Enhanced Semantic Segmentation of Retinal Microlesions through R2U-Net Architecture

Alejandro Pereira[1], Carlos Santos[2], Marilton Aguiar[1], Daniel Welfer[3], Marcelo Dias[1], Rafaela de Menezes[2], Douglas Santana[4]

[1]Programa de Pós-Graduação em Computação – Universidade Federal de Pelotas (UFPel)
Pelotas – RS – Brasil

[2]Instituto Federal de Educação, Ciência e Tecnologia Farroupilha (IFFar)
Alegrete – RS – Brasil

[3]Departamento de Computação Aplicada – Universidade Federal de Santa Maria (UFSM)
Santa Maria – RS – Brasil

[4]Instituto Federal de Educação, Ciência e Tecnologia de Goiás (IFG)
Goiânia – GO – Brasil

{adspereira, marilton, marcelo.sdias}@inf.ufpel.edu.br

carlos.santos@iffarroupilha.edu.br, rafaela.2022302825@aluno.iffar.edu.br

daniel.welfer@ufsm.br

douglas.santana@ifg.edu.br

Abstract. *Diabetic Retinopathy (DR) is a microvascular complication related to diabetes that affects approximately 33% of individuals with this condition and, if not detected and treated early, can lead to irreversible vision loss. Fundus lesions such as Hard and Soft Exudates, Hemorrhages, and Microaneurysms typically identify DR. The development of computational methods to segment these lesions plays a fundamental role in the early diagnosis of the disease. This article proposes a new approach that uses an R2U-Net combined with data augmentation techniques for segmenting fundus lesions. We trained, adjusted, and evaluated the proposed approach in the DDR dataset, achieving an accuracy of 99.87% and an $mIoU$ equal to 59.69%. Furthermore, we assessed it in the IDRiD dataset, achieving an $mIoU$ of 49.92%. The results obtained in the experiments highlight the potential contribution of the approach in generating lesion annotations in creating new DR datasets, which is essential given the scarcity of annotations in publicly available datasets.*

1. Introduction

Diabetic Retinopathy (DR) represents the main ocular complication associated with Diabetes Mellitus, occurring in approximately 33% of diabetic patients. On a global scale, DR affects more than 100 million people, making it one of the leading causes of blindness and visual impairment if healthcare professionals do not identify and treat it early [Tan and Wong 2023]. The severity of DR manifests through the presence of various types of retinal lesions, including Hard Exudates (EX), Hemorrhages (HE), Soft Exudates (SE), and Microaneurysms (MA) in the fundus of the eye [Kanimozhi et al. 2021].

Recently, researchers have emphasized the remarkable feature learning capability of deep learning in image segmentation. This way, encoder-decoder architectures based on U-Net can efficiently segment the previously mentioned lesions, demonstrating promising results in medical image segmentation applications [Wang et al. 2023].

However, obtaining DR data for model training is challenging since the different public datasets available in the literature present a small number of labeled lesions and problems associated with the quality of the images. These problems often negatively affect the results obtained by predictors based on deep neural networks.

In this context, the main contribution of this article is to present a new method for segmenting fundus lesions with greater precision using an R2U-Net [Alom et al. 2018] together with data augmentation techniques. Therefore, the proposed work aims to assist in diagnosing Diabetic Retinopathy and optimize the lesion annotation process to enable the creation of new DR datasets.

We organize the remainder of the article as follows: Initially, we present a review of related work, followed by the base architecture used in this work. We provide details on the materials and methods used later in the article, followed by presenting results obtained by the proposed work. Finally, we will describe the final considerations.

2. Related Work

Given the challenges presented in the previous Section, the following will review works available in the literature whose purpose is the segmentation of fundus lesions. Among the pieces found is the one proposed by Porwal et al. [Porwal et al. 2020], which presents results of deep learning models used for classification, detection, and segmentation of objects in fundus images, wherein the segmentation challenges the U-Net architecture [Ronneberger et al. 2015] has been explored. The main contribution of the work was the availability of the IDRiD public image set (*Indian Diabetic Retinopathy Image Dataset*).

The study by Li et al. [Li et al. 2019] reported the introduction of a new dataset for Diabetic Retinopathy named DDR (Dataset for Diabetic Retinopathy). This study also evaluated deep learning models for retinal lesions' classification, detection, and segmentation. The segmentation task employed the DeepLab-v3+ [Chen et al. 2018] and HED [Xie and Tu 2015] models. The results presented demonstrate the challenge faced by the models in identifying retinal lesions, mainly in Microaneurysms, where they were obtained in the test set IoU values equal to 0.0325 and 0.0110, respectively, in the DeepLab-v3+ and HED models, highlighting the complexity of the task.

In Anand et al. [Anand and Sundaram 2023], the base U-Net architecture received some modifications to improve the accuracy in the segmentation of blood vessels, Hard Exudates, and Microaneurysms in fundus images. The study utilized the following public datasets: IDRiD, DIARETDB1, STARE, ChaseDB1, DRIVE, and HRF. Although the work did not segment Hemorrhages and Soft Exudates, the results presented suggest great potential in the segmentation of retinal lesions with the U-Net architecture, highlighting the accuracies obtained in the IDRiD set of 99.86% in the segmentation of Exudates Hard and 99.9% in targeting Microaneurysms.

Nair et al. [Nair et al. 2023] proposed a new neural network called MesU-Net for retinal image segmentation. This model combines U-Net and MesNet. The authors segmented Hard Exudates, Hemorrhages, Microaneurysms, and blood vessels, obtaining accuracies of 97.6%, 98.1%, 99.2%, and 83.7%, respectively. Furthermore, they used public datasets to train the models, with the IDRiD set dedicated to segmentation fundus lesions and the DRIVE dataset focused on vessel segmentation.

3. R2U-Net Architecture

R2U-Net was developed based on the Deep Residual Network [He et al. 2016], RCNN (Recurrent Convolutional Neural Network) [Liang and Hu 2015] and U-Net [Ronneberger et al. 2015] models in an attempt to optimize the advantages of these three deep learning models. Among the main factors that differ the architecture from R2U-Net to U-Net, the following stand out: (1) Use of RCLs (*Recurrent Convolutional Layers*) and RCLs with residual units; (2) Efficient method of accumulating characteristics in RCL units; and, (3) Elimination of the cutting and copying unit present in U-Net.

Adopting RCLs and RCLs with residual units replacing conventional convolutional layers gives the architecture a more efficient and deeper structure. Furthermore, the efficient feature accumulation method in RCL units incorporated by R2U-Net improves convergence during the training and testing phases. Such an implementation significantly contributes to enhanced feature extraction and representation. Finally, the cut and copy unit in U-Net is replaced exclusively by concatenation operations. These architectural modifications aim to optimize the effectiveness of R2U-Net, exploring state-of-the-art strategies to improve the model's performance, especially in medical image segmentation.

Figure 1 illustrates the architectural model of the neural network that makes up the proposed approach. This structure comprises two main parts: (1) Coding, represented by modules in green, and (2) Decoding, highlighted by modules in blue. Additionally, the architecture incorporates recurring and residual operations. Recurrent operations enable the reuse of information in different areas of the network, promoting a more comprehensive understanding of contexts. On the other hand, residual processes introduce shortcuts between layers to reduce challenges associated with performance degradation in deeper networks. Experimental results presented by the authors indicate that such operations positively impact the training and testing phases of the model without implying an increase in the number of network parameters.

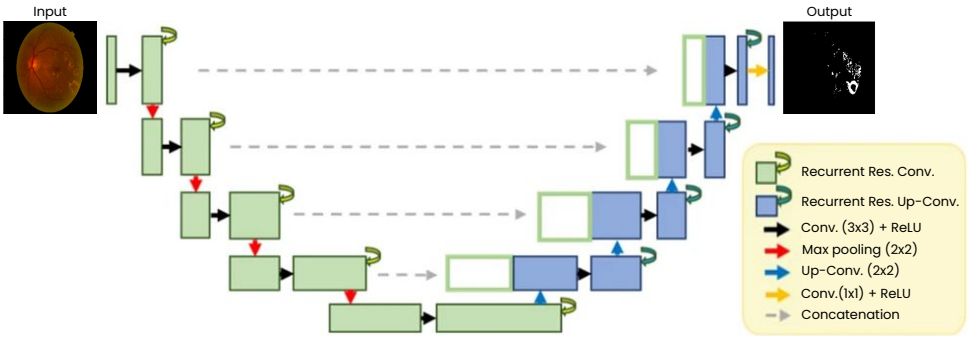

Figure 1. R2U-Net architecture diagram. Source: Adapted from [Shim et al. 2022].

4. Materials and Methods

The proposed work used an R2U-Net as a basis to segment retinal lesions associated with DR. We conducted both experiments on hardware consisting of an AMD Ryzen 5 2600X Six-Core @ 12x 3.6GHz processor, an NVIDIA TITAN Xp GPU video card with 12 GB of VRAM, and 16 GB of RAM. To further enhance the accuracy of the proposed work, we incorporated a data augmentation step.

4.1. Datasets

Two datasets of Diabetic Retinopathy were used in this work, both publicly available and consisting of color images of the fundus of the eye obtained from eye exams. The DDR dataset [Li et al. 2019] has 13,673 fundus images, but 757 have lesion annotations at the pixel level. They annotated the following lesions: Hard Exudates, Soft Exudates, Hemorrhages, and Microaneurysms. The IDRiD dataset was created from clinical exams in an Indian ophthalmology clinic. This set has 516 images of the fundus of the eye. Of these images, 81 have annotations of the lesions at the pixel level.

The DDR set was selected to train and evaluate the proposed work in the experiments, as it has a more significant number of annotated images. We exclusively utilized the IDRiD dataset to evaluate the generalization capacity of the model across various Diabetic Retinopathy datasets. The Hold-out validation method [Lee et al. 2018] was employed on both datasets, allocating 50% for training, 20% for validation, and 30% for testing. Table 1 presents the number of images in each set after divisions.

Table 1. The number of images in each dataset after Hold-out validation.

Dataset	Training	Validation	Testing	Total
DDR	383	149	225	757
IDRiD	42	17	22	81

4.2. Data Augmentation

Among the main challenges faced in constructing computational models for segmenting fundus lesions associated with DR, the small number of lesions labeled in retinal images stands out. This limitation is standard among DR datasets, including the DDR dataset used in training the proposed work. To overcome this problem, we performed a data augmentation step on the images in the training set, aiming to improve the effectiveness of the model in the segmentation of retinal lesions.

At this stage, we applied several techniques to expand the available data in the library, named Albumentations: Fast and Flexible Image Augmentations [Buslaev et al. 2020]. We employed the following methods: Horizontal Flip, Vertical Flip, Elastic Transform, Grid Distortion, and Optical Distortion. We applied the Horizontal and Vertical Flip methods as an alternative to mitigate problems related to positional biases in the input data. The Horizontal Flip method flips the input image around the y axis, in contrast to the Vertical Flip, which reverses the input image along the x axis.

We utilized the Elastic Transform and Grid Distortion techniques to address the structural variations and distortions present in the fundus images. Elastic Transform is an approach that introduces complex wave-like distortions into the image structure, generating non-linear variations in the shape and configuration of objects. On the other hand, Grid Distortion causes variations in the distribution of pixels, creating an illusion of changes in the image's perspective, as if seen from above, with different regions exhibiting varying relative distances.

Additionally, the Optical Distortion technique was employed to simulate visual distortions, mimicking the effects caused by lenses in photography. Exposing the model to these variations makes it possible to improve its generalization capacity with different lesion patterns and structures present in fundus images. In Table 2, we present the parameters used for each technique.

Table 2. Parameter values used when applying data augmentation techniques on fundus images.

Augmentation Type	Parameters	Values
Elastic Transform	alpha, alpha affine, sigma, probability	120, 3.6, 6, 1
Grid Distortion	percentage	1
Horizontal Flip	probability	1
Optical Distortion	distort limit, shift limit, probability	2, 0.5, 1
Vertical Flip	probability	1

Figure 2 depicts the application of data augmentation techniques to a fundus image. These techniques created five additional pictures for each original image in the training set, resulting in 2,298 images. For training the proposed model, we used this augmented dataset.

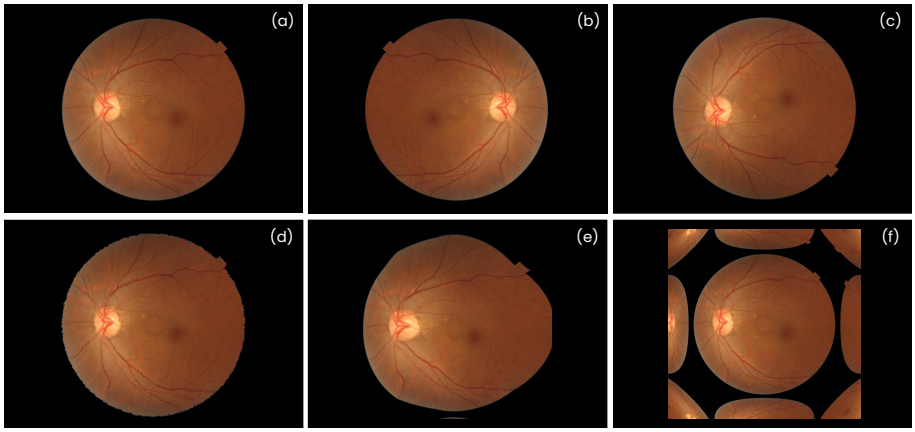

Figure 2. Example of data augmentation over a fundus image from the DDR set. In (a) the original fundus image, (b) Horizontal Flip, (c) Vertical Flip, (d) Elastic Transform, (e) Grid Distortion, and (f) Optical Distortion. Source: Own authorship.

4.3. Architecture Training and Adjustment

For semantic segmentation of fundus lesions, an R2U-Net was used, as mentioned previously, thus training a model for each type of lesion. In the training processes, we divided

images and masks with lesion annotations from the DDR set into Training, Validation, and Test sets in a proportion of 50:20:30, respectively. Furthermore, we configured the network input size to accommodate images with dimensions of 256×256×3 in the RGB (Red, Green, Blue) color space, and we resized both images accordingly.

We trained each model for 50 epochs, using a learning rate 0.001 and a batch size of 4. Furthermore, the optimizer was Adam, combined with the ReLU activation function. We exclusively utilize the validation set to adjust the proposed models, aiming to optimize the hyperparameters and achieve improved results. Table 3 presents the hyperparameters used in the training of this work.

Table 3. Hyperparameters used in the proposed work.

Hyperparameters	Value
Activation Function	ReLU
Batch Normalization	True
Batch Size	4
Filters for each down and upsampling levels	[64, 128, 256, 512, 1024]
Input Size	(256, 256, 3)
Learning Rate	0.001
Number of Epochs	50
Optimizer	Adam
Output Activation	Sigmoid

4.4. Performance Metrics

We evaluated the proposed approach using the metrics: Accuracy (Acc – Equation 1), Sensitivity (Sen – Eq. 2), Precision (Pre – Eq. 3), Dice Coefficient (DC – Eq. 4), and Intersection over Union (IoU – Eq. 5). For a deeper understanding of these metrics, it is crucial to understand the concepts of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN).

TP represents cases where the model correctly classified an instance as positive. TN indicates points where the model was correct when ranking an example as unfavorable. On the other hand, FP occurs when the model mistakenly classifies an instance as positive, while FN represents cases in which the model fails to identify a positive example correctly.

Accuracy is a broad metric that evaluates the overall precision of the model, indicating the proportion of correct predictions about the total number of predictions. This metric is valuable for getting an overview of the model's performance, considering both TP and TN. The Sensitivity metric measures the proportion of positive cases adequately identified by the model, which is especially useful when detecting positives is crucial.

Precision calculates the ratio between the number of examples correctly classified as positive and the total number of samples classified as positive, placing greater emphasis on errors related to false positives (FP). On the other hand, the index is a statistical validation metric used to evaluate the performance and accuracy of spatial overlap between two samples. It is widely used in image analysis and serves as a crucial indicator to

measure the agreement and quality of segmentation, highlighting the similarity between the ground truth (gt) and the model prediction.

Another metric commonly used to evaluate models is IoU. This metric quantifies the level of overlap between the gt bounding boxes and the bounding boxes predicted by the model. The metric ranges from 0 to 1, where 0 indicates no overlap between the boxes, and 1 shows a perfect overlap between the predicted bounding box (pd) and the Ground Truth. The calculation of the evaluation metrics follows the equations below:

$$Acc = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{1}$$

$$Sen = \frac{TP}{(TP + FN)} \tag{2}$$

$$Pre = \frac{TP}{(TP + FP)} \tag{3}$$

$$DC = \frac{2TP}{(2TP + FN + FP)} \tag{4}$$

$$IoU = \frac{Area\ Overlap}{Area\ Union} = \frac{Area\ (gt \cap pd)}{Area\ (gt \cup pd)} \tag{5}$$

5. Results

We compared the proposed work with distinct state-of-the-art models that perform semantic segmentation of objects. In the experiments, we utilized the following models: (1) HED [Li et al. 2019]; (2) DeepLab-v3+ [Li et al. 2019]; (3) U-Net [Ronneberger et al. 2015]; (4) U-Net++ [Zhou et al. 2018]; (5) Attention U-net [Oktay et al. 2018]; and, (6) R2U-Net [Alom et al. 2018]. Furthermore, the U-Net models, U-Net++ and Attention U-net, were also evaluated using pre-trained weights on the ImageNet [Russakovsky et al. 2015] dataset, publicly available on Google Cloud Storage[1].

The Table 4 presents the results obtained in the segmentation of Hard Exudates, Hemorrhages, Soft Exudates, and Microaneurysms using the metrics Acc, Sen, Pre, DC, and IoU in dataset DDR validation set. The values highlighted in bold represent the best results obtained in each metric in the different types of lesions. When analyzing Table 4, it is evident that the proposed work achieved better results than other models in most metrics. Accuracies of 0.9997, 0.9998, 0.9992, and 0.9997 were achieved for EX, HE, SE, and MA, respectively, with a mean Intersection over Union ($mIoU$) for the classes equal to 0.5969.

[1] `https://storage.googleapis.com/tensorflow/keras-applications/vgg16/vgg16_weights_tf_dim_ordering_tf_kernels_notop.h5`

Table 4. Results obtained in the segmentation of Hard Exudates, Hemorrhages, Soft Exudates, and Microaneurysms compared to other models using the metrics Acc, Sen, Pre, DC, and IoU in the validation set of the DDR dataset.

Models	EX					HE				
	Acc	Sen	Pre	DC	IoU	Acc	Sen	Pre	DC	IoU
HED [Li et al. 2019]	-	-	-	-	0.0948	-	-	-	-	0.2183
DeepLab-v3+ [Li et al. 2019]	-	-	-	-	0.2910	-	-	-	-	0.2819
U-Net [Ronneberger et al. 2015]	0.9994	0.1666	0.3064	0.1218	0.5429	0.9984	0.0000	0.0000	0.0000	0.4990
U-Net++ [Zhou et al. 2018]	0.9995	0.0000	0.0000	0.0000	0.4995	0.9984	0.0000	0.0000	0.0000	0.4990
Attention U-net [Oktay et al. 2018]	0.9995	0.1434	0.3737	0.1279	0.5448	0.9984	0.0000	0.0000	0.0000	0.4990
R2U-Net [Alom et al. 2018]	0.9992	0.3631	0.2948	0.2073	0.5720	0.9962	0.0740	0.1270	0.0509	0.5137
U-Net with imageNet weights	0.9994	**0.4639**	0.2295	0.2119	0.5757	0.9205	0.0690	0.4838	0.0020	0.0035
U-Net++ with imageNet weights	0.9995	0.0000	0.0000	0.0000	0.4995	0.9984	0.0000	0.0000	0.0000	0.4990
Attention U-net with imageNet weights	0.9993	0.4197	0.1966	0.1842	0.5644	0.9982	0.0247	0.6314	0.0412	0.5118
Proposed work	**0.9997**	0.3234	**0.4809**	**0.2705**	**0.6053**	**0.9988**	**0.2284**	**0.7585**	**0.2991**	**0.6094**
Models	**SE**					**MA**				
	Acc	Sen	Pre	DC	IoU	Acc	Sen	Pre	DC	IoU
HED [Li et al. 2019]	-	-	-	-	0.0379	-	-	-	-	0.0204
DeepLab-v3+ [Li et al. 2019]	-	-	-	-	0.2756	-	-	-	-	0.0429
U-Net [Ronneberger et al. 2015]	0.9990	0.0000	0.0000	0.0000	0.4992	0.9997	0.0000	0.0000	0.0000	0.4999
U-Net++ [Zhou et al. 2018]	0.9990	0.0000	0.0000	0.0000	0.4992	0.9997	0.0000	0.0000	0.0000	0.4999
Attention U-net [Oktay et al. 2018]	0.9990	0.0149	0.5928	0.0238	0.5066	0.9997	0.0000	0.0000	0.0000	0.4999
R2U-Net [Alom et al. 2018]	0.9991	0.1313	0.7759	0.1682	0.5601	0.9997	0.0408	**0.5528**	0.0629	0.5200
U-Net with imageNet weights	**0.9992**	**0.2656**	0.7786	**0.3054**	**0.6184**	0.9997	0.0319	0.2575	0.0465	0.5158
U-Net++ with imageNet weights	0.9990	0.0000	0.0000	0.0000	0.4992	0.9997	0.0000	0.0000	0.0000	0.4999
Attention U-net with imageNet weights	**0.9992**	0.2357	0.7889	0.2912	0.6091	0.9997	0.0115	0.1909	0.0177	0.5052
Proposed work	**0.9992**	0.2230	**0.8604**	0.2796	0.6072	**0.9997**	**0.2298**	0.3418	**0.2185**	**0.5656**

Table 5. Results obtained in the segmentation of Hard Exudates, Hemorrhages, Soft Exudates, and Microaneurysms compared to other models using the metrics Acc, Sen, Pre, DC, and IoU in the test set of the DDR dataset.

Models	EX					HE				
	Acc	Sen	Pre	DC	IoU	Acc	Sen	Pre	DC	IoU
HED [Li et al. 2019]	-	-	-	-	0.1874	-	-	-	-	0.0524
DeepLab-v3+ [Li et al. 2019]	-	-	-	-	0.3118	-	-	-	-	0.1425
U-Net [Ronneberger et al. 2015]	0.9976	0.1745	0.5455	0.1936	0.5668	0.9951	0.0001	0.5000	0.0001	0.4973
U-Net++ [Zhou et al. 2018]	0.9972	0.0000	0.0000	0.0000	0.4983	0.9951	0.0000	0.0000	0.0000	0.4973
Attention U-net [Oktay et al. 2018]	0.9976	0.1474	0.5935	0.1693	0.5576	0.9951	0.0011	0.0356	0.0007	0.4975
R2U-Net [Alom et al. 2018]	0.9976	0.4077	0.4872	0.3314	0.6143	0.9902	0.0910	0.1586	0.0597	0.5137
U-Net with imageNet weights	0.9978	**0.4868**	0.4611	**0.3791**	0.6312	0.9110	0.0640	0.0052	0.0079	0.4577
U-Net++ with imageNet weights	0.9972	0.0000	0.0000	0.0000	0.4983	0.9951	0.0000	0.0000	0.0000	0.4973
Attention U-net with imageNet weights	0.9977	0.4441	0.4214	0.3400	0.6206	0.9950	0.0200	0.6546	0.0340	0.5083
Proposed work	**0.9980**	0.3566	**0.6960**	0.3777	**0.6466**	**0.9959**	**0.1686**	**0.8138**	**0.2319**	**0.5778**
Models	**SE**					**MA**				
	Acc	Sen	Pre	DC	IoU	Acc	Sen	Pre	DC	IoU
HED [Li et al. 2019]	-	-	-	-	0.0782	-	-	-	-	0.0110
DeepLab-v3+ [Li et al. 2019]	-	-	-	-	0.2295	-	-	-	-	0.0325
U-Net [Ronneberger et al. 2015]	0.9997	0.0001	0.3333	0.0002	0.4994	**0.9998**	0.0000	0.0000	0.0000	0.4998
U-Net++ [Zhou et al. 2018]	0.9997	0.0000	0.0000	0.0000	0.4993	**0.9998**	0.0000	0.0000	0.0000	0.4998
Attention U-net [Oktay et al. 2018]	0.9997	0.0104	0.1224	0.0094	0.5025	**0.9998**	0.0000	0.0000	0.0000	0.4998
R2U-Net [Alom et al. 2018]	0.9997	0.0805	0.2077	0.0619	0.5215	**0.9998**	0.0430	**0.2287**	0.0381	0.5135
U-Net with imageNet weights	0.9997	**0.1994**	0.2619	0.1322	0.5533	**0.9998**	0.0077	0.0485	0.0077	0.5026
U-Net++ with imageNet weights	0.9997	0.0000	0.0000	0.0000	0.4993	**0.9998**	0.0000	0.0000	0.0000	0.4998
Attention U-net with imageNet weights	0.9997	0.1840	0.2126	0.0998	0.5378	**0.9998**	0.0029	0.0531	0.0039	0.5018
Proposed work	**0.9998**	0.1704	**0.3953**	**0.1446**	**0.5584**	**0.9998**	**0.1899**	0.1801	**0.1099**	**0.5336**

Table 5 presents the segmentation fundus lesions EX, HE, SE, and MA results in the DDR test set. The table includes metrics such as Acc, Sen, Pre, DC, and IoU, highlighting the superior performance of the proposed work compared to other methods.

Notably, in the test set, the accuracies were slightly lower compared to the validation set, reaching 0.9980 for EX, 0.9959 for HE, 0.9998 for SE, 0.9998 for MA, and a $mIoU$ of 0.5791. These results indicate the ability of the proposed work to deal with data never observed before, reinforcing its reliability and adaptability.

Figure 3 compares the predictions made by the proposed approach during the segmentation of fundus lesions and the ground truth of these images from the test set of the DDR dataset. When comparing columns (b) and (c), despite not having detected some lesions, it is notable that the proposed work managed to segment most of the instances in the selected images accurately.

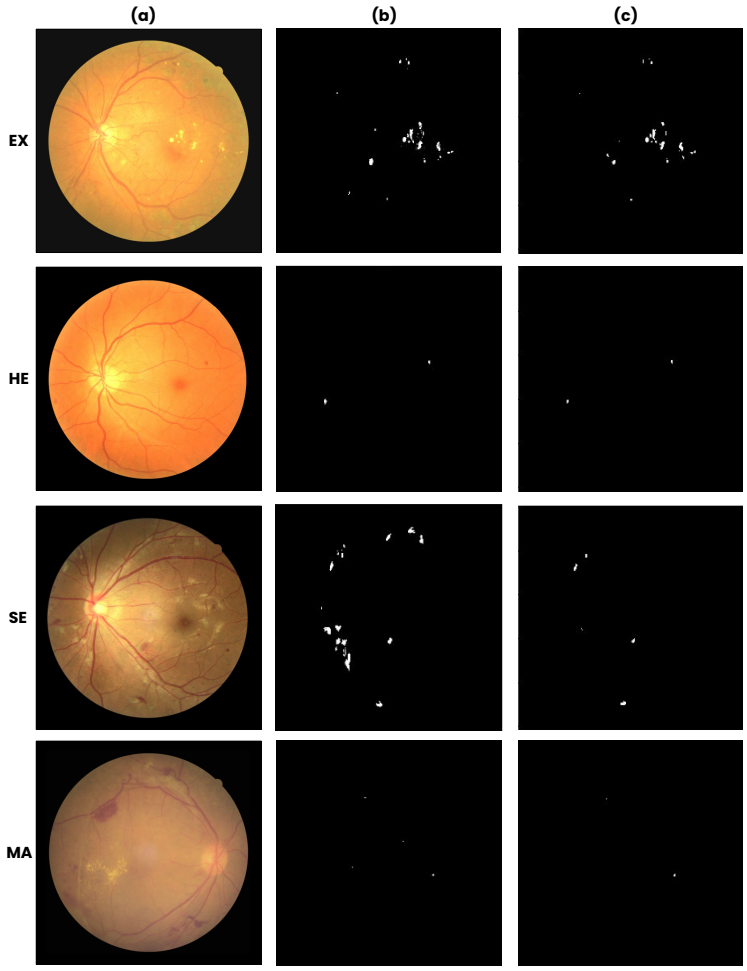

Figure 3. Visual comparison between the fundus lesion instance segmentations performed by the proposed work with ground truth in images from the test set of the DDR dataset. (a) Original images; (b) Ground Truth; and (c) Proposed work. Source: Own authorship.

Finally, experiments were conducted on the IDRiD set to assess the accuracy of the proposed work on diverse datasets. Thus, our model trained on the DDR set was used to segment the lesions contained in the *dataset* IDRiD validation set. Table 6 presents the results obtained with the metrics IoU and $mIoU$. The proposed work achieved IoU values of 0.4983, 0.4992, 0.4993, and 0.4998 for the classes EX, HE, SE, and MA, with a $mIoU$ of 0.4992.

Table 6. Results obtained in the segmentation of fundus lesions compared to the other models using the metrics IoU e $mIoU$ in the IDRiD dataset validation set.

Models	IoU				$mIoU$
	EX	**HE**	**SE**	**MA**	
U-Net	0.4987	**0.4999**	**0.4999**	0.0000	0.3746
U-Net++	0.0000	0.0000	0.0000	0.0000	0.0000
Attention U-net	**0.4989**	**0.4999**	0.4995	0.0000	0.3745
R2U-Net	0.4963	0.4956	0.4994	**0.4999**	0.4978
U-Net with imageNet weights	0.4975	0.4595	0.4994	**0.4999**	0.4890
U-Net++ with imageNet weights	0.0000	0.0000	0.0000	0.0000	0.0000
Attention U-net with imageNet weights	0.4978	0.4997	0.4996	**0.4999**	**0.4992**
Proposed work	0.4983	0.4992	0.4993	0.4998	**0.4992**

Table 7 shows the results obtained in the *dataset* IDRiD test set with the metrics IoU and $mIoU$. In this evaluation stage, the proposed work got a $mIoU$ equal to 0.4990 and IoU equal to 0.4978, 0.4990, 0.4992, and 0.4998 in classes EX, HE, SE, and MA, respectively. In both evaluation stages in the IDRiD set, the proposed work obtained a $mIoU$ higher than the other models, validating the generalization capacity of the proposed work in segmentation fundus lesions associated with DR.

Table 7. Results obtained in the segmentation of fundus lesions compared to the other models using the metrics IoU and $mIoU$ in the IDRiD dataset test set.

Models	IoU				$mIoU$
	EX	**HE**	**SE**	**MA**	
U-Net	0.4986	0.4995	**0.4999**	0.0000	0.3745
U-Net++	0.0000	0.0000	0.0000	0.0000	0.0000
Attention U-net	**0.4988**	**0.4999**	0.4998	0.0000	0.3746
R2U-Net	0.4954	0.4933	0.4995	**0.4999**	0.4970
U-Net with ImageNet weights	0.4966	0.4608	0.4996	**0.4999**	0.4892
U-Net++ with ImageNet weights	0.0000	0.0000	0.0000	0.0000	0.0000
Attention U-net with ImageNet weights	0.4966	0.4994	0.4997	**0.4999**	0.4989
Proposed work	0.4978	0.4990	0.4992	0.4998	**0.4990**

6. Conclusions

This work introduced a convolutional neural network model for the segmentation of instances of fundus lesions associated with Diabetic Retinopathy using an R2U-Net combined with data augmentation techniques, which improved the accuracy of the model, including the use of small batch sizes. The best results obtained by the proposed work were in the dataset DDR validation set, with accuracies equal to 99.97%, 99.98%, 99.92%, and 99.97% for Exudates Hard, Hemorrhages, Soft Exudates, and Microaneurysms, respectively, totaling an average class accuracy of 99.87% and a $mIoU$ of 59.69%. Furthermore,

the proposed work was superior to the other models in the IDRiD dataset, reaching $mIoU$ equal to 49.92% and 49.90%, respectively, in the validation and test sets.

These results demonstrate significant potential in utilizing encoder-decoder architectural models based on U-Net for segmenting retinal lesions related to DR. While our work shows promise, improving model accuracy remains a priority for enhancing early Diabetic Retinopathy diagnosis and optimizing lesion annotation. Our research group aims to develop a new publicly available dataset comprising fundus images, lesion binary masks, and corresponding annotations. This dataset will aid in training deep learning algorithms for segmentation and object detection tasks. Our method automates lesion mask generation, benefiting medical experts by reducing their annotation workload. Experts need only validate and potentially correct these masks, streamlining the process and optimizing medical resource allocation.

In future work to achieve these objectives, it will be necessary to explore new architectural models, incorporate pre-processing steps to improve the extraction of image features within the neural network, and explore ensemble techniques combining different encoder-decoder deep neural network models to create an even more precise approach.

Acknowledgements

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

References

Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T. M., and Asari, V. K. (2018). Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv preprint arXiv:1802.06955*.

Anand, M. and Sundaram, A. M. (2023). Channel and spatial attention aware unet architecture for segmentation of blood vessels, exudates and microaneurysms in diabetic retinopathy. *Preprint*.

Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., and Kalinin, A. A. (2020). Albumentations: Fast and flexible image augmentations. *Information (Switzerland)*, 11(2).

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11211 LNCS:833–851.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2016-Decem, pages 770–778, Las Vegas, NV, USA, 27–30 June 2016.

Kanimozhi, J., Vasuki, P., and Roomi, S. M. M. (2021). Fundus image lesion detection algorithm for diabetic retinopathy screening. *Journal of Ambient Intelligence and Humanized Computing*, 12:7407–7416.

Lee, L. C., Liong, C.-Y., and Jemain, A. A. (2018). Validity of the best practice in splitting data for hold-out validation strategy as performed on the ink strokes in the context of forensic science. *Microchemical Journal*, 139:125–133.

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.

Liang, M. and Hu, X. (2015). Recurrent convolutional neural network for object recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3367–3375.

Nair, A. T., ML, A., and MN, A. K. (2023). Segmentation of retinal images using improved segmentation network, mesu-net. *International Journal of Online & Biomedical Engineering*, 19(15).

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., et al. (2018). Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.

Porwal, P., Pachade, S., Kokare, M., and Deshmukh, G. (2020). IDRiD: Diabetic Retinopathy – Segmentation and Grading Challenge. *Medical Image Analysis*, 59.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9351:234–241.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115:211–252.

Shim, J.-H., Kim, W. S., Kim, K. G., Yee, G. T., Kim, Y. J., and Jeong, T. S. (2022). Evaluation of u-net models in automated cervical spine and cranial bone segmentation using x-ray images for traumatic atlanto-occipital dislocation diagnosis. *Scientific Reports*, 12(1):21438.

Tan, T.-E. and Wong, T. Y. (2023). Diabetic retinopathy: Looking forward to 2030. *Frontiers in Endocrinology*, 13:1077669.

Wang, H., Cao, P., Yang, J., and Zaiane, O. (2023). Mca-unet: multi-scale cross co-attentional u-net for automatic medical image segmentation. *Health Information Science and Systems*, 11(1):10.

Xie, S. and Tu, Z. (2015). Holistically-nested edge detection. *Proceedings of the IEEE International Conference on Computer Vision*, 2015 Inter:1395–1403.

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., and Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, pages 3–11. Springer.