefficient and the Intersection over Union (IoU). These metrics are particularly suitable for measuring performance in diabetic retinopathy lesion segmentation, as they accurately evaluate the overlap between predicted segmentations and reference masks. Additionally, the usage of metrics such as accuracy is a limitation of other works in the literature, since datasets for medical image segmentation are highly imbalanced. As observed in Figures 2 and 3, the region of interest (ROI) occupies a very small portion of the image, which makes it possible to achieve high accuracy values with a naive model that claims all pixels are background.

1) Intersection over Union (IoU): IoU, also known as the Jaccard Index, is a fundamental metric in segmentation tasks that quantifies the ratio between the area of overlap and the area of union between the model's prediction and the reference mask. Mathematically, it is expressed by:

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

Where:

- TP (True Positive): pixels correctly classified as lesion
- FP (False Positive): pixels incorrectly classified as lesion
- FN (False Negative): pixels incorrectly classified as background

IoU ranges between 0 and 1, where 1 represents perfect segmentation with complete overlap between prediction and reference, while 0 indicates a total absence of overlap.

In this study, we chose to calculate the IoU considering the dataset as a whole, instead of calculating it individually for each image and then obtaining the average. This approach was chosen to avoid distortions in the metric caused by images without lesions, which would result in undefined IoU values, impairing the representativeness of the evaluation [9].

2) Dice Coefficient: Dice Coefficient [31] is a statistical metric that measures the similarity between two sets. In the context of image segmentation, it is calculated by the formula:

$$\text{Dice} = \frac{2TP}{2TP + FP + FN} \tag{2}$$

It is strongly correlated with IoU but gives greater weight to the intersection, making it more sensitive to small variations in the overlap between prediction and reference. Like IoU, it ranges between 0 and 1, with 1 indicating perfect segmentation.

An important characteristic of the Dice Coefficient is its mathematical relationship with IoU:

$$\text{Dice} = \frac{2 \times \text{IoU}}{1 + \text{IoU}} \tag{3}$$

This relationship shows that Dice always presents higher values than IoU for the same segmentation, except in extreme cases (0 or 1), where they coincide. This makes Dice particularly useful for evaluation in imbalanced datasets, as is the case in diabetic retinopathy lesion segmentation, where lesion areas are significantly smaller than background areas.

C. Preprocessing

Data augmentation techniques are fundamental for artificially expanding the training dataset, especially in medical applications such as diabetic retinopathy segmentation, where the quantity of annotated images is frequently limited [4]. Each applied technique has specific objectives and distinct impacts on the segmentation process:

1) Horizontal Flip: This transformation creates a horizontally mirrored image, maintaining the structural properties of exudates and other lesions. By using this technique, the model is expected to learn that lesions can occur in different lateral positions of the retina, increasing the robustness of segmentation without introducing artificial artifacts. It is particularly valuable for preserving the realistic appearance of the retina.

2) ShiftScaleRotate: This technique combines three geometric transformations:

- Translation: Shifts the image by up to 5%, simulating different camera positions during capture.
- Scale: Modifies the image size by up to 10%, helping the model recognize lesions at different magnifications.
- Rotation: Rotates the image by up to 30 degrees, teaching the model to identify exudates regardless of orientation.

3) RGBShift: This transformation independently adjusts the values of the R, G, and B bands by up to 30%, simulating different lighting conditions and variations between capture equipment. In retinopathy segmentation, this technique is particularly relevant because exudates and other structures have specific chromatic characteristics that can subtly vary between patients and imaging conditions.

4) RandomBrightnessContrast: Randomly modifies the brightness and contrast of the image by up to 30%, simulating variations in lighting conditions during capture and equipment settings. In the context of DR, this technique is important because the contrast between exudates and the retinal background can vary significantly between patients, directly influencing the segmentation of lesions.

5) Elastic Transformation: Widely used in medical images [32], this technique introduces elastic distortions to simulate realistic deformations. The process consists of generating random displacement fields, which are applied to the input image. To ensure the deformations have a more natural appearance, these fields are smoothed using a Gaussian filter. The parameters used (alpha=1.0, sigma=10, alpha_affine=10) were

carefully adjusted to produce realistic deformations without generating artifacts that would compromise segmentation.

All techniques were implemented with a 50% probability, meaning that each transformation had an equal chance of being applied or not during training, ensuring diversity in the augmented samples.

These transformations were previously explored [9], but without evaluating whether any method could harm the performance of others. For example, using translation, scale, and rotation can generate regions without information, negatively impacting model learning. Therefore, a systematic comparison of each data augmentation technique will be conducted, both individually and in combination, to identify the set that maximizes performance metrics.

The combination of Data Augmentation techniques has shown promising results in computer vision tasks, as seen in RandAugment [33] and TrivialAugment [34]. However, these methods focus on generic image classification (e.g., CIFAR, ImageNet) and do not address the challenges of medical image segmentation. They also lack a systematic analysis of individual and combined augmentation effects across different architectures. This work aims to fill this gap by empirically evaluating how specific augmentation combinations impact diabetic retinopathy lesion segmentation using five encoders.

D. Architecture

UNet++ [8] represents a significant evolution of the original U-Net architecture, distinguished by three main characteristics: (1) insertion of convolutional layers in the skip paths to reduce semantic differences between encoder and decoder feature maps; (2) use of dense connections in the skip paths to improve gradient flow; and (3) application of deep supervision, enabling model pruning. These modifications make UNet++ particularly suitable for segmentation tasks in medical images.

In previous work [9], four pre-trained encoders on the ImageNet dataset were compared with the UNet++ architecture: MobileNet-v2 [35], VGG16 [15], EfficientNet-B3 [36], and ResNet50 [18]. Analysis of these results revealed that the VGG16 encoder showed superior performance in most cases, ranking only behind MobileNet-v2 in the segmentation of hard exudates from the IDRiD dataset, with an approximate difference of 6%.

In addition to the architectures tested above, this work also intends to use the VGG11 architecture [15]. The choice of VGG11 is based on the hypothesis that a smaller architecture could offer comparable or superior performance in scenarios with limited data, reducing the risk of overfitting often associated with more complex models.

Both VGG11 and VGG16 architectures share structural characteristics, such as the use of 3×3 convolutional filters, five 2×2 max-pooling layers, and three fully connected layers ending with a softmax layer. The fundamental difference between them is in depth: while VGG16 has 16 convolutional layers, VGG11 has only 11, resulting in a reduction in the number of parameters (from 138 million to 133 million) and, consequently, in computational complexity.

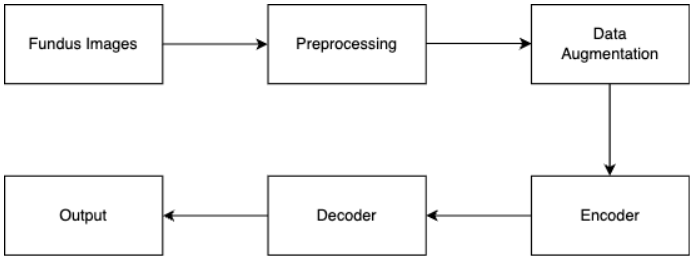

Fig. 1: Flowchart of the proposed methodology

To systematically compare the performance of these five encoders (the four original ones plus VGG11) in conjunction with different data augmentation techniques, the methodological flowchart illustrated in Figure 1 was adopted. This approach allows us to evaluate not only the relative effectiveness of each encoder but also its interaction with different data augmentation strategies, determining optimal configurations for the segmentation of DR lesions.

IV. Experiments

This section presents the results of a comparative analysis of different data augmentation techniques and encoders in the segmentation of hard exudates in images from the IDRiD database. The choice of this specific lesion is justified by the fact that exudates are one of the main signs of diabetic retinopathy that can be prevented through early assessment and detection [37]. All experiments were conducted with the following parameters: 100 epochs, using early stopping with a patience of 10 epochs, batch size of 6, and a learning rate of 0.001. The evaluation was performed using the model checkpoint that achieved the highest IoU on the validation dataset.

A. Impact of Data Augmentation and Comparison Between Encoders

Initially, we compared the performance of the five encoders using all data augmentation techniques simultaneously, as presented in Table I.

TABLE I: Metrics with Data Augmentation

Metrics (%)	IoU	Dice
MobileNetV2	0.496	0.663
EfficientNet-B3	**0.500**	**0.667**
Resnet50	0.470	0.639
VGG11	**0.500**	**0.667**
VGG16	0.488	0.656

EfficientNet-B3 and VGG11 presented the best results, both achieving an IoU of 0.500 and a Dice coefficient of 0.667. Although VGG16 is a deeper and theoretically more robust architecture, it did not show a statistically significant advantage over VGG11. These findings indicate that lighter models can achieve comparable performance when trained on limited datasets.

To evaluate the importance of data augmentation, we conducted experiments without applying any data augmentation technique, as shown in Table II

TABLE II: Metrics without Data Augmentation

Metrics (%)	IoU	Dice
MobileNetV2	0.220	0.361
EfficientNet-B3	**0.362**	**0.532**
Resnet50	0.266	0.421
VGG11	0.312	0.475
VGG16	0.320	0.485

The absence of data augmentation resulted in a significant performance drop for all encoders, with EfficientNet-B3 suffering the smallest reduction. This result confirms the critical importance of data augmentation for segmentation tasks in medical images, especially with limited datasets.

B. Analysis of Individual Data Augmentation Techniques

To evaluate the specific contribution of each technique, tests were conducted separately with the five data augmentation approaches. The results are summarized in Table III.

TABLE III: Metrics for each Data Augmentation technique

Technique	Encoder	IoU	Dice
Horizontal Flip	**MobileNetV2**	**0.495**	0.662
	EfficientNet-B3	0.491	0.659
	Resnet50	0.424	0.596
	VGG11	**0.495**	**0.670**
	VGG16	0.466	0.662
ShiftScaleRotate	MobileNetV2	0.317	0.481
	EfficientNet-B3	**0.429**	**0.600**
	Resnet50	0.255	0.406
	VGG11	0.292	0.452
	VGG16	0.285	0.444
RGBShift	MobileNetV2	0.278	0.436
	EfficientNet-B3	0.349	0.518
	Resnet50	**0.386**	**0.557**
	VGG11	0.344	0.518
	VGG16	0.306	0.469
RandomBrightnessContrast	MobileNetV2	0.320	0.485
	EfficientNet-B3	0.331	0.498
	Resnet50	0.291	0.451
	VGG11	**0.393**	**0.564**
	VGG16	0.379	0.550
ElasticTransform	MobileNetV2	0.452	0.622
	EfficientNet-B3	0.467	0.637
	Resnet50	0.415	0.587
	VGG11	**0.490**	**0.658**
	VGG16	0.463	0.633

Individual analysis revealed that:

1) Horizontal Flip was consistently the most effective technique, with VGG11 achieving the best result.
2) ElasticTransform proved to be the second most effective technique, with performance very close to Horizontal Flip.
3) ShiftScaleRotate presented the worst overall performance, except for EfficientNet-B3, where it outperformed RGBShift and RandomBrightnessContrast.
4) VGG11 demonstrated greater compatibility with most techniques, leading in three of the five approaches tested.

Notably, the performance of the Horizontal Flip technique in isolation was very close to that obtained with all techniques combined for all architectures, suggesting that simpler techniques can be equally effective without the additional computational cost.

C. Analysis of Combinations of Data Augmentations

Based on individual results, we systematically evaluated specific combinations of techniques, always including Horizontal Flip due to its superior performance. As the ShiftScaleRotate transformation presented the worst results in isolation, it was excluded from the combinatorial analysis with the other techniques. Table IV presents the results.

TABLE IV: Metrics for each Data Augmentation combination. HF represents Horizontal Flip transformation, RGB represents RGBShift transformation, RBC represents RandomBrightness-Contrast transformation, and ET represents Elastic Transformation.

Technique	Encoder	IoU	Dice
HF and RGB	**MobileNetV2**	**0.416**	**0.588**
	EfficientNet-B3	0.384	0.555
	Resnet50	0.353	0.522
	VGG11	0.358	0.512
	VGG16	0.344	0.510
HF and RBC	MobileNetV2	0.233	0.378
	EfficientNet-B3	**0.433**	**0.605**
	Resnet50	0.310	0.473
	VGG11	0.325	0.491
	VGG16	0.360	0.529
HF and ET	MobileNetV2	0.492	0.651
	EfficientNet-B3	0.491	0.658
	Resnet50	0.481	0.649
	VGG11	**0.496**	**0.663**
	VGG16	0.487	0.655
HF, RGB and RBC	MobileNetV2	0.286	0.445
	EfficientNet-B3	**0.443**	**0.614**
	Resnet50	0.339	0.506
	VGG11	0.279	0.437
	VGG16	0.331	0.498
HF, ET and RGB	MobileNetV2	0.331	0.498
	EfficientNet-B3	**0.484**	**0.652**
	Resnet50	0.270	0.426
	VGG11	0.311	0.474
	VGG16	0.313	0.477
HF, ET and RBC	MobileNetV2	0.477	0.646
	EfficientNet-B3	0.487	0.655
	Resnet50	0.478	0.647
	VGG11	0.504	0.670
	VGG16	**0.510**	**0.676**
HF, ET, RGB and RBC	MobileNetV2	0.347	0.515
	EfficientNet-B3	**0.378**	**0.549**
	Resnet50	0.342	0.510
	VGG11	0.311	0.474
	VGG16	0.373	0.544

The combination of Horizontal Flip + Elastic Transformation + RandomBrightnessContrast with the VGG16 encoder emerged as the most effective configuration, achieving an IoU of 0.510 and Dice of 0.676. This result is particularly important for three reasons: (1) it outperformed by 1.6 percentage points the best result obtained with the same architecture using all techniques combined; (2) it used only three of the five available techniques, reducing computational cost; and (3) it demonstrated that the careful selection of complementary transformations produces superior results to more comprehensive approaches.

When analyzing the behavior of different encoders with this specific combination, a consistent improvement was observed in all of them. VGG11 obtained an IoU of 0.504 and Dice of

0.670, followed by EfficientNet-B3 (IoU: 0.487, Dice: 0.655) and MobileNet-v2 (IoU: 0.477, Dice: 0.646). The fact that the performance variation between encoders is relatively small suggests that this combination of techniques provides robust benefits, regardless of the architecture used.

The in-depth analysis of the various combinations also revealed significant patterns of interaction between each of the techniques:

1) The inclusion of RGBShift in any combination resulted in worse performance. In the configuration with all techniques (HF+ET+RBC+RGB), EfficientNet-B3 suffered a drop of 10.9 percentage points in IoU (from 0.487 to 0.378) compared to the combination without RGB. This effect was even more pronounced in VGG11, where the IoU plummeted from 0.504 to 0.311 (a reduction of 19.3 percentage points), suggesting that the chromatic modifications introduced by RGBShift may interfere with the model's ability to correctly identify the distinctive features of hard exudates.
2) The combination of HF with ET proved effective, outperforming the individual performance of each technique in almost all encoders. For VGG11, this combination achieved an IoU of 0.496 and Dice of 0.663, values that exceed the performance of HF alone (IoU: 0.495, Dice: 0.670) and ET alone (IoU: 0.490, Dice: 0.658). This synergy can be explained by the complementary nature of the transformations: while HF introduces variations in horizontal orientation, ET simulates structural deformations that can naturally occur in the lesions.
3) A pattern of saturation of benefits was observed as more techniques were added, with the exception of the specific HF+ET+RBC combination. For example, the HF+RGB+RBC combination resulted in significantly lower metrics (VGG11: IoU 0.279, Dice 0.437) when compared to the simple application of HF (VGG11: IoU 0.495, Dice 0.670). This suggests that not all data augmentation techniques work additively, and may even interfere with each other when combined indiscriminately.
4) Each encoder demonstrated distinct sensitivities to augmentation combinations. ResNet50, for example, obtained modest gains with the HF+ET+RBC combination (IoU: 0.478, Dice: 0.647) compared to HF+ET (IoU: 0.481, Dice: 0.649), suggesting that RandomBrightnessContrast did not provide additional benefits for this specific architecture. VGG16, on the other hand, presented opposite behavior, with significant gains when adding RBC. These differences highlight the importance of considering the intrinsic characteristics of each architecture when selecting data augmentation techniques.

These results provide concrete evidence that more complex combinations do not always result in better performance, confirming our initial hypothesis. The key to optimizing the segmentation of hard exudates seems to lie in the careful selection of complementary data augmentation techniques, rather than the indiscriminate application of multiple transformations.

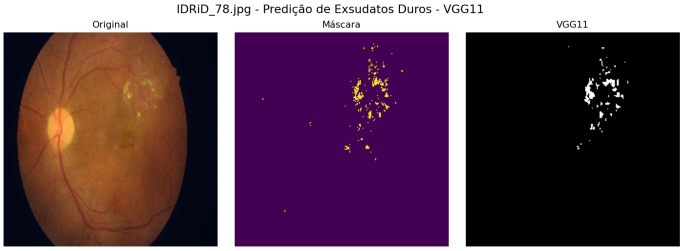

Fig. 2: Combination of Horizontal Flip, ElasticTransform and RandomBrightnessContrast

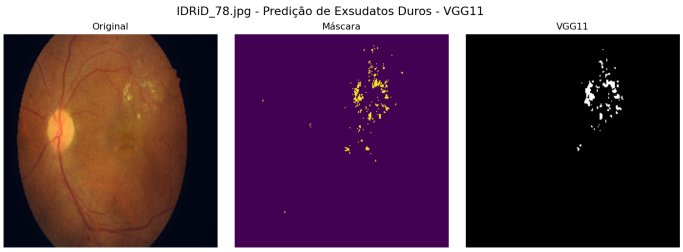

Fig. 3: Combination of Horizontal Flip, ShiftScaleRotate, RandomBrightnessContrast and RGBShift

D. Visualization

Figures 2 and 3 present a visual comparison of the predictions obtained using different data augmentation combinations. The HF+ET+RBC combination (Figure 2) demonstrated greater capability in detecting sparse regions of hard exudates compared to the original combination of all techniques (Figure 3), corroborating the view that selective and specific combinations of data augmentation techniques are more effective than approaches that indiscriminately apply multiple transformations. Moreover, the HF+ET+RBC configuration produced more detailed and coherent segmentation masks, which may be particularly beneficial in clinical scenarios. By generating more precise lesion outlines, this approach has the potential to support medical professionals in the diagnosis of diabetic retinopathy based on model-assisted image analysis.

V. Conclusion

This work conducted a systematic analysis of data augmentation techniques applied to the segmentation of hard exudates in retinal images, using the IDRiD database and different convolutional neural network encoders. The results showed that applying various combined techniques does not always improve metrics and, in some cases, can even hinder model performance.

Among all tested techniques, horizontal flip alone stood out for being simple and efficient, achieving good results with less computational cost. When combined with Elastic Transform and RandomBrightnessContrast, the results were even better, surpassing even more complex combinations.

Based on these results, it is concluded that simpler and more targeted approaches may be more effective than complex combinations of data augmentation techniques. Additionally, the importance of carefully testing each technique becomes evident, taking into account the type of image and the problem being solved.

For future work, we propose the inclusion of statistical hypothesis testing to verify whether the performance differences observed between encoders are statistically significant, thus reinforcing the comparative analysis conducted in this study.

REFERENCES

[1] A. M. Bernardo and E. L. L. Rodrigues, "Segmentação de exsudatos em imagens de retina utilizando redes neurais convolucionais," in *Anais do XV Simpósio Brasileiro de Automação Inteligente*, pp. 8, 2021. doi: 10.20906/sbai.v1i1.2776.

[2] S. Basu, S. Mukherjee, A. Bhattacharya, and A. Sen, "Segmentation of Blood Vessels, Optic Disc Localization, Detection of Exudates, and Diabetic Retinopathy Diagnosis from Digital Fundus Images," in *Proceedings of Research and Applications in Artificial Intelligence*, pp. 173–184, 2021. doi: 10.1007/978-981-16-1543-6_16.

[3] N. Shaukat, J. Amin, M. Sharif, S. Kadry, and L. Ševčík, "Classification and Segmentation of Diabetic Retinopathy: A Systemic Review," *Applied Sciences*, vol. 13, p. 3108, Feb. 2023. doi: 10.3390/app13053108.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., Cham: Springer, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[5] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv preprint arXiv:1704.04861*, Apr. 2017.

[6] Y. Zhou, B. Wang, L. Huang, S. Cui, and L. Shao, "A Benchmark for Studying Diabetic Retinopathy: Segmentation, Grading, and Transferability," *IEEE Trans. Med. Imaging*, vol. 40, no. 3, pp. 818–828, Mar. 2021. doi: 10.1109/TMI.2020.3037771.

[7] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, and F. Meriaudeau, "Indian Diabetic Retinopathy Image Dataset (IDRiD)," *IEEE Dataport*, 2018. doi: 10.21227/H25W98.

[8] Z. Zhou, M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, D. Stoyanov, Z. Taylor, G. Carneiro, T. Syeda-Mahmood, A. Martel, L. Maier-Hein, J. M. R. S. Tavares, A. Bradley, J. P. Papa, V. Belagiannis, J. C. Nascimento, Z. Lu, S. Conjeti, M. Moradi, H. Greenspan, and A. Madabhushi, Eds. Cham: Springer, 2018, pp. 3–11. doi: 10.1007/978-3-030-00889-5_1.

[9] A. E. P. Alves, P. C. Cortez, D. F. Assis, B. R. Silva, and A. G. Moreira, "Detecção de Lesões de Retinopatia Diabética em Imagens de Fundoscopia Utilizando UNet++," in *Anais do XVI Congresso Brasileiro de Inteligência Computacional (CBIC'2023)*, Salvador, BA, pp. 1–9, 2023. doi: 10.21528/CBIC2023-165.

[10] E. Goceri, "Medical Image Data Augmentation: Techniques, Comparisons and Interpretations," *Artificial Intelligence Review*, vol. 56, no. 11, pp. 12561–12605, Nov. 2023. doi: 10.1007/s10462-023-10453-z.

[11] G. Latha, P. Priya, and V. Smitha, "Enhanced Diabetic Retinopathy Detection and Exudates Segmentation Using Deep Learning: A Promising Approach for Early Disease Diagnosis," *Multimedia Tools and Applications*, vol. 83, pp. 1–24, Feb. 2024. doi: 10.1007/s11042-024-18629-7.

[12] L. Lin, M. Li, Y. Huang, P. Cheng, H. Xia, K. Wang, J. Yuan, and X. Tang, "The SUSTech-SYSU Dataset for Automated Exudate Detection and Diabetic Retinopathy Grading," *Scientific Data*, vol. 7, no. 1, p. 409, Nov. 2020. doi: 10.1038/s41597-020-00755-0.

[13] A. Albelaihi and D. M. Ibrahim, "DeepDiabetic: An Identification System of Diabetic Eye Diseases Using Deep Neural Networks," *IEEE Access*, vol. 12, pp. 10769–10789, 2024. doi: 10.1109/ACCESS.2024.3354854.

[14] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, and K. E. Muller, "Contrast-Limited Adaptive Histogram Equalization: Speed and Effectiveness," in *Proc. 1st Conf. Visualization in Biomedical Computing*, pp. 337–345, May 1990. doi: 10.1109/VBC.1990.109340.

[15] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *3rd Int. Conf. Learning Representations (ICLR)*, 2015.

[16] Y. Liu, J. Gao, and H. Zhu, "Diabetic Retinopathy Image Classification Method Based on GreenBen Data Augmentation," *arXiv preprint* arXiv:2410.09444, 2024.

[17] R. Liu, X. Wang, Q. Wu, L. Dai, X. Fang, T. Yan, J. Son, S. Tang, J. Li, Z. Gao, A. Galdran, J. M. Poorneshwaran, H. Liu, J. Wang, Y. Chen, P. Porwal, G. S. W. Tan, X. Yang, C. Dai, H. Song, M. Chen, H. Li, W. Jia, D. Shen, B. Sheng, and P. Zhang, "DeepDRiD: Diabetic Retinopathy—Grading and Image Quality Estimation Challenge," *Patterns*, vol. 3, no. 6, p. 100512, Jun. 2022. doi: 10.1016/j.patter.2022.100512.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[19] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Montreal, QC, Canada, Oct. 2021, pp. 9992–10002. doi: 10.1109/ICCV48922.2021.00986.

[20] S. V. Deshmukh, A. Roy, and P. Agrawal, "Retinal Image Segmentation for Diabetic Retinopathy Detection Using U-Net Architecture," *International Journal of Image, Graphics and Signal Processing*, vol. 15, no. 1, pp. 79–88, Feb. 2023.

[21] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imaging*, vol. 23, no. 4, pp. 501–509, Apr. 2004. doi: 10.1109/TMI.2004.825627.

[22] A. D. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans. Med. Imaging*, vol. 19, no. 3, pp. 203–210, Mar. 2000. doi: 10.1109/42.845178.

[23] A. Budai, R. Bock, A. Maier, J. Hornegger, and G. Michelson, "Robust Vessel Segmentation in Fundus Images," *Int. J. Biomed. Imaging*, vol. 2013, no. 1, p. 154860, 2013. doi: 10.1155/2013/154860.

[24] M. Manan, T. Khan, A. Saadat, M. Arsalan, and S. Naqvi, "A Residual Encoder-Decoder Network for Segmentation of Retinal Image-Based Exudates in Diabetic Retinopathy Screening," *arXiv preprint* arXiv:2201.05963, 2022. doi: 10.48550/arXiv.2201.05963.

[25] E. Decencière, G. Cazuguel, X. Zhang, G. Thibault, J.-C. Klein, F. Meyer, B. Marcotegui, G. Quellec, M. Lamard, R. Danno, D. Elie, P. Massin, Z. Viktor, A. Erginay, B. Laÿ, and A. Chabouis, "TeleOphta: Machine learning and image processing methods for teleophthalmology," *IRBM*, vol. 34, no. 2, pp. 196–203, Apr. 2013. doi: 10.1016/j.irbm.2013.01.010.

[26] L. Giancardo, F. Meriaudeau, T. P. Karnowski, Y. Li, S. Garg, K. W. Tobin, and E. Chaum, "Exudate-based diabetic macular edema detection in fundus images using publicly available datasets," *Med. Image Anal.*, vol. 16, no. 1, pp. 216–226, Jan. 2012. doi: 10.1016/j.media.2011.07.004.

[27] T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen, L. Lensu, I. Sorri, A. Raninen, R. Voutilainen, H. Uusitalo, H. Kalviainen, and J. Pietila, "The DIARETDB1 diabetic retinopathy database and evaluation protocol," in *Proc. British Machine Vision Conf.*, Warwick, UK, 2007, pp. 15.1–15.10. doi: 10.5244/C.21.15.

[28] R. İncir and F. Bozkurt, "A Study on Effective Data Preprocessing and Augmentation Method in Diabetic Retinopathy Classification Using Pre-Trained Deep Learning Approaches," *Multimedia Tools and Applications*, vol. 83, no. 4, pp. 12185–12208, Jan. 2024. doi: 10.1007/s11042-023-15754-7.

[29] E. Dugas, J. [Last Name], J. [Last Name], and W. Cukierski, "Diabetic Retinopathy Detection," 2015. Kaggle competition.

[30] M. Tan and Q. V. Le, "EfficientNetV2: Smaller Models and Faster Training," in *Int. Conf. Mach. Learn.*, Apr. 2021.

[31] L. R. Dice, "Measures of the Amount of Ecologic Association Between Species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945. doi: 10.2307/1932409.

[32] A. V. Buslaev, A. Parinov, E. Khvedchenya, V. I. Iglovikov, and A. A. Kalinin, "Albumentations: Fast and Flexible Image Augmentations," *arXiv preprint* arXiv:1809.06839, 2018.

[33] E. D. Cubuk, B. Zoph, J. Shlens, and Q. V. Le, "RandAugment: Practical automated data augmentation with a reduced search space," 2019. ArXiv preprint.
[34] S. G. Müller and F. Hutter, "TrivialAugment: Tuning-free Yet State-of-the-Art Data Augmentation," 2021. ArXiv preprint.
[35] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 4510–4520. doi: 10.1109/CVPR.2018.00474.
[36] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *ArXiv*, May 2019.
[37] A. M. Bernardo, "Auxílio ao diagnóstico da retinopatia diabética por segmentação semântica utilizando rede neural U-NET em imagens de retina," M.S. thesis, Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2022. doi: 10.11606/D.18.2022.tde-17012024-153112.