Diabetes detection through retinal images utilizing artificial intelligence

Ammar Asad, Zeinab Khadra, Mohammed Hammoud, Ola Haydar, Sergey S. Lupin

***Abstract*— Diabetes is a prevalent chronic condition impacting millions globally, with complications such as diabetic retinopathy posing substantial risks, including vision loss. This article critically examines the transformative role of artificial intelligence (AI) in the early detection and diagnosis of diabetic retinopathy through retinal image analysis. Leveraging advanced deep learning algorithms, AI systems can analyze high-resolution retinal images to identify subtle pathological changes often overlooked by human observers. By integrating extensive datasets and employing sophisticated neural networks, these technologies enhance diagnostic accuracy and enable timely interventions, thereby mitigating visual impairment among diabetics. Additionally, we address challenges related to data diversity, ethical concerns, and the necessity for clinical validation of AI tools. As we advance into a new era of medical diagnostics, the collaboration between AI and ophthalmology is revolutionizing patient care, making early detection standard in managing diabetes complications. This study utilizes three widely recognized retina datasets—CHASE_DB1, DRIVE—and employs the U-Net model to develop our neural network while discussing the results.**

***Keywords*— *Artificial Intelligence, Dataset, Convolutional, Diabetes, Model, Segmentation, Blood vessels, Medical image, Drive, CHASE-DB1*.**

I. INTRODUCTION

Diabetic retinopathy is a condition that affects the blood vessels in the retina due to high blood sugar levels. Early symptoms include blurred vision and, if left untreated, the condition can lead to complete vision loss. The retina contains several layers of cells that detect light and transmit information to the brain. When the blood vessels in the retina are compromised, fluid can leak or burst, resulting in damage to the light-sensitive cells.

The number of people with diabetes increased from 108 million in 1980 to 422 million by 2014. Between 2000 and 2019, diabetes mortality rates This rise in prevalence has been more pronounced in low- and middle-income countries compared to high-income countries increased by 3% across all age groups[1].

Doctors analyze various clinical signs in retinal images to understand how diabetes affects vision. One of the most noticeable changes is observed in the blood vessels. These vessels often appear dilated or exhibit a zigzag pattern, indicating how diabetes impacts blood flow. The images may also reveal blockages in certain vessels, which can disrupt blood supply to the retina.

Bleeding is another critical sign; myelogenous bleeding can manifest as small red spots, while superficial bleeding presents as thin red lines beneath the retina. Such bleeding is indicative of deterioration in blood vessel health. Additionally, macular edema is a significant sign of diabetic effects, as fluid accumulates in the macular area, potentially leading to a loss of central vision.

Due to a lack of oxygen, new tissue may form, known as abnormal vascular formation, which further complicates the condition. Chromosomal changes may also be visible in the retina, with color changes or pigment spots signaling a worsening state.

The effectiveness and reliability of artificial intelligence models heavily depend on the volume of input data, especially images from a large cohort of patients. This extensive training is essential for achieving accurate outcomes. However, obtaining comprehensive medical data poses significant challenges in the healthcare field. Factors such as human error, stemming from inaccuracies in imaging or equipment, can adversely affect decision-making processes. This is particularly true for images of the eye's retinal blood vessels, where subtle details may be easily overlooked.

In this context, deep learning techniques, particularly those focused on vascular segmentation, play a crucial role. They enhance image clarity and emphasize intricate details that may not be easily recognizable to all medical professionals specializing in ophthalmology. By leveraging these advanced technologies, we can improve diagnostic accuracy and support better clinical decisions.

Retinography is a delicate procedure that necessitates advanced technology and skilled operators. However, several medical errors can occur during the process. One common issue is focus errors, which happen when the camera is not adjusted properly, resulting in blurred images that make it difficult for doctors to interpret the results accurately. Mispositioning the camera may lead to errors in identifying the target area of the retina, resulting in a loss of crucial information.

Additionally, insufficient or excessive lighting can negatively impact image quality. Furthermore, improper handling of the patient—such as failing to stabilize or adjust the patient's head during imaging—can also cause blurry pictures and lead to misinterpretation of the results. Even when images are clear, some doctors may misinterpret them, which can result in misdiagnosis.

[1] https://www.who.int/home

Using inappropriate or damaged lenses can further compromise image quality. To avoid these mistakes, it is essential to implement thorough training and adhere to careful protocols, ensuring accurate and reliable results.

Eye segmentation using deep learning is a rapidly evolving field with significant potential across various domains. Advances in neural network architectures and training methodologies are continuously improving the accuracy and efficiency of eye segmentation systems. Due to the limited availability of data, we utilized three well-known medical datasets that are frequently employed in scientific research: CHASE_DB, STARE, and DRIVE. Our studies focused on the U-Net architecture to achieve effective retinal segmentation.

The main goal of this work is to improve the results concerning the detection of diabetic retinopathy by applying deep learning techniques, focused on segmenting the blood vessels of the retina using a U-Net architecture. By leveraging advanced data augmentation strategies and analyzing established retinal image datasets, this research seeks to improve diagnostic precision and support clinical decision-enabling better clinical decisions in the management of diabetic retinopathy.

The rest of the paper is organized as follows: Section II reviews related work; Section III describes the implementation details, including the dataset and evaluation metrics; and finally, Section IV discusses the results obtained.

II. RELATED WORKS

A deep learning method utilizing convolutional neural networks (CNNs) has been developed for the detection of diabetic retinopathy (DR) with visual interpretation. This approach employs a regression activation map (RAM) to highlight essential areas in retinal images, thereby enhancing the accuracy of early detection. Experiments have shown that this method outperforms traditional techniques, providing better support for doctors in diagnosing the disease [1].

By implementing a U-Net architecture along with data augmentation techniques, significant improvements in segmentation accuracy were achieved, particularly evident in the results from the DRIVE dataset [2].

Additionally, the Patho-GAN model has been designed to detect diabetic retinopathy and generate retinal images. This model incorporates Convolution-BatchNorm-LeakyReLU blocks and employs a Generative Adversarial Network (GAN) strategy to improve interaction between the generator and discriminator. The training process utilizes a combination of adversarial loss, perceptual loss, and intensity loss, all aimed at enhancing the quality of the generated images [3]. Furthermore, the RAVIR dataset was introduced to analyze retinal blood vessels using infrared imaging, thereby improving analysis accuracy.

The SegRAVIR methodology was developed to effectively distinguish between arteries and veins, showing improved segmentation accuracy and vessel diameter measurement compared to existing models. Notable performance enhancements were observed on the DRIVE, STARE, and CHASE_DB1 datasets, highlighting the effectiveness of this proposed approach [4]. To improve retinal vessel segmentation, new techniques were introduced, including encoding-enhancing blocks and bottleneck optimization, which help reduce feature loss. Additionally, an adaptive thresholding algorithm was proposed for precise pixel-level classification. These methods have led to significant advancements in the extraction of fine vessels. Performance evaluations using the DRIVE, CHASE-DB1, and STARE datasets demonstrated competitive results. Cross-validation assessments further indicate a strong generalization capability across different datasets [5].

Various methods for extracting vascular centers have been discussed, with a focus on deformable and geometric models. The challenges associated with vascular segmentation in pathological cases, such as low image quality and noise, were also reviewed, underscoring the need for performance improvements [6]. The importance of validating scientific research and enhancing the reproducibility of studies was emphasized. The difficulties researchers encounter in achieving reproducible results were highlighted. Several initiatives aimed at reducing the rates of irreproducibility were examined, along with the impact these issues have on the scientific community. The discussion stressed the necessity of promoting transparency and adopting best practices in scientific research [7].

Deep convolutional neural networks (CNNs) have been successfully employed to recognize human activities by analyzing multichannel time signals. A highly effective method has been developed to automatically extract features from raw data, and results demonstrate that this approach surpasses traditional algorithms in several experiments [8].

IterNet is an advanced model based on the UNet architecture designed for segmenting blood vessels in retinal images. It incorporates replicated miniature units to enhance the detection of intricate details that might otherwise be overlooked. This model achieves outstanding performance, receiving an AUC score of 0.9816 on the DRIVE dataset. To prevent overfitting, it utilizes techniques such as weight sharing and intermittent connections. This research highlights the importance of effective communication in presenting segmentation results, which in turn assists healthcare professionals in vascular analysis [9].

The DUCK-Net model is an innovative convolutional neural network designed to enhance the accuracy of medical image segmentation. It utilizes a codec-decoding structure combined with a residual size reduction mechanism. Experiments conducted on popular datasets for segmenting appendages demonstrated superior performance in various metrics, including the DIESS coefficient and the Jaccard index, indicating the model's strong ability to generalize even with limited training data [10].

The significance of the Global Average Pooling (GAP) layer in convolutional neural networks (CNNs) was highlighted, showcasing how it improves the model's capacity to accurately identify locations, even when training is based solely on image labels. Additionally, the model incorporates Class Activation Mapping (CAM) technology, which aids in

identifying key areas within images, thus paving the way for new possibilities in image recognition applications.

The results indicated that the system achieved a low error rate of 37.1% in positioning on the ILSVRC 2014 dataset, without requiring bounding box explanations [11]. DFFR-U-NET has introduced a new method to improve the division accuracy of diabetic retinopathy lesions such as HM, HE, and OD. The results showed that the proposed method achieves an accuracy of up to 98% in the HM division and 99% in the HE and OD division, superior to the previous methods that used the IDRiD data set [12].

The SA-UNet network offers an innovative approach to retinal vascular fragmentation by integrating the Spatial Attention Unit, enhancing segmentation accuracy. The network relies on convolution blocks with orderly projection to reduce over-adaptation, and its performance was evaluated on DRIVE and CHASE_DB1 datasets, showing superior results compared to current methods. SAUNet is a lightweight and efficient model, making it an ideal choice for retinal vascular segmentation applications [13]. The approach integrates supervised and unsupervised algorithms to enhance segmentation performance, utilizing multi-scale filters alongside U-Net models [14]. A structured drop block technique has been proposed to address this issue. Additionally, training neural networks with dynamic weights have been employed to tackle challenges like the under-segmentation of faint vessels and inaccuracies around edge pixels [15].

III. METHODS AND METHODOLOGY

A. Dataset

The photographs in the DRIVE database were sourced from a diabetic retinopathy screening program conducted in the Netherlands. The screening involved 400 diabetic participants, aged between 25 and 90 years. From this population, 40 photographs were randomly selected: 33 of these images showed no signs of diabetic retinopathy, while 7 images indicated mild early diabetic retinopathy, all with a resolution of 565×584 pixels. Additionally, the CHASE_DB1 dataset is designed for retinal vessel segmentation and includes 28 color retinal images, each with a size of 999×960 pixels2. These images were collected from the left and right eyes of 14 school children. Each image has been annotated by two independent human experts3. Several samples from the DRIVE dataset are illustrated in Figure 1a, while Figure 1b displays a selection of samples from the CHASE_DB1 dataset.

B. Data Augmentation

Data augmentation is a technique in machine learning and deep learning that enlarges a training dataset by creating modified versions of existing data. This approach is especially beneficial when the original dataset is small or imbalanced, as it enhances the model's ability to generalize and adapt to new data.

Common methods of augmentation include rotating images at various angles, flipping them horizontally or vertically, zooming in or out, cropping random sections, and adjusting attributes such as brightness, contrast, saturation, and color.

Additionally, random noise can be introduced into images, and transformations like panning, cropping, or elastic distortions can be applied.

(a)

(b)

Fig. 1. some samples from each dataset

Data augmentation is a valuable strategy for improving model performance and is widely used across various fields, particularly in computer vision and natural language processing. Numerous influential research papers have demonstrated how data augmentation can enhance the performance of neural network models in classifying medical images, including X-rays and MRI scans. Researchers have utilized techniques such as rotation, scaling, cropping, flipping, and adding noise to expand the dataset size and improve the model's generalization capabilities.

Below is Table 1, which presents the mechanisms employed in the data augmentation process along with their corresponding values.

TABLE. I.LIST OF APPLIED AUGMENTATIONS

Type	Value
Vertical Flip	0.5
Random Rotate90	0.5
Horizontal Flip	1
Transpose	1
Grid Distortion	1

C. Model Architecture

U-Net is a deep learning model specifically designed for medical image segmentation tasks, although it has also become popular in various other fields. First introduced in 2015 by researchers at the University of Freiburg, its primary goal was to enhance the accuracy of segmenting medical images, particularly in areas such as X-ray and MRI analysis.

The U-Net architecture features a distinct design that utilizes an encoder-decoder structure, enabling it to effectively

[2] https://www.kaggle.com/datasets/andrewmvd/drive-digital-retinal-images-for-vessel-extraction

[3]https://www.kaggle.com/datasets/rashasarhanalharthi/chase-db1

capture spatial and contextual information. The encoder portion of U-Net functions similarly to a traditional convolutional neural network (CNN), progressively reducing the spatial dimensions of the input image while increasing the depth of feature representations. This process is accomplished through convolutional layers followed by pooling layers, which help extract essential features from the image.

U-Net is widely employed in various domains, including medicine for segmenting tissues and organs in medical images, computer vision for object detection, and agricultural image analysis for identifying crops or affected areas. The decoder component aims to reconstruct the segmented output by increasing the size of the feature maps. U-Net stands out due to its skip connections, which link corresponding layers in the encoder and decoder. These connections allow the model to retain high-resolution features from earlier layers, a crucial aspect for accurate localization in segmentation tasks. This architecture is particularly effective for biological image analysis, where precise delineation of structures is essential. Moreover, it is designed to learn from a relatively limited number of training images, making it robust in scenarios where data may be scarce.

Overall, U-Net's innovative approach to integrating feature extraction with spatial information makes it a powerful tool for various applications, including remote sensing image analysis and artistic style transfer. Its flexibility and efficiency have made it a popular choice among practitioners in deep learning. Figure 2 below illustrates the architecture of U-Net.

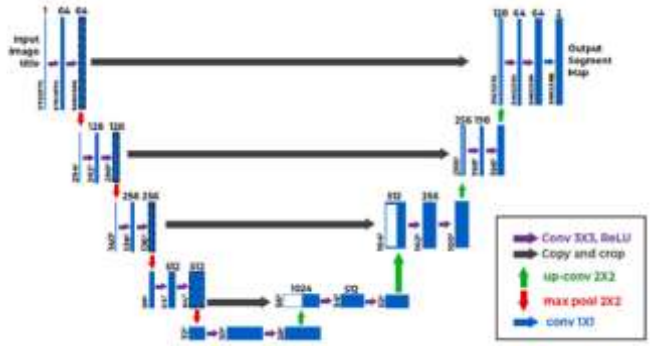

Fig. 2. U-Net Architecture Module

D. Implementation Details

for model training. The models are optimized with the Adam optimizer, using a learning rate of 1e-4, without implementing any learning rate scheduling algorithms. For our experiments on the DRIVE and CHASE_DB1 datasets, we employ a batch size of 8. Our architecture consists of four convolutional layers. The training process uses binary cross-entropy loss, along with a combination of binary cross-entropy loss functions.

To enhance the training dataset, we applied data augmentation techniques, increasing the total number of images to 600 for both datasets. Table 1 presents the various data augmentation methods employed in this study. We implemented all our models in PyTorch and conducted the training on a single NVIDIA RTX 3060 GPU. Each model was trained for a total of 50 epochs, and notably, we did not use pre-trained weights in any of our experiments.

Our network architecture features four layers each for the encoders and decoders. Each encoder layer comprises two convolutional layers followed by a BatchNorm2d layer. BatchNorm2d, which stands for "Batch Normalization for 2D data," is designed to normalize the outputs of 2D convolutional layers that use a kernel size of 3 and padding of 1. We employed the ReLU (Rectified Linear Unit) activation function to introduce non-linearity into the model, ensuring computational efficiency.

Similarly, each decoder layer consists of two convolutional layers and a BatchNorm2d layer, with the same kernel size and padding values. However, we utilized transposed convolution (ConvTranspose2d) in the decoders to achieve effective upsampling and reconstruction.

E. Metrics

F1 Score: The F1 score is a metric that balances precision and recall by calculating their harmonic mean. The optimal value of the F1 score is 1, which indicates perfect precision and recall, while the lowest possible value is 0, reflecting a scenario where either precision or recall (or both) is zero. The F1 score treats precision and recall equally, meaning that an increase in one will proportionately affect the overall score. This emphasizes the importance of a balanced approach in performance evaluation. The formula for the F1 score is given in Equation 1.

$$F1_{score} = \frac{2*TP}{2*TP+FP+FN} = 2* \frac{\text{Precision}}{\text{Precision} + \text{Recall}} \quad (1)$$

In this context, TP represents the count of true positives, FN denotes the count of false negatives, and FP indicates the count of false positives. By default, the F1 score is set to 0.0 when there are no true positives, false negatives, or false positives. This situation signifies that the model has failed to correctly identify any positive instances and highlights its inability to effectively distinguish between the classes being evaluated.

Jaccard Score: The Jaccard index, also known as the Jaccard similarity coefficient, is a statistical measure used to compare the similarity and diversity of sample sets. It is calculated by dividing the size of the intersection of two sets by the size of their union. The formula is presented in Equation 2.

$$J = \frac{A \cap B}{A \cup B} \quad (2)$$

The expression $|A \cap B|$ represents the number of elements that are common to both sets A and B, while $|A \cup B|$ denotes the total number of elements in either set. The Jaccard index ranges from 0 to 1, where a value of 0 indicates no similarity between the sets, and a value of 1 signifies complete similarity.

Precision and recall are crucial metrics for evaluating the effectiveness of predictions, particularly in scenarios where there is a significant class imbalance. In the context of information retrieval, precision measures the proportion of

relevant items within the set of retrieved items. In other words, it assesses how many of the returned items are relevant.

Conversely, **Recall** measures the proportion of relevant items that were successfully retrieved out of all the relevant items available, as expressed in Equation 3. This means it evaluates how many relevant items were identified in the results. Here, "relevant" refers to positively labeled items, including both true positives and false negatives.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

Precision is defined as the number of true positives (TP) divided by the sum of true positives and false positives (FP), as expressed in Equation 4.

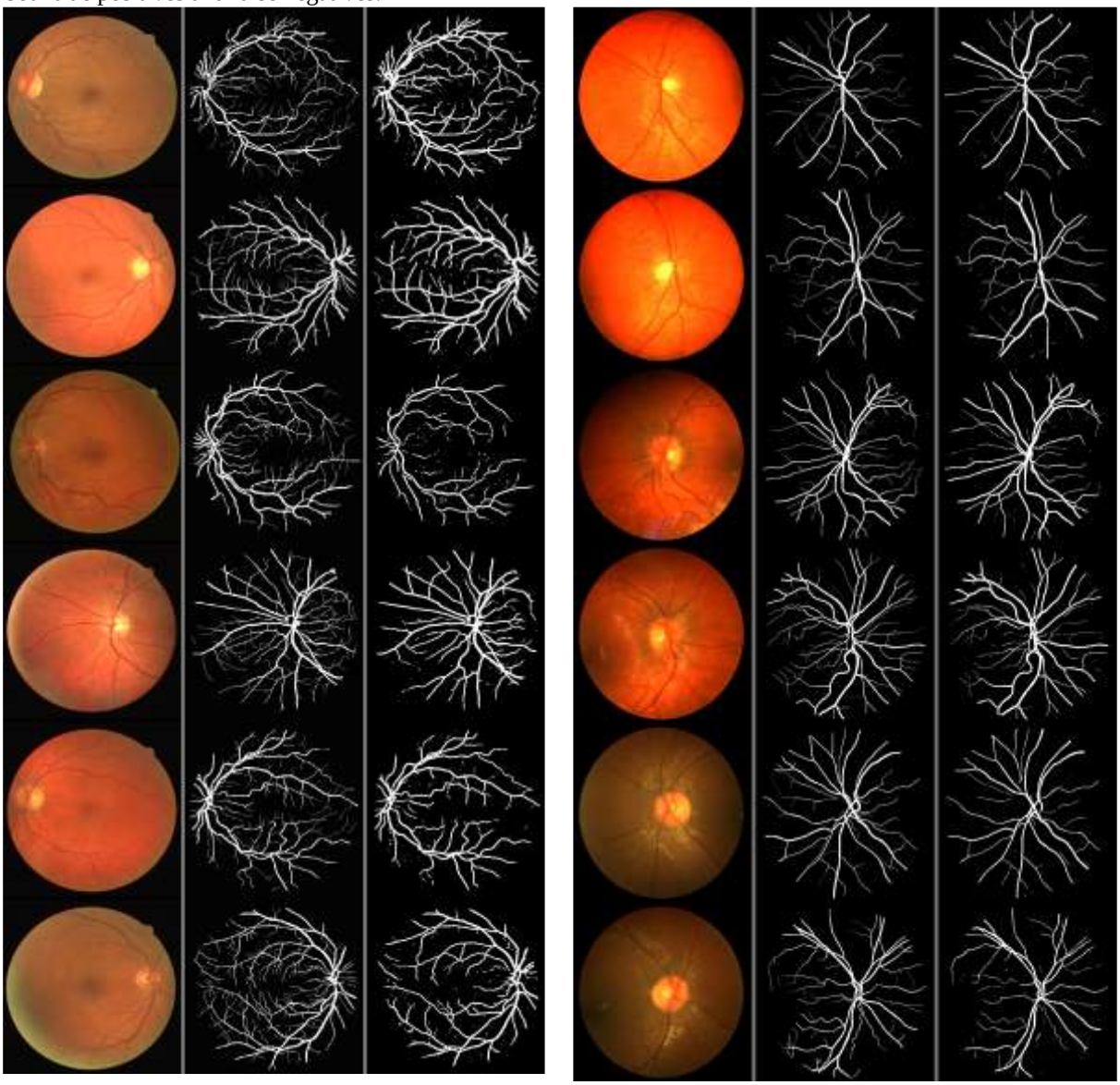

Fig. 3. some samples from each dataset

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

Accuracy is a metric that measures the proportion of correctly predicted instances (both true positives and true negatives) out of the total instances in a dataset. It is calculated as Equation 5:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

Where TP is true positives, TN is true negatives and TOT is total instances. This metric is commonly used to evaluate the performance of classification models.

IV. Results and Discussion

In Figure 3, we present the results for specific samples organized into two distinct sections. Each section includes a set of output images corresponding to the input samples, along with their respective masks. Specifically, Figure 3b displays output samples derived from the CHASEDB1 dataset, while Figure 3a showcases results from the Drive dataset. Each set of images illustrates the output for the input samples alongside their corresponding masks.

Table 2 and Table 3 list the results we obtained from the training with the results of other articles showing an improvement in some results. The model was developed by systematically adding new augmentation techniques in each iteration. These diverse augmentation methods provided the model with fresh information, leading to improved performance. It's worth noting that we did not utilize any pre-trained models or modify the architecture of the models we used, which contrasts with the methodologies employed in other studies. This difference may explain the variations in results when compared to those reported in the articles. Furthermore, we must recognize several negative factors that impacted the outcomes, including the type of camera used for image capture, inaccuracies in slope angle measurements, and potential human errors. All of these aspects significantly influenced the results.

TABLE. II.TESTING RESULTS FOR DRIVE

Study	**Jaccard**	**F1**	**Recall**	**Precision**	**Accuracy**
[15]	-	0.814	0.754	0.885	0.953
[16]	-	0.817	0.782	-	0.956
[17]	0.692	0.818	0.788	-	0.970
Ours	0.644	0.783	0.747	0.829	0.966

TABLE. III.TESTING RESULTS FOR CHASE-DB1

Study	**Jaccard**	**F1**	**Recall**	**Precision**	**Accuracy**
[18]	-	0.814	0.829	0.734	0.958
[19]	0.991	0.795	0.829	-	0.976
[17]	0.988	0.804	0.797	-	0.976
Ours	0.985	0.797	0.797	0.783	0.974

V. CONCLUSION

In conclusion, this paper presents a segmentation model that effectively utilizes a wide range of data augmentation techniques to enhance the performance of the standard U-Net architecture. We have addressed the challenges posed by fundus camera inputs and environmental factors by implementing a comprehensive data augmentation strategy that employs both individual techniques and their combinations for optimal results. Additionally, our approach takes into account the limitations of the U-Net architecture by incorporating various rotated versions of the augmented images. This strategy helps preserve valuable information and reduces potential losses associated with pooling operations. Overall, our findings demonstrate that tailored data augmentation significantly improves segmentation accuracy in retinal image analysis.

REFERENCES

[1]. Z. Wang and J. Yang, "Diabetic retinopathy detection via deep convolutional networks for discriminative localization and visual explanation," in Workshops at the thirty-second AAAI conference on artificial intelligence, 2018.

[2]. E. S. Uysal, M. ¸S. Bilici, B. S. Zaza, M. Y. ¨ Ozgen¸c, and O. Boyar, "Exploring the limits of data augmentation for retinal vessel segmentation," arXiv preprint arXiv:2105.09365, 2021.

[3]. Y. Niu, L. Gu, Y. Zhao, and F. Lu, "Explainable diabetic retinopathy detection and retinal image generation," IEEE journal of biomedical and health informatics, vol. 26, no. 1, pp. 44–55, 2021.

[4]. A. Hatamizadeh, H. Hosseini, N. Patel, J. Choi, C. C. Pole, C. M. Hoeferlin, S. D. Schwartz, and D. Terzopoulos, "Ravir: A dataset and methodology for the semantic segmentation and quantitative analysis of retinal arteries and veins in infrared reflectance imaging," IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 7, pp. 3272–3283, 2022.

[5]. M. N. Getahun, O. Y. Rogov, D. V. Dylov, A. Somov, A. Bouridane, and R. Hamoudi, "Fs-net: Full scale network and adaptive threshold for improving extraction of micro-retinal vessel structures," arXiv preprint arXiv:2311.08059, 2023.

[6]. S. Moccia, E. De Momi, S. El Hadji, and L. S. Mattos, "Blood vessel segmentation algorithms—review of methods, datasets and evaluation metrics," Computer methods and programs in biomedicine, vol. 158, pp. 71–91, 2018.

[7]. M. Voets, K. Møllersen, and L. A. Bongo, "Reproduction study using public data of: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," PloS one, vol. 14, no. 6, p. e0217541, 2019.

[8]. J. Yang, M. N. Nguyen, P. P. San, X. Li, and S. Krishnaswamy, "Deep convolutional neural networks on multichannel time series for human activity recognition." in Ijcai, vol. 15. Buenos Aires, Argentina, 2015, pp. 3995–4001.

[9]. L. Li, M. Verma, Y. Nakashima, H. Nagahara, and R. Kawasaki, "Iternet: Retinal image segmentation utilizing structural redundancy in vessel networks," in Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2020, pp. 3656–3665.

[10]. R.-G. Dumitru, D. Peteleaza, and C. Craciun, "Using duck-net for polyp image segmentation," Scientific reports, vol. 13, no. 1, p. 9803, 2023.

[11]. B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2921–2929.

[12]. M. Alharbi and D. Gupta, "Segmentation of diabetic retinopathy images using deep feature fused residual with u-net," Alexandria Engineering Journal, vol. 83, pp. 307–325, 2023.

[13]. C. Guo, M. Szemenyei, Y. Yi, W. Wang, B. Chen, and C. Fan, "Sa-unet: Spatial attention u-net for retinal vessel segmentation," in 2020 25th international conference on pattern recognition (ICPR). IEEE, 2021, pp. 1236–1242.

[14]. Y. Ma, Z. Zhu, Z. Dong, T. Shen, M. Sun, and W. Kong, "Multichannel retinal blood vessel segmentation based on the combination of matched filter and u-net network," BioMed research international, vol. 2021, no. 1, p. 5561125, 2021.

[15]. [A. Khanal and R. Estrada, "Dynamic deep networks for retinal vessel segmentation," Frontiers in computer science, vol. 2, p. 35, 2020.

[16]. Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "Dunet: A deformable network for retinal vessel segmentation," Knowledge-Based Systems, vol. 178, pp. 149–162, 2019.

[17]. 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: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer, 2015, pp. 234–241.

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

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

Paper received 16 April 2026.

Ammar Asad, Ph.D. student, Moscow Polytechnic University, (e-mail: ammarasad09@gmail.com).

Zeinab Khadra, Ph.D. student, Moscow Polytechnic University, (e-mail: zeinabkhadrah@gmail.com).
Mohammed Hammoud, Ph.D. student, Skolkovo Institute of Science and Technology, (e-mail: Mohammed.Hammoud@skoltech.ru).
Ola Haydar, Ph.D. student, National Research University of Electronic Technology «MIET», (e-mail: oulahaider@gmail.com).
Sergey S. Lupin, Ph.D., Ass. Professor, National Research University of Electronic Technology «MIET», (e-mail: Sergeylupin.jr@gmail.com);