Article

A Comprehensive Approach for an Interpretable Diabetic Macular Edema Grading System Based on ConvUNext

Zaira Garcia-Nonoal [1], Atoany Fierro-Radilla [2] and Mariko Nakano [1,*]

1 Graduate Section, Mechanical Electronic Engineering School, Instituto Politecnico Nacional, Av. Santa Ana 1000, Mexico City 04440, Mexico; zgarcian2100@alumno.ipn.mx
2 Mechatronics Department, Cuernavaca Campus, Tecnologico de Monterrey 2, Cuernavaca 62790, Morelos, Mexico; afierror@tec.mx
* Correspondence: mnakano@ipn.mx

Abstract: Diabetic macular edema (DME) is a leading cause of vision impairment in diabetic patients, necessitating a timely and accurate diagnosis. This paper proposes a comprehensive system for DME grading using retinal fundus images. Our approach integrates multiple deep learning modules, each designed to address key aspects of the diagnostic process. The first module employs the ConvUNeXt model for segmenting hard exudates (HaEx), crucial indicators of DME. The second module uses RetinaNet for precise optic disc (OD) localization, which is essential for subsequent macula localization. The third module refines macula localization, leveraging preprocessing techniques to enhance image clarity. Finally, our system consolidates these results to provide interpretable DME grading. Experimental evaluations were conducted on the Messidor dataset, demonstrating the system's robust performance. The HaEx segmentation module achieved a mean IoU of 70.5% and a Dice coefficient of 0.64. The OD localization module showed perfect accuracy, recall, and precision at 1.0. For macula localization, our method satisfied the 1R criterion with 99.38% accuracy. The DME grading module achieved an overall accuracy of 91.12%, with an AUC of 0.9334. Our method offers a balanced performance across accuracy, sensitivity, and specificity compared to other non-interpretable and partially interpretable methods.

Keywords: diabetic macular edema; fundus images; deep learning; hard exudates; optic disc; macula; DME grading

1. Introduction

Diabetes mellitus, colloquially termed diabetes, is a persistent medical ailment caused by the body's incapacity to effectively modulate blood glucose levels. As per the International Diabetes Federation, approximately 171 million adults globally were reported to be living with diabetes in 2000. Projections anticipate a surge to 366 million by the year 2030 and 700 million by the year 2045 [1,2]. Diabetic retinopathy (DR) stands as a significant complication capable of culminating in partial vision loss and, in severe cases, total blindness. Initially, DR may not exhibit symptoms, but over time, it can exacerbate and lead to the loss of vision [3]. DR can manifest various abnormalities in the retina, such as microaneurysms, hard and soft exudates, hemorrhages, neovascularization, and macular edema. Generally, DR is categorized into four stages according to its severity, which are mild non-proliferative DR, moderate non-proliferative DR, severe non-proliferative DR, and proliferative DR (PDR) [4].

Diabetic macular edema (DME), an associated complication of DR, can be regarded as an additional stage of DR. DME typically results from the leakage of tissue fluid from macular vessels or retinal thickening, manifesting at any DR stage [5]. The leakage of tissue fluid causes Hard Exudates (HaExs), which is a typical lesion related to DME. This lesion forms clusters with varying sizes and shapes, and it can appear in any position of

the retina. Following the clinical grading standard for DME, the positional relationship between HaExs and the macular center serves as a critical criterion for severity grading. The severity of DME has three grades: grade 0, grade 1, and grade 2, representing no visual DME, non-clinically significant macular edema (NCSME), and clinically significant macular edema (CSME), respectively [6]. The DME grading provided by the Messidor website is given in Table 1. Independent of the DR grade, the early and timely detection of CSME is very important to avoid partial and total vision loss because the macula region is responsible for acquiring clear visual information, and any damage to this region would cause important visual impairment. The ophthalmologist first checks whether the quality of the fundus image is sufficient to make a diagnosis. Once it is determined that the image quality is adequate for diagnosis, the ophthalmologist examines the presence of HaExs within the macular area to determine if the grade is CSME or not to decide on an appropriate treatment [7].

Table 1. DME grading criteria provided by [5].

Grade	Severity	Details
0	Healthy	No visible HaExs near the macula
1	NCSME	Distance between macula and HaExs > one papilla diameter
2	CSME	Distance between the macula and HaExs <= one papilla diameter

In the literature, there are two approaches for DME grading: one is handcrafted feature-based methods [8–11] and the other involves deep neural networks [5,12–20]. In the early stage of the DME grading schemes, almost all proposals belong to the former approach, in which the two distinct stages, the feature extraction stage and the grade assignment stage using the extracted features, are required to carry out the DME grading. Recently, with the rise of deep learning, several proposals have begun to use different structures of deep neural networks for DME grading tasks. This approach uses a so-called end-to-end strategy, integrating both processes within the algorithm itself. Generally, algorithms based on deep neural networks provide higher grading accuracy compared with handcrafted-based methods. However, algorithms based on deep neural networks can be seen as a black box for users [21,22] and the decision-making process taken by AI-based algorithms is somewhat opaque. In computer-aided diagnostics in the medical field, obtained classification or grading results must be interpretable for clinicians, who finally make the decision to provide the most appropriate treatment to his/her patients.

In this paper, we propose an interpretable DME grading system to support ophthalmologists in making a decision about the most adequate treatment. The proposed system is composed of four stages. In the first stage, HaExs, key lesions of DME, are segmented using optimized ConvUNext, and an optic disc (OD) is detected using RetinaNet, which is the second stage. Then, using OD positional information obtained in the second stage, the macula region is localized in the third stage. Finally, as the fourth stage, combining all results of the three previous stages, a DME grade is assigned. Considering real application, in which retinal images are captured by different fundus cameras under different environments, we applied two different preprocessing to enhance the input image. The majority voting scheme is applied to obtain the final DME grade with a high confidence rate, obtaining a performance accuracy of 91.12%, sensitivity of 91.12%, and specificity of 93.00%.

While Optical Coherent Tomography (OCT) has become a reliable method for efficiently detecting DME, we opted to utilize retinal images as input data due to the high cost of OCT equipment. In many developing countries, this equipment is only accessible in major hospitals or specialized eye clinics located in urban areas. Given the potential

application of our proposed system in remote areas or within telemedicine frameworks, retinal images serve as more suitable input data.

The rest of the paper is organized as follows. In Section 2, we briefly describe some related works previously reported in the literature and a detailed description of the proposed system is provided in Section 3. Experimental results are shown in Section 4, and in Section 5, we provide a discussion about some limitations and challenges of the proposed algorithm. Finally, the conclusions and future works derived from this paper are provided in Section 6. Additionally, mathematical developments of some concepts are provided in the Appendix A.

2. Related Works

In this section, we describe previous works on the DME grading system. As mentioned before, previous works are categorized into two approaches: a handcrafted feature-based approach and a deep learning-based approach. Some works try to increase the interpretability of the results obtained using the deep learning-based approach. We also mention these works in this section.

2.1. Handcrafted Feature-Based Approach

Almost all the early works rely on handcrafted feature-based approaches. Despite their limited performance, these approaches are similar to the methodology used by ophthalmologists to determine the severity of DME. In this sense, the results obtained can be interpretable for them. The authors of [8] proposed an automatic DME grading system utilizing the Particle Swarm Optimization (PSO) algorithm for HaExs segmentation. The detection of the OD and fovea relies on mathematical morphology, and the severity of the DME is determined based on the locations of lesions HaExs with respect to the OD. In reference [9], texture features are extracted from the area surrounding the macula, and DME grading is performed using a Support Vector Machine (SVM), achieving an accuracy rate of 86%. The authors of [10] utilized CLAHE (Contrast Limit Adaptive Histogram Equalization) for image enhancement. Following this, HaExs were identified through a segmentation technique, and the DME grading was conducted based on the spatial location of HaExs within the macula region. Another handcrafted-based approach outlined in [11] involves localizing the macula using its anatomical features and determining its location with respect to the OD. Subsequently, HaExs are detected using a vector quantization technique and a semi-supervised method, followed by DME grading based on the location of HaExs and the macula coordinates.

2.2. Deep Learning-Based Approach

Deep learning, especially convolutional neural networks (CNNs), is gaining widespread popularity and finding increased application in various domains, notably in medical image processing. This approach, distinct from traditional feature engineering techniques described in Section 2.1, addresses challenges such as medical image segmentation, detection, and classification, achieving substantial success. To diagnose various ocular diseases from fundus images, different algorithms based on deep learning have been developed, generally showing high performance [23–27]. The deep learning-based approach lies in its ability to automatically learn feature extraction through a backbone network, primarily composed of convolutional and pooling layers. An antecedent study employing deep learning in DME grading is found in [12], where the authors presented a deep neural network comprising 13 convolutional layers for feature extraction and two dense layers dedicated to DME grading. The model demonstrated noteworthy performance, achieving accuracy, sensitivity, and specificity values of 88.8%, 74.7%, and 96.5%, respectively.

Some other authors demonstrated that ensemble methods perform well for DME grading, as in the work outlined in [18], where the authors advocated a two-stage approach. In the initial stage, the algorithm discerns the presence or absence of DME. Upon confirming the presence of DME, the image is then subjected to the second stage, where

grading is conducted based on the severity of the condition. Both stages integrate the proposed technique, termed the Hierarchical Ensemble of Convolutional Neural Networks (HE-CNN). The results demonstrated notable accuracy, sensitivity, and specificity values of 95.47%, 94.68%, and 97.19%, respectively.

In due course, natural language processing algorithms found applications in computer vision, leveraging their notable performance in natural language applications. Consequently, certain researchers began implementing these algorithms for DME grading. A notable example is the research documented in [5], where the authors introduced an innovative end-to-end architecture incorporating ResNet50 along with channel attention (SENet) for feature extraction. Moreover, they introduced a disease attention module to enhance disease-specific information related to DME. The outcomes demonstrated an accuracy value of 97.06%. In other research presented in [20], the authors introduced a three-component system encompassing the backbone, the neck, and the head. The Swin Transformer served as the framework's backbone, incorporating window multi-head self-attention (W-MSA) and mobile window multi-head self-attention (SW-MSA). The neck of the framework was composed of Global Average Pooling (GAP), while the head utilized linear CLS (class token). This modular design contributes to the model's simplicity and ease of training, as it ensures a clear mapping between features and categories. The results were 98.66%, 98.66%, 99.32%, and 98.66% for accuracy, sensitivity, specificity, and F1-score, respectively. It is worth mentioning that the above four methods use the public Messidor data set to grade the three severity degrees of DME. Other methods [13,15,16,19] use private data sets or their classification is binary (presence or absence of DME).

2.3. Interpretable Machine Learning Approach

Instead of treating a deep learning model as a black box, some authors have pursued an approach where deep learning algorithms are only employed as extractors of relevant information about the DME. They utilize these extracted features to conduct DME grading, employing a conventional machine learning classifier. This method offers some degree of interpretability since the DME grade assignment is based on the retinal anatomical features extracted in the first stage.

For example, in [14], the authors advocated the utilization of a deep learning model to extract features from fundus images. The training process involved incorporating these multi-scale features from the image, such as hard exudate masks, macular masks, and macular images. Subsequently, XGBoost is used for DME grading, which introduces these data. This work can be considered partially interpretable because input data, such as the positional relationship between HaExs and macula, is interpretable data, which is generally used by ophthalmologists to determine the severity of the DME. However, the decision-making performed by XGBoost is still a black box.

In reference [17], the authors developed an end-to-end deep fusion model based on EfficientNet as a backbone network. This network is used not only for binary classification, DME or non-DME but also HaExs localization, combined with a bi-directional feature pyramid network (BiFPN). This fusion model was trained using a loss function, which is a linear combination of classification loss and detection loss. Although the decision about the presence or absence of DME was taken by a black box like CNN, the fact that a part of the same CNN is used for the detection of HaExs increases the interpretability of results. Additionally, the authors detected optic discs and macula using YOLOv3 to provide additional information for ophthalmologists, although the system does not provide the three grades of the DME shown in Table 1.

3. Proposed System

The proposed system consists of four modules: the HaEx segmentation module, optic disc localization module, macula localization module, and finally the DME grading module that uses all the information provided by the previous three modules. The input image obtained from ophthalmoscopy is pre-processed to crop the unwanted background which

is introduced into three modules: the HaEx segmentation module, optic disc localization module, and macular localization module. Each module executes different tasks with the intention of combining the results to create an interpretable diagnosis decision to help ophthalmologists determine the grade of DME from eye fundus images that are the input of our system. Figure 1 shows a diagram of the proposed system with all modules and their relationship. Each module will be described in detail in the following subsections.
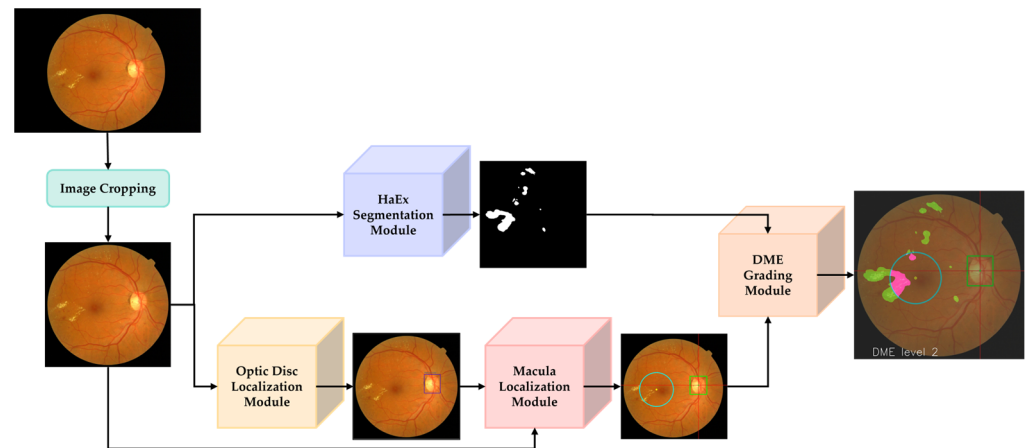

Figure 1. Proposed System Diagram. The DME grading is performed using the macular region (circle) and optic disc (square).

3.1. Hard Exudates (HaExs) Segmentation Module

The objective of the first module is to segment HaEx lesions utilizing CNN from RGB fundus images to obtain a binary mask of these lesions. Figure 2 presents the processes inside the module in detail.
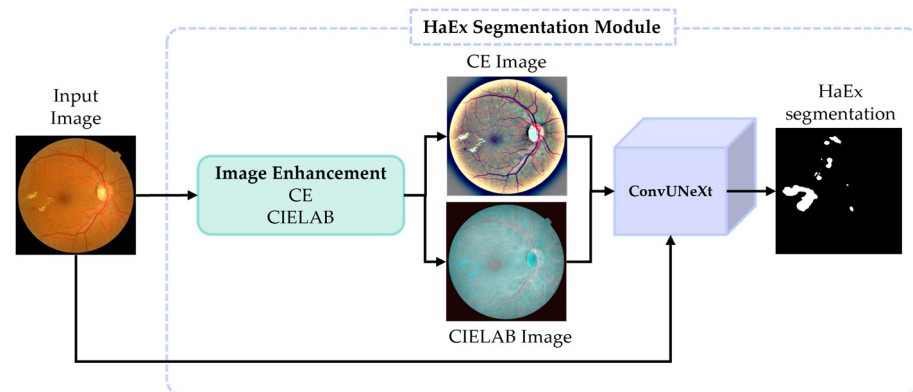

Figure 2. HaEx segmentation module subprocesses.

For dataset preparation, three datasets, Messidor [28], Retinal Lesions [29], and DDR [30] were considered. Retinal Lesions and DDR are public datasets with RGB retinal images and ground truth (GT) masks of DR lesions, which are labeled based on the consensus of several ophthalmologists [29,30]. In the case of Messidor, the GT masks are privately generated by two ophthalmologists based on the Messidor RGB images. The final GT masks were decided by their consensus. Each dataset provides different types of ground truth masks for HaExs. For example, Messidor and Retinal Lesions provide rough segmentation, grouping agglomeration of exudate spots. Meanwhile, in the DDR dataset, each exudate spot is segmented separately. Figure 3 shows examples of images and corresponding GT masks from three datasets. We decided to use the Messidor dataset and Retinal Lesions dataset to train and evaluate the HaEx segmentation module because the amount and shape of the leak of lipids play an important role in causing HaEx lesions identified by

yellowish-white deposits [31]. Therefore, rough segmentation used in Messidor and Retinal Lesion is more adequate for the segmentation module.
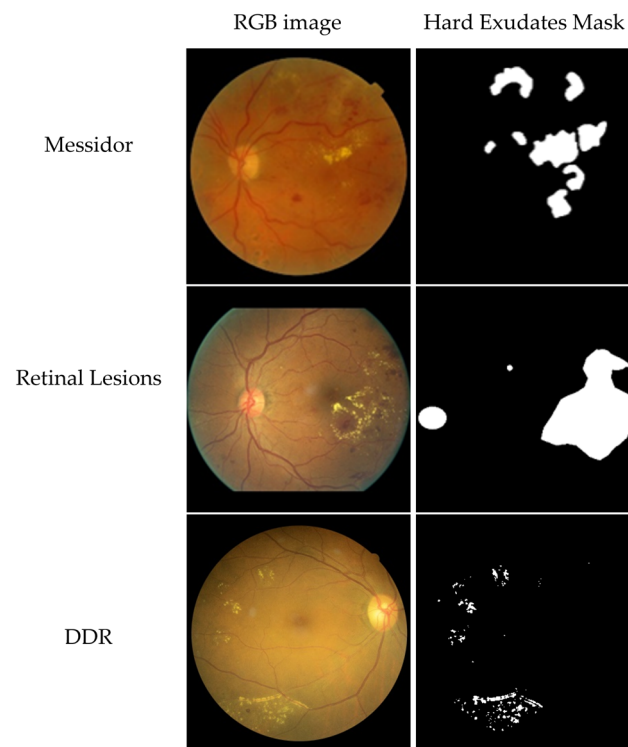

Figure 3. Example of an image for each dataset and its respective GT mask for HaEx segmentation.

A cropping process was employed on the images to keep only the region of interest, removing the greater black background as shown in Figure 1. Then the datasets are divided into two sets: train and test with approximately 80% and 20% of the images. Table 2 presents the number of images in training and test sets.

Table 2. Datasets and their respective division into training and test sets.

Dataset	Training Set	Test Set
Messidor [28]	900	236
Retinal Lesions [29]	400	96

ConvUNeXt [32] was used to achieve our goal of accurate HaEx segmentation because it is a model based on UNet for medical image segmentation with the advantage of reducing the number of parameters and enhancing the convolution blocks by using large convolution kernels and depth-wise separable convolution. This parameter reduction allows ConvUNeXt to achieve a lightweight design while maintaining segmentation performance. It also integrates a gating mechanism for feature fusion and a lightweight attention mechanism (Attention Gate) to filter out noise and highlight relevant information. These mechanisms enhance the model's ability to capture both high-level and low-level semantic information, resulting in more accurate segmentation outcomes [32]. The attention gate mechanism controls the importance of features at different spatial locations, allowing the model to suppress irrelevant areas in the input image while highlighting relevant features in specific local areas, leading to improved accuracy in medical image segmentation tasks, as shown by a review paper [33].

The quality of retinal images is highly variable and influenced by several factors, including the type of ophthalmoscopy employed, the precision of the ophthalmoscopic focus, and the environmental conditions during image acquisition, such as the level of

illumination. Therefore, in some cases, the preprocessing stage to enhance HaEx lesions is required to obtain accurate segmentation. We consider two image-enhancing methods: contrast enhancement (CE) [34] and the conversion in CIELAB color space from RGB. The CE preprocessing is given by (1).

$$I_{CE}(x, y) = \alpha I(x, y) + \beta G(x, y; \sigma) \circledast I(x, y) + \mu \tag{1}$$

where I_{CE} is the enhanced image and I is the original image, the values of α, β, σ and μ are constant values set as 4, -4, $300/30$, and 128, respectively, and $G(x, y; \sigma)$ is a Gaussian filter with the variance σ. The operation $\circledast$ means convolution. This process involves subtracting the Gaussian-filtered image from the original image to enhance contrast, with μ serving as a baseline shift for the grayscale.

The second preprocessing method started by converting the RGB images to CIELAB color space. From the A and B channels, we calculated the minimum and maximum pixel values and the a and b variables, respectively. We applied a linear transformation to enhance the contrast of the image including a gamma correction term (2).

$$y = \left(\frac{x - a}{b - a} \right)^{\gamma} (d - c) + c \tag{2}$$

where γ is the gamma value used for the transformation. If $\gamma > 1$, the transformation compresses the range of pixel values towards the lower end, making the image appear darker. The variable x corresponds to the input in channels A or B, and the c and d values are set to 0 and 255. The final image I_t is obtained concatenating two times the transformed channel B, y_B and one the transformed channel A, y_A as shown by (3).

$$I_t = concat(y_B, y_B, y_A) \tag{3}$$

Because the level of distortion of the incoming retinal images cannot be known in advance, we trained three ConvUNeXt for each input image: input image without preprocessing, input image enhanced by CE, and input image in CIELAB color space. Hyperparameters of each of the three ConvUNeXt are adjusted empirically, which are shown in Table 3. We denominated these three networks as ConvUNeXt-Original, ConvUNeXt -CE, and ConvUNeXt-CIELAB.

Table 3. Hyperparameters used to train three ConvUNeXts according to three treatments of input images.

Hyperparameter	ConvUNeXt-Original	ConvUNeXt-CE	ConvUNeXT-CIELAB
Learning rate	1×10^{-4}	1×10^{-3}	1×10^{-3}
Weight decay	5×10^{-5}	5×10^{-4}	5×10^{-4}
Epochs	100	100	100
Loss function	Combination of Cross-Entropy and Dice loss		
Optimizer	AdamW		
Batch size	2		
Data augmentations	Random resize, random crop, horizontal and vertical flip		

3.2. Optic Disc (OD) Localization Module

This module aims to localize the Optic Disc and obtain its coordinates to use in the next module for calculating the macula area. To conduct the experiments, the DRIVE [35] and Messidor datasets were selected under the supervision of ophthalmologists.

In this case, we define RetinaNet as a model due to its strong performance in detecting small objects and capability for accurate object localization with high precision and recall rates [36]. RetinaNet achieves this by utilizing a feature pyramid network (FPN) due to its ability to provide multi-scale feature representation, enhance localization accuracy, improve

efficiency, adapt to object scales, and capture contextual information. By leveraging the hierarchical structure of the pyramid-shaped network, object detectors like RetinaNet can achieve high accuracy and robust performance in detecting objects of varying sizes and complexities, as well as a Focal Loss function that introduces a modulating factor to the cross-entropy loss given by (4).

$$FL(p_t) = -(1 - p_t)^\gamma \log(p_t) \tag{4}$$

where p_t is the probability of the ground truth class, which is the output of the classification model for the true class label, $(1 - p_t)^\gamma$ is the Modulating factor that down-weights the loss for well-classified examples (where $p_t > 0.5$) and amplifies the loss for hard, misclassified examples (where $p_t \leq 0.5$) and γ is the focusing parameter that controls the modulating term. When $\gamma = 0$, the Focal Loss reduces to the standard cross-entropy loss.

The Focal Loss function is crucial in object detection tasks due to its ability to address class imbalance, focus on hard examples, improve learning dynamics, enhance object localization, and provide flexibility for better generalization. By incorporating this function, object detectors like RetinaNet can achieve higher accuracy, robustness, and efficiency in detecting objects in challenging real-world scenarios. These characteristics make RetinaNet suitable for applications where accurate object localization is essential, and it is also suitable for real-time applications.

For the training phase, we used DRIVE, split into training and test sets with 32 and 8 images, respectively. The hyperparameters used to train the model were: a learning rate of 1×10^{-4}, batch size of 4, 8 steps, and 50 epochs with a backbone of resnet50 using pre-trained weights on the CoCo dataset [37].

After the detection of OD, the center of OD and diameter of OD (DD) are calculated as follows.

$$COD_x = BB_{xmin} + \frac{(BB_{xmax} - BB_{xmin})}{2} \tag{5}$$

$$COD_y = BB_{ymin} + \frac{(BB_{ymax} - BB_{ymin})}{2} \tag{6}$$

where $\left(COD_x, COD_y\right)$ is the coordinate of the center of OD, $\left(BB_{xmin}, BB_{ymin}\right)$ and $\left(BB_{xmax}, BB_{ymax}\right)$ are coordinates of the lower-left and upper-right of the bounding box obtained by RetinaNet.5

The OD diameter (DD) was calculated according to (7), using BB_{xmin} and BB_{xmax}:

$$DD = BB_{xmax} - BB_{xmin} \tag{7}$$

3.3. Macula Localization Module

As mentioned before, the functionality of this module relies upon the outcomes generated by the OD Localization Module to conduct calculations pertinent to macula localization. Figure 4 illustrates the subprocesses inside the module.

Due to the variations in color intensity and luminosity within fundus images, which consequently impact the visibility of the macula, we decided to employ two preprocessing methodologies aimed at enhancing the color representation of the macula region. First, we selected an image from the DRIVE dataset with suitable color and brightness for macula discrimination. Secondly, we applied the iterative robust homomorphic surface fitting (IRHSF) preprocessing [38] to the selected image to create a new one that serves as a reference to perform a histogram specification across the entire training dataset. The creation of the reference image was performed only once, and it is used to apply the histogram specification to any other retinal images. The details of the IRHSF preprocessing and histogram specification are described in the Appendix A.
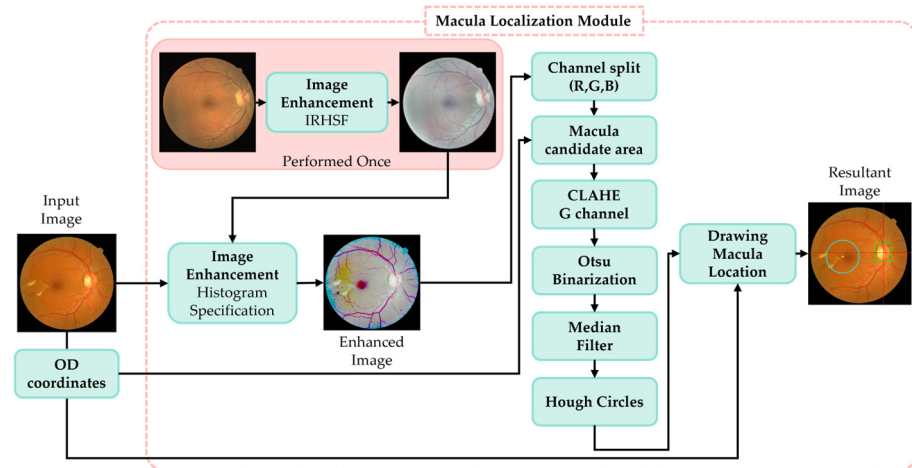
Figure 4. Macula Localization module subprocesses indicated by a red dot-line box.

The initial images or their preprocessed counterparts, the histogram specification results, and coordinates corresponding to the OD were employed as inputs for this module to determine the macula area. To begin with, the calculations of the images, the OD center coordinates, and the DD were parsed, followed by the pre-processed images to split them into three channels: red, green, and blue.

According to the literature the center of the macula, the fovea, is located at approximately 2.5 DD from the center of the OD with a little horizontal angle and is the darkest area of a fundus image. For this reason, a candidate area for its localization is defined as a rectangle of 2 DD length and 2.5 DD width as shown in Figure 5 [39].
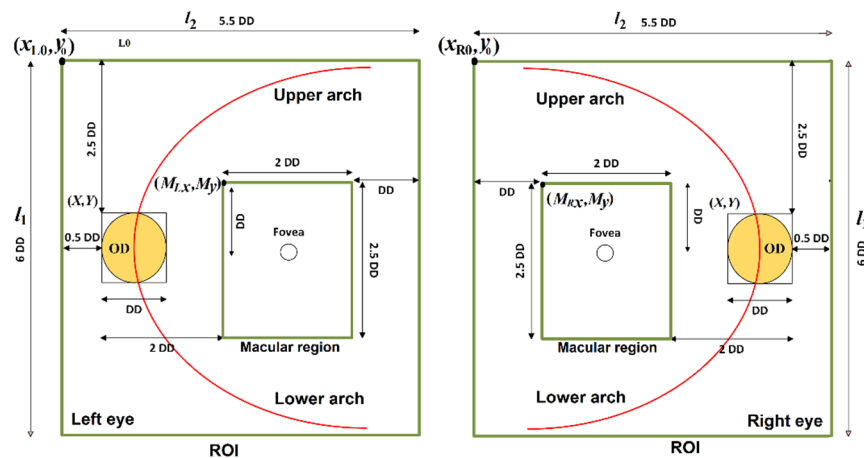

Figure 5. Simplified anatomical structure of human retinal image [39].

Additionally, image processing techniques were used to highlight the exact location of the fovea within the designated rectangle. This entails employing contrast enhancement methods, such as CLAHE, in the green channel of the image to produce a new version. Subsequently, a binary image of the previous outcome was created using Otsu binarization to choose the optimal threshold value for each image. Finally, a median blur filter was applied. The Hough circles algorithm detected the circles and their centers within the result pre-processed designated area. In cases where multiple circles were identified, they were averaged to obtain a single result. This one was considered the center of the fovea. To end the process in the original fundus image the quadrants, OD, OD center, the macula area defined as a circle of 1 DD, and fovea were drawn.

3.4. DME Grading Module

This module consolidates the outcomes obtained from all preceding components to assign the DME grading. To perform an interpretable DME grading, HaEx masks segmented by the HaEx segmentation module, and the macula area localized by the macula localization module are combined. The macula area is inside of a circle whose center is the fovea and whose diameter is 1DD as mentioned in Section 3.3. The assessment was conducted by determining whether the distance between the segmented HaEx masks and the position of the macula signifies a potential danger based on the criterion presented in Table 1. It means that if some pixels of the HaEx mask are located within the macula area, grade 2 (CSME) is assigned. Grade 1 (NCSME) is assigned if all pixels of the HaEx mask are located outside the macula area and grade 0 (Healthy) if no HaEx mask is presented in the image.

Once the grading was performed, HaExs within the macular region were accentuated using different colors corresponding to the designated DME severity grade, as well as the plotted macular area and OD position bounding box. Thus, facilitating comprehensible grading outcomes for ophthalmologists.

4. Experimental Results

In this section, we present the findings from our proposed system for DME grading based on retinal fundus images. The results were analyzed in terms of accuracy, precision, recall, and other relevant metrics. The following subsections detail the performance of each module, demonstrating the effectiveness and robustness of our approach. The results highlight the segmentation accuracy of HaEx lesions, the precision of OD detection and macula localization, and the reliability of the DME grading module. Visual examples and quantitative metrics are provided to support our findings. These results are crucial for validating the efficacy of our system and its potential application in assisting ophthalmologists with DME diagnosis.

4.1. Hard Exudates (HaEx) Segmentation Module Results

The HaEx segmentation module, employing the ConvUNeXt model, demonstrated effective performance in segmenting lesions from RGB fundus images. The evaluation was conducted using the Messidor dataset, described in the methodology section.

In Figure 6, we can observe examples of test results evaluating Messidor's original images with the ConvUNeXt-trained segmentation model.
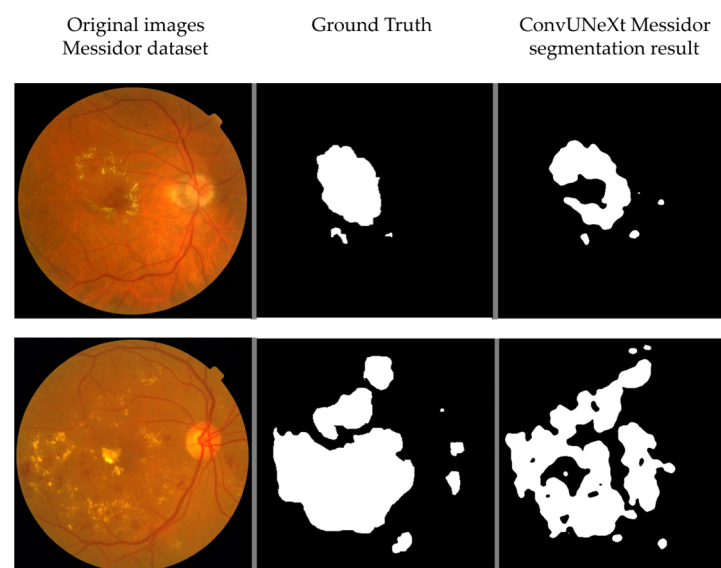

Figure 6. Messidor dataset results after training ConvUNext.

The results of both preprocessing methods, Contrast Enhancement (CE) and CIELAB, improve identifying areas with hard exudates as seen in the first row of Figure 7, while in images without exudates, such as those shown in the second row, there are no highlighted areas in yellow or blue depending on the preprocessing applied.
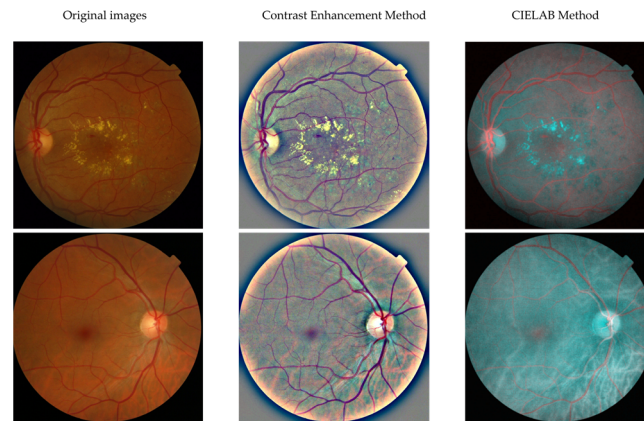
Original images   Contrast Enhancement Method   CIELAB Method

Figure 7. Results of HaEx enhancement.

The metric values of the three models are shown in Table 4. Both preprocessing methods, CE and CIELAB, improve the IoU, indicating enhanced capability to detect affected areas.

Table 4. Metrics results for HaEx segmentation of three models.

Model	Mean IoU	Dice Coefficient
ConvUNeXt-Original	69.5	0.64
ConvUNeXt-CE	70.4	0.34
ConvUNeXt-CIELAB	70.5	0.42

4.2. Optic Disc (OD) Localization Module Results

The OD localization module utilizing RetinaNet demonstrated robust performance in detecting the OD, as evidenced by the metrics in Table 5. Figure 8a–d provides examples of the OD localization results from the Messidor dataset.

Table 5. Metrics results for OD localization using IoU threshold of 0.5.

Method	mAP	Accuracy	Recall	Precision
Basit, A. [40]	-	0.9861	-	-
Ali, H.M. [41]	0.996	1.0	0.996	0.996
Our Method	1.0	1.0	1.0	1.0

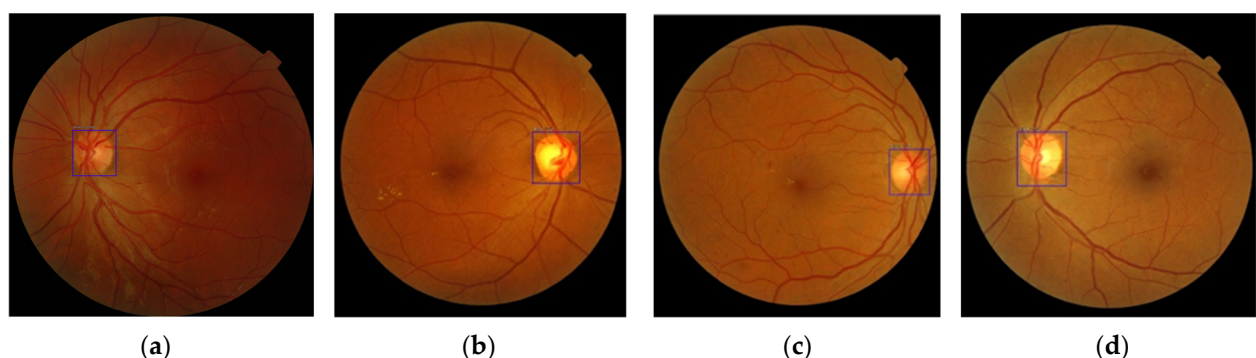(**a**)   (**b**)   (**c**)   (**d**)

Figure 8. OD localization Module results. ODs are localized correctly by the square boxes.

4.3. Macula Localization Module Results

To assess the performance of the macula localization, we utilized the macula annotations provided by [42] for a subset of 1136 images. For quantitative results, we calculated the 1R criterion, which refers to a specific method used in medical imaging, particularly in ophthalmology, for identifying the macula in retinal images. The "1R" criterion refers to the difference between the ground truth (GT) foveal position and the predicted foveal position, which should be within one optic disc (OD) radius. We computed the 1R criterion using the ground truth annotations and predictions. Table 6 contains the 1/2R, 1R, and 2R results and a comparison with two previous methods that used the same 1136 images.

Table 6. Comparison of macula detection scores in percentage.

Method	1/2R	1R	2R
Aquino, A. [43]	91.28	98.24	99.56
Molina-Casado, J. [44]	96.08	98.58	99.50
Our Method	79.40	99.38	99.91

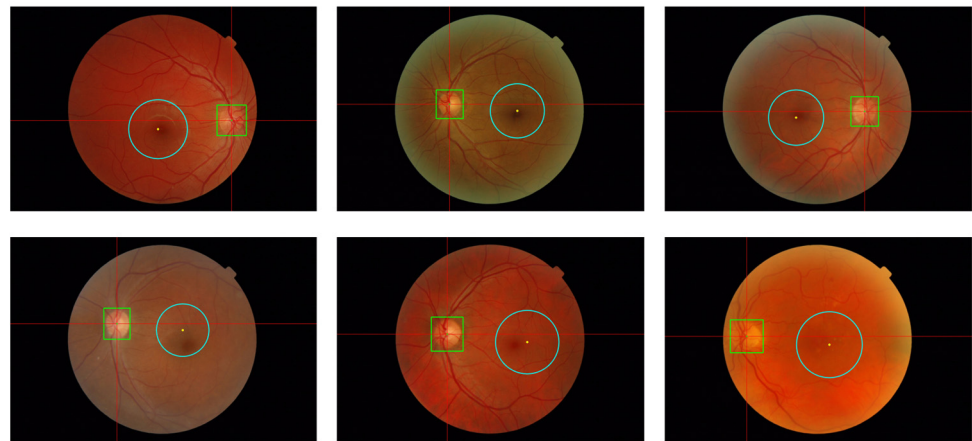Figure 9 illustrates examples of the macula localization outcomes using the Messidor dataset.

Figure 9. Images whose macula localization satisfies the 1R and 1/2R criterion. The circles indicate the detected macular regions and squares indicate the ODs.

The value achieved by our method significantly ensures precise identification of the macula, which is crucial for the accurate grading of DME. Such a high accuracy provides consistent and reliable results, reducing variability and enhancing the dependability of diagnoses across different images and patients. Furthermore, automated systems with this level of precision can screen large volumes of retinal images swiftly and efficiently, aiding in the early detection and management of DME. By minimizing errors in identifying the macula and grading DME, these systems yield more reliable results. Finally, precise macula detection facilitates consistent monitoring of the progression or regression of DME, enabling effective follow-up care and treatment adjustments.

4.4. DME Grading Module Results

We employed the Risk of Macular Edema data from the Messidor dataset to evaluate our results. Table 7 displays the performance metrics obtained from three experiments conducted with different preprocessing methods: No preprocessing (Original), Contrast Enhancement (CE), and CIELAB preprocessing of images. Majority voting is used to calculate metric values, especially in ensemble methods. In this approach, multiple models are trained on the same dataset, and their predictions are combined to make a final decision. This method has benefits, such as improving prediction accuracy by leveraging the strengths

of multiple models and reducing the impact of noise or incorrect predictions from individual models.

Table 7. DME grading metrics results.

Metric	Value
Accuracy	91.12%
Precision	91.31%
Recall	91.12%
F1-Score	90.72%
Specificity	93.00%
Sensitivity	91.12%
AUC	0.9334

Figure 10 presents examples of the grading results obtained with the proposed system for each case.
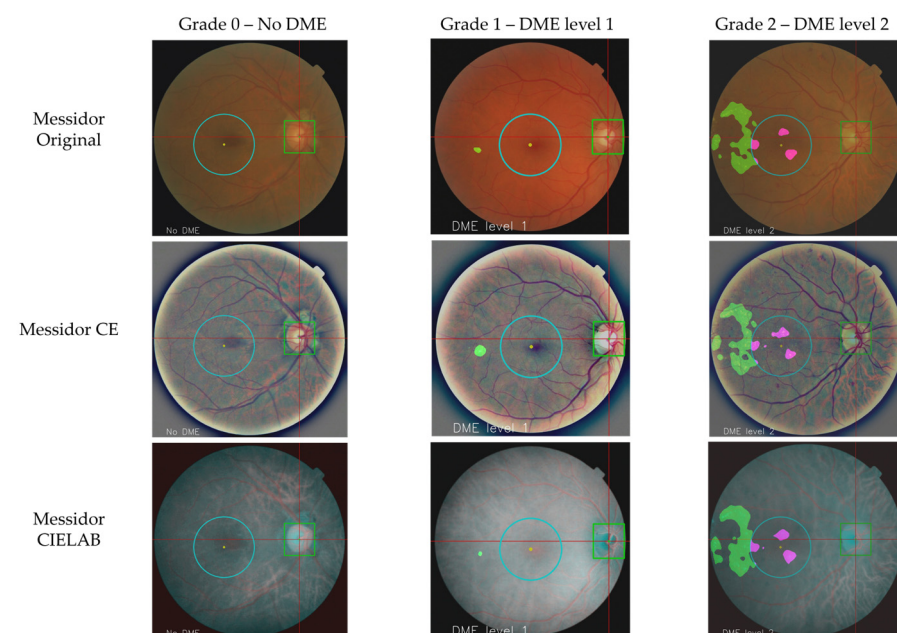

Figure 10. Results of HaEx enhancement. The circles indicate the detected macular regions and squares indicate the ODs. HaExs within the macular region are depicted in pink and outside are depicted in green.

Finally, a comparison with other non-interpretable and interpretable methods is presented in Table 8.

Table 8. Comparison of the proposed method with other non-interpretable and interpretable DME grading methods on MESSIDOR.

Method	Accuracy	Sensitivity	Specificity
Non-interpretable methods			
Al-Bander, B. [12]	88.8%	74.4%	96.5%
Singh, R.K. [18]	95.47%	94.68%	97.19%
Fu, Y. [5]	97.06%	-	-
Yao, Z. [20]	98.66%	98.66%	99.32%
Partially Interpretable Methods			
He, X. [14]	96.33%	-	-
Wang, T.Y [17]	86.3%	79.5%	87.7%
Totally Interpretable Method			
Our Method	91.12%	91.12%	93.00%

The results show that non-interpretable methods generally perform better in accuracy, sensitivity, and specificity. While our method shows slightly lower accuracy than some of the non-interpretable methods, it maintains balanced sensitivity and specificity, suggesting a robust performance in DME grading. However, interpretable methods like ours demonstrate competitive performance while providing insights into the modules, which may be advantageous in certain contexts, such as clinical decision-making or model interpretability.

4.5. Ablation Studies

In this section, we assessed the significance and impact of various image preprocessing techniques in the HaEx Segmentation and Macula Localization modules on the overall performance of the DME grading system. For the HaEx Segmentation module, consider a scenario where the original fundus images have poor contrast and varying illumination. Without preprocessing, the segmentation model struggles to distinguish HaEx from other retinal structures, resulting in suboptimal performance. Enhancing the quality of the input images using contrast enhancement (CE) and CIELAB preprocessing methods leads to more accurate segmentation. These results are shown in Table 2.

Because input image quality is unknown, we cannot select an enhancement method for all possible input images. To alleviate the negative impact of the low-quality input image on the DME grading accuracy, we introduced the majority voting mechanism to the DME grading. Table 9 shows the DME grading accuracy and AUC obtained using two enhancement methods individually and without any enhancement method which are compared with the proposed majority voting. From Table 9, we can observe that the majority voting mechanism is necessary.

Table 9. Ablation study about the preprocessing effect on the DME grading.

Model with Preprocessing	Accuracy (%)	AUC
ConvUNeXt-Original only	80.98	0.9334
ConvUNeXt-CE only	78.86	0.8836
ConvUNeXt-CIELAB only	88.12	0.9069
Majority Voting	91.12	0.9334

In the Macula Localization Module, several preprocessing techniques were applied sequentially to the original fundus images for accurate detection of the macula area. This sequence of preprocesses is composed of IRHSF, histogram specification, CLAHE, and Otsu binarization in this order. It is not possible to eliminate one of these processes from the sequence because the output of the previous process is the input of the next process. Table 10 shows the percentage of Messidor images whose macular area can be detected with/without the sequence of preprocesses. From Table 10, we can conclude that the sequence of preprocesses is indispensable for macula area detection. If this sequence is eliminated, only 43.36% of Messidor images can be evaluated in the DME grading module.

Table 10. Ablation study about the impact of the sequence of preprocesses on macula area detection.

	With the Sequence	Without the Sequence
Percentage of images	100%	43.46%

5. Discussion

As mentioned before, the main objective of the proposed system is to provide an interpretable DME grading system that can be used in any local region, including low-resource areas and in telemedicine frameworks. Considering this context, it is desirable for the proposed system to work on a mobile device or any common personal computer. The time elapsed from introducing an input retinal image to obtaining the DME grade is measured on a personal computer with an Intel® i7 processor and 16 GB RAM, obtaining

approximately 3.8 s per input image. This processing time may be short enough, but if multiple images are required per patient, the processing time must be reduced.

Although the proposed system employs several image enhancement techniques to mitigate the impact of low-quality images on the accuracy of the grading, the relatively inexpensive ophthalmoscopies used in low-resource areas typically provide low-resolution images that may lead to misclassification of DME grading. Considering that the quality of some retinal images can be too low to diagnose, these low-quality images must be detected before their introduction to the proposed system because adapting the proposed algorithm to any low-quality retinal images is a very challenging task.

Recently, optical coherence tomography (OCT) has been used to detect various eye diseases such as DME, macular holes, age-related macular degeneration, and so on. Applying deep learning techniques to images generated by spectral-domain OCT (SD-OCT) can result in a high DME detection rate of over 99% [37]. Undoubtedly, the combination of OCT technology and deep learning will become the main methodology for diagnosing DME. However, at present, OCT equipment is still expensive for any local hospital in developing countries compared to traditional ophthalmoscopies, making it difficult to use OCT.

In segmentation tasks, constructing accurate GT masks is crucial during both the training and testing phases. As mentioned before, the shapes of the GT masks for Messidor and Retinal Lesions are very similar, allowing for proper training and testing. To evaluate possible bias produced by the private GT masks, we obtained Mean IoU using test images, without enhancement, from only the Retinal Lesions dataset. The obtained mean IoU is 66.4, which is 3 points lower than the mean IoU obtained from all test images (69.5) as shown in Table 4. In this sense, there is a small bias due to private GT masks or different types of retinal images used in Retinal Lesions.

6. Conclusions

The proposed system offers a comprehensive approach to diabetic macular edema (DME) grading, addressing the critical need for early detection and classification of this condition. With diabetes becoming increasingly prevalent worldwide, the risk of DME-related vision loss emphasizes the urgency for accurate diagnostic tools. Leveraging advances in medical imaging and deep learning, our system integrates multiple modules to provide interpretable and clinically relevant DME grading.

In the first module, our system employs ConvUNeXt to segment hard exudates (HaEx), a key indicator of DME, from retinal fundus images. Through rigorous experimentation and dataset comparisons, we demonstrate the effectiveness of our segmentation approach across diverse datasets, ensuring robust performance under varying conditions. After that, our system proceeds to localize the optic disc (OD), a crucial reference point for subsequent macula localization. Utilizing RetinaNet, we achieve precise OD localization, laying the groundwork for accurate macula area calculations. The macula localization module refines our system's capabilities by employing preprocessing techniques to enhance macula visibility and optimize image representations to facilitate accurate macula localization, which is crucial for DME severity assessment.

In the final stage, our system consolidates the outcomes from all preceding modules to assign DME grades, providing practical insights for ophthalmologists. By integrating segmentation results, OD coordinates, and macula localization information, our system offers interpretable grading outcomes, facilitating informed clinical decision-making.

Experimental results demonstrate the effectiveness of our system across multiple performance metrics, with competitive accuracy, sensitivity, and specificity compared to existing methods. While non-interpretable methods often excel in certain metrics, our system maintains the balance between performance and interpretability, offering valuable insights into the decision-making process.

As a future work, we will consider adding retinal image quality assessment into the proposed algorithm to discard some low-quality images that make it difficult to perform accurate DME grading.

Author Contributions: Conceptualization, Z.G.-N. and A.F.-R.; methodology, M.N.; software, Z.G.-N.; validation, A.F.-R. and M.N.; formal analysis, M.N.; investigation, Z.G.-N. and A.F.-R.; resources, M.N.; data curation, Z.G.-N.; writing—original draft preparation, Z.G.-N.; writing—review and editing, A.F.-R. and M.N.; visualization, Z.G.-N.; supervision, A.F.-R. and M.N.; project administration, M.N.; funding acquisition, M.N. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable. All images used belong to public databases.

Data Availability Statement: The image databases used for this research are public databases, which are available in: Messidor Dataset: https://www.adcis.net/en/third-party/messidor2/ (accessed on 22 June 2024), Retinal Lesions Dataset: https://www.kaggle.com/c/diabetic-retinopathy-detection (accessed on 22 June 2024), DDR Dataset: https://github.com/nkicsl/DDR-dataset (accessed on 22 June 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

In this section, we provide mathematical development of IRHSF preprocessing and histogram specification, which are used in the "Macula Localization module".

IRHSF preprocessing is used to reduce the illumination variation of an image (intra-image variation). It considers an image as a product of two components: illumination I_s which can be regarded as the background, and I_r, the fundus image without illumination, as shown in (A1).

$$I(x, y, \lambda) = I_s(x, y, \lambda) \times I_r(x, y, \lambda) \tag{A1}$$

where (x, y) are the image coordinates and λ is the light wavelength, which is determined for each color channel, R, G, and B. Applying the natural logarithm to (A1) we have $I_L(x, y, \lambda) = \log(I_s(x, y, \lambda)) + \log(I_r(x, y, \lambda))$. According to [38] the value of the coordinates (x, y) of I_L can be calculated with a polynomial function of order 4 with respect to the coordinates of I_L.

$$I_L = SP \tag{A2}$$

where $I_L = \begin{bmatrix} I_L(0,0) \\ I_L(0,1) \\ \vdots \\ I_L(x,y) \\ \vdots \\ I_L(N,M) \end{bmatrix}$ is the logarithm of the image I with size NxM ordered in vector

form by the coordinate, S is a matrix that represents (x, y) variable of the 4th-order polynomial and P is the coefficient vector of the 15 elements of the polynomial, which are as follows:

$$S = \begin{bmatrix} 0^0 0^0 & 0^1 0^0 & \cdots & 0^1 0^3 & 0^0 0^4 \\ 0^0 1^0 & 0^1 1^0 & \cdots & 0^1 1^3 & 0^0 1^4 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ x^0 y^0 & x^1 y^0 & \cdots & x^1 y^3 & x^0 y^4 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ N^0 M^0 & N^1 M^0 & \cdots & N^1 M^3 & N^0 M^4 \end{bmatrix}, P = [p_1, p_2, p_3, \cdots, p_{14}, p_{15}]^T \tag{A3}$$

To obtain P considering the anatomical structures of the retina, such as blood vessels and optic disc, a diagonal matrix W of size, NMxNM with the value 1, whose pixel positions belong to anatomical structures, is considered. Multiplying $S^T W$ in (A2), we obtain $S^T W I_L = S^T W S P$ and solving with respect to the coefficient vector P, we obtain (A4).

$$P = \left(S^T W S\right)^{-1}\left(S^T W I_L\right) \tag{A4}$$

Once P is obtained, the background illumination image without anatomical elements $I_s(x, y, \lambda)$ and the retina image without considering background illumination variation $I_r(x, y, \lambda)$ can be obtained in (A5).

$$\begin{aligned} I_s(x, y, \lambda) &= \exp(SP) \\ I_r(x, y, \lambda) &= \exp(I_L(x, y, \lambda) - SP) \end{aligned} \tag{A5}$$

The histogram specification process can be described as shown in (A6).

$$M(i) = min\left\{j \middle| G_{target}(j) \geq G_{input}(i)\right\} \tag{A6}$$

where $G_{input}(i)$ represents the Cumulative Distribution Function (CDF) calculated using the input image histogram, $G_{target}(j)$ is the CDF calculated using the target image histogram, and $M(i)$ is the mapping function that maps intensity levels in the input image to intensity levels in the target histogram. This is typically performed by matching the CDF values of the input image to the closest CDF values of the target histogram.

References

1. Shaban, M.; Ogur, Z.; Mahmoud, A.; Switala, A.; Shalaby, A.; Khalifeh, H.A.; Ghazal, M.; Fraiwan, L.; Giridharan, G.; Sandhu, H. A Convolutional Neural Network for the Screening and Staging of Diabetic Retinopathy. *PLoS ONE* **2020**, *15*, e0233514. [CrossRef] [PubMed]
2. Mathews, M.R.; Anzar, S.M. A Comprehensive Review on Automated Systems for Severity Grading of Diabetic Retinopathy and Macular Edema. *Int. J. Imaging Syst. Technol.* **2021**, *31*, 2093–2122. [CrossRef]
3. Yasashvini, R.; Vergin Raja Sarobin, M.; Panjanathan, R.; Graceline Jasmine, S.; Jani Anbarasi, L. Diabetic Retinopathy Classification Using CNN and Hybrid Deep Convolutional Neural Networks. *Symmetry* **2022**, *14*, 1932. [CrossRef]
4. Mookiah, M.R.K.; Acharya, U.R.; Chua, C.K.; Lim, C.M.; Ng, E.Y.K.; Laude, A. Computer-Aided Diagnosis of Diabetic Retinopathy: A Review. *Comput. Biol. Med.* **2013**, *43*, 2136–2155. [CrossRef] [PubMed]
5. Fu, Y.; Lu, X.; Zhang, G.; Lu, Q.; Wang, C.; Zhang, D. Automatic Grading of Diabetic Macular Edema Based on End-to-End Network. *Expert. Syst. Appl.* **2023**, *213*, 118835. [CrossRef]
6. Antal, B.; Hajdu, A. An Ensemble-Based System for Automatic Screening of Diabetic Retinopathy. *Knowl. Based Syst.* **2014**, *60*, 20–27. [CrossRef]
7. Gulshan, V.; Peng, L.; Coram, M.; Stumpe, M.; Wu, D.; Narayanaswamy, A.; Venugopàlan, S.; Widner, L.; Madams, T.; Cuadron, J.; et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* **2016**, *316*, 2402–2410. [CrossRef] [PubMed]
8. Sreejini, K.S.; Govindan, V.K. Automatic Grading of Severity of Diabetic Macular Edema Using Color Fundus Images. In Proceedings of the 2013 3rd International Conference on Advances in Computing and Communications, ICACC 2013, Cochin, India, 29–31 August 2013; pp. 177–180.
9. Kunwar, A.; Magotra, S.; Sarathi, M.P. Detection of High-Risk Macular Edema Using Texture Features and Classification Using SVM Classifier. In Proceedings of the 2015 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2015, Kochi, India, 10–13 August 2015; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2015; pp. 2285–2289.
10. Zubair, M.; Ahmad, J.; Alqahtani, F.; Khan, F.; Shah, S.A.; Abbasi, Q.H.; Jan, S.U. Automated Grading of Diabetic Macular Edema Using Color Retinal Photographs. In Proceedings of the 2022 2nd International Conference of Smart Systems and Emerging Technologies, SMARTTECH 2022, Riyadh, Saudi Arabia, 9–11 May 2022; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2022; pp. 1–6.
11. Ren, F.; Cao, P.; Zhao, D.; Wan, C. Diabetic Macular Edema Grading in Retinal Images Using Vector Quantization and Semi-Supervised Learning. *Technol. Health Care* **2018**, *26*, S389–S397. [CrossRef]
12. Al-Bander, B.; Al-Nuaimy, W.; Al-Taee, M.A.; Williams, B.M.; Zheng, Y. *Diabetic Macular Edema Grading Based on Deep Neural Networks*; The University of Iowa: Iowa City, IA, USA, 2017; pp. 121–128.
13. Sahlsten, J.; Jaskari, J.; Kivinen, J.; Turunen, L.; Jaanio, E.; Hietala, K.; Kaski, K. Deep Learning Fundus Image Analysis for Diabetic Retinopathy and Macular Edema Grading. *Sci. Rep.* **2019**, *9*, 10750. [CrossRef]

14. He, X.; Zhou, Y.; Wang, B.; Cui, S.; Shao, L. DME-Net: Diabetic Macular Edema Grading by Auxiliary Task Learning. In Proceedings of the Medical Image Computing and Computer Assisted Intervention (MICCAI), Shenzhen, China, 13–17 October 2019.

15. Wang, Z.; Zhong, Y.; Yao, M.; Ma, Y.; Zhang, W.; Li, C.; Tao, Z.; Jiang, Q.; Yan, B. Automated Segmentation of Macular Edema for the Diagnosis of Ocular Disease Using Deep Learning Method. *Sci. Rep.* **2021**, *11*, 13392. [CrossRef]

16. Wu, J.; Zhang, Q.; Liu, M.; Xiao, Z.; Zhang, F.; Geng, L.; Liu, Y.; Wang, W. Diabetic Macular Edema Grading Based on Improved Faster R-CNN and MD-ResNet. *Signal Image Video Process.* **2021**, *15*, 743–751. [CrossRef]

17. Wang, T.Y.; Chen, Y.H.; Chen, J.T.; Liu, J.T.; Wu, P.Y.; Chang, S.Y.; Lee, Y.W.; Su, K.C.; Chen, C.L. Diabetic Macular Edema Detection Using End-to-End Deep Fusion Model and Anatomical Landmark Visualization on an Edge Computing Device. *Front. Med.* **2022**, *9*, 851644. [CrossRef] [PubMed]

18. Singh, R.K.; Gorantla, R. DMenet: Diabetic Macular Edema Diagnosis Using Hierarchical Ensemble of CNNs. *PLoS ONE* **2020**, *15*, e0220677. [CrossRef]

19. Li, F.; Wang, Y.; Xu, T.; Dong, L.; Yan, L.; Jiang, M.; Zhang, X.; Jiang, H.; Wu, Z.; Zou, H. Deep Learning-Based Automated Detection for Diabetic Retinopathy and Diabetic Macular Oedema in Retinal Fundus Photographs. *Eye* **2022**, *36*, 1433–1441. [CrossRef]

20. Yao, Z.; Yuan, Y.; Shi, Z.; Mao, W.; Zhu, G.; Zhang, G.; Wang, Z. FunSwin: A Deep Learning Method to Analysis Diabetic Retinopathy Grade and Macular Edema Risk Based on Fundus Images. *Front. Physiol.* **2022**, *13*, 961386. [CrossRef] [PubMed]

21. Holzinger, A.; Langs, G.; Denk, H.; Zatloukal, K.; Müller, H. Causability and Explainability of Artificial Intelligence in Medicine. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2019**, *9*, e1312. [CrossRef] [PubMed]

22. Singh, A.; Jothi Balaji, J.; Rasheed, M.A.; Jayakumar, V.; Raman, R.; Lakshminarayanan, V. Evaluation of Explainable Deep Learning Methods for Ophthalmic Diagnosis. *Clin. Ophthalmol.* **2021**, *15*, 2573–2581. [CrossRef] [PubMed]

23. Yeonwoo, J.; Yu-Jin, H.; Jae-Ho, H. Review of Machine Learning Applications using Retinal Fundus Images. *Diagnostics* **2022**, *12*, 134. [CrossRef] [PubMed]

24. Sebastian, A.; Elharrouss, O.; Al-Maadeed, S.; Al-Maadeed, N. A survey on Deep-Learning-Based Diabetic Retinopathy Classification. *Diagnostics* **2023**, *13*, 345. [CrossRef]

25. Alawad, M.; Aljouie, A.; Alamri, S.; Alghamdi, M.; Alabdulkader, B.; Alkanhal, N.; Almazroa, A. Machine Learning and Deep Learning Techniques for Optic Disc and Cup Segmentation—A Review. *Clin. Ophthalmol.* **2022**, *16*, 747–764. [CrossRef]

26. Jihyoung, R.; Mobeen, R.; Imran, N.; Kil, C. SegR-Net: A Deep Learning Framework with Multi-Scale Feature Fusion for Robust Retinal Vessel Segmentation. *Comput. Biol. Med.* **2023**, *163*, 107132. [CrossRef]

27. Wan, C.; Zhou, X.; You, Q.; Sun, J.; Shen, J.; Zhu, S.; Jiang, Q.; Yang, W. Retinal Image Enhancement using Cycle-Constraint Adversarial Network. *Front. Med.* **2022**, *8*, 793726. [CrossRef] [PubMed]

28. Decencière, E.; Zhang, X.; Cazuguel, G.; Lay, B.; Cochener, B.; Trone, C.; Gain, P.; Ordóñez-Varela, J.-R.; Massin, P.; Erginay, A. Feedback on a Publicly Distributed Image Database: The Messidor Database. *Image Anal. Stereol.* **2014**, *33*, 231–234. [CrossRef]

29. Wei, Q.; Li, X.; Yu, W.; Zhang, X.; Zhang, Y.; Hu, B.; Mo, B.; Gong, D.; Chen, N.; Ding, D.; et al. Learn to Segment Retinal Lesions and Beyond. In Proceedings of the 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021.

30. Li, T.; Gao, Y.; Wang, K.; Guo, S.; Liu, H.; Kang, H. Diagnostic Assessment of Deep Learning Algorithms for Diabetic Retinopathy Screening. *Inf. Sci.* **2019**, *501*, 511–522. [CrossRef]

31. Shen, Y.; Wang, H.; Fang, J.; Liu, K.; Xu, X. Novel Insights into the Mechanisms of Hard Exudate in Diabetic Retinopathy: Findings of Serum Lipidomic and Metabolomics Profiling. *Heliyon* **2023**, *9*, e15123. [CrossRef] [PubMed]

32. Han, Z.; Jian, M.; Wang, G.-G. Convunext: An Efficient Convolution Neural Network for Medical Image Segmentation. *Knowl.-Based Syst.* **2022**, *253*, 109512. [CrossRef]

33. Yutong, X.; Bing, Y.; Qingbiao, G.; Jianpeng, Z.; Qi, W.; Yong, X. Attention Mechanisms in Medical Image Segmentation: A Survey. *arXiv* **2023**, arXiv:2305.17937.

34. Khojasteh, P.; Aliahmad, B.; Kumar, D. Fundus Images Analysis Using Deep Features for Detection of Exudates, Hemorrhages and Microaneurysms. *BMC Ophthalmol.* **2018**, *18*, 288. [CrossRef]

35. Staal, J.; Abràmoff, M.D.; Niemeijer, M.; Viergever, M.A.; van Ginneken, B. Ridge-based Vessel Segmentation in Color Images of the Retina. *IEEE Trans. Med. Imaging* **2004**, *23*, 501–509. [CrossRef]

36. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [CrossRef]

37. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In *Computer Vision—ECCV*; Springer: Cham, Switzerland, 2014; pp. 740–755.

38. Narasimha-Iyer, H.; Can, A.; Roysam, B.; Stewart, V.; Tanenbaum, H.L.; Majerovics, A.; Singh, H. Robust Detection and Classification of Longitudinal Changes in Color Retinal Fundus Images for Monitoring Diabetic Retinopathy. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 1084–1098. [CrossRef]

39. Calderon-Auza, G.; Carrillo-Gomez, C.; Nakano, M.; Toscano-Medina, K.; Perez-Meana, H.; Leon, A.G.-H.; Quiroz-Mercado, H. A Teleophthalmology Support System Based on the Visibility of Retinal Elements Using the Cnns. *Sensors* **2020**, *20*, 2838. [CrossRef]

40. Basit, A.; Fraz, M.M. Optic Disc Detection and Boundary Extraction in Retinal Images. *Appl. Opt.* **2015**, *54*, 3440–3447. [CrossRef]

41. Ali, H.M.; El Abbadi, N.K. Optic Disc Localization in Retinal Fundus Images Based on You Only Look Once Network (YOLO). *Int. J. Intell. Eng. Syst.* **2023**, *16*, 332–342. [CrossRef]

42. Gegúndez-Arias, M.E.; Marin, D.; Bravo, J.M.; Suero, A. Locating the Fovea Center Position in Digital Fundus Images Using Thresholding and Feature Extraction Techniques. *Comput. Med. Imaging Graph.* **2013**, *37*, 386–393. [CrossRef]
43. Aquino, A. Establishing the Macular Grading Grid by Means of Fovea Centre Detection Using Anatomical-based and Visual-based Features. *Comput. Biol. Med.* **2014**, *55*, 61–73. [CrossRef]
44. Molina-Casado, J.; Carmona, E.; García-Feijoó, J. Fast Detection of the Main Anatomical Structures in Digital Retinal Images Based on Intra- and Inter-structure Relational Knowledge. *Comput. Methods Programs Biomed.* **2017**, *149*, 55–68. [CrossRef]