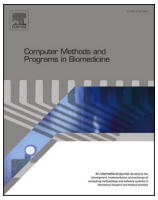
Computer Methods and Programs in Biomedicine 249 (2024) 108160

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine



# Attention-based deep learning framework for automatic fundus image processing to aid in diabetic retinopathy grading

Roberto Romero-Oraá [a,b,1,*], María Herrero-Tudela [a], María I. López [a,b], Roberto Hornero [a,b], María García [a,b]

[a] Biomedical Engineering Group, University of Valladolid, Valladolid, 47011, Spain
[b] Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Spain

ARTICLE INFO

*Keywords:*
Diabetic retinopathy grading
Fundus images
Deep learning
Attention mechanism
Explainable artificial intelligence

ABSTRACT

*Background and objective:* Early detection and grading of Diabetic Retinopathy (DR) is essential to determine an adequate treatment and prevent severe vision loss. However, the manual analysis of fundus images is time consuming and DR screening programs are challenged by the availability of human graders. Current automatic approaches for DR grading attempt the joint detection of all signs at the same time. However, the classification can be optimized if red lesions and bright lesions are independently processed since the task gets divided and simplified. Furthermore, clinicians would greatly benefit from explainable artificial intelligence (XAI) to support the automatic model predictions, especially when the type of lesion is specified. As a novelty, we propose an end-to-end deep learning framework for automatic DR grading (5 severity degrees) based on separating the attention of the dark structures from the bright structures of the retina. As the main contribution, this approach allowed us to generate independent interpretable attention maps for red lesions, such as microaneurysms and hemorrhages, and bright lesions, such as hard exudates, while using image-level labels only.
*Methods:* Our approach is based on a novel attention mechanism which focuses separately on the dark and the bright structures of the retina by performing a previous image decomposition. This mechanism can be seen as a XAI approach which generates independent attention maps for red lesions and bright lesions. The framework includes an image quality assessment stage and deep learning-related techniques, such as data augmentation, transfer learning and fine-tuning. We used the architecture Xception as a feature extractor and the focal loss function to deal with data imbalance.
*Results:* The Kaggle DR detection dataset was used for method development and validation. The proposed approach achieved 83.7 % accuracy and a Quadratic Weighted Kappa of 0.78 to classify DR among 5 severity degrees, which outperforms several state-of-the-art approaches. Nevertheless, the main result of this work is the generated attention maps, which reveal the pathological regions on the image distinguishing the red lesions and the bright lesions. These maps provide explainability to the model predictions.
*Conclusions:* Our results suggest that our framework is effective to automatically grade DR. The separate attention approach has proven useful for optimizing the classification. On top of that, the obtained attention maps facilitate visual interpretation for clinicians. Therefore, the proposed method could be a diagnostic aid for the early detection and grading of DR.

## 1. Introduction

### 1.1. Background

Diabetic retinopathy (DR) is a primary cause of blindness and vision loss globally [1]. There is enough scientific evidence that most visual loss can be prevented through early detection and adequate treatment [1,2]. However, this condition is initially asymptomatic and usually remains undetected until an advanced vision-threatening stage [3]. Therefore, regular DR screening programs aimed at early diagnosis,

* Corresponding author at: Biomedical Engineering Group, E.T.S. Ingenieros de Telecomunicación, Universidad de Valladolid, Campus Miguel Delibes, Paseo Belén 15, 47011 - Valladolid, Spain.
*E-mail address:* roberto.romero@uva.es (R. Romero-Oraá).
[1] URL: www.gib.tel.uva.es

https://doi.org/10.1016/j.cmpb.2024.108160
Received 6 June 2023; Received in revised form 26 January 2024; Accepted 30 March 2024
Available online 3 April 2024

surveillance and timely treatment of DR are required [1]. In these programs, DR detection and DR severity grading are performed by trained specialists through visual inspection of fundus images [2]. The main problem is that, in practice, the manual analysis of these images involves a work overload due to the increasing prevalence of the DR and the limited resources in personnel and technology [4]. Additionally, the level of agreement among specialists when it comes to DR grading tends to be moderate (around 11 % discrepancy), which implies a certain subjectivity related to diagnosis [5]. In this context, computer-assisted diagnostic (CAD) systems may play an important role in analyzing fundus images to assist ophthalmologists [3]. This way, DR diagnosis can be improved in terms of accuracy, speed and confidence, while reducing the workload of specialists and the health costs [6]. Thus, the limited number of human graders could care for a larger number of patients.

The analysis of fundus images is aimed at the detection of retinal lesions, and the severity of the DR can be graded based on the type and the quantity of the detected lesions [7]. According to the International Clinical DR Scale, DR severity can be classified in 5 stages [7]:

- Stage 0. No apparent Retinopathy: it presents no visible signs of abnormalities.
- Stage 1. Mild Non-Proliferative Diabetic Retinopathy (NPDR): it is characterized by the presence of microaneurysms (MAs) only, which are the first sign of DR.
- Stage 2. Moderate NPDR: it shows more than just MAs but less than severe NPDR.
- Stage 3. Severe NPDR: it is considered when there are either more than 20 intraretinal hemorrhages (HEs), venous beading or prominent intraretinal microvascular abnormalities and no signs of proliferative DR.
- Stage 4. Proliferative DR: it presents either or both of neovascularization vitreous or pre-retinal HE.

Taking this scale into account, referable DR is defined as moderate or worse DR or referable diabetic macular edema. Conversely, nonreferable DR is defined as no retinopathy or mild NPDR and no apparent macular edema [8].

Although the clinical signs are well defined for each stage, in practice, DR severity grading is performed as a global estimation based on the type and extension of the overall retinal lesions [2]. Thus, the individual detection of retinal lesions is not as relevant as the severity degree in the full eye. In the end, the decision about the treatment is based on this graduation and, consequently, DR grading is the ultimate goal for CADSs [2].

### 1.2. Related work

In the literature, several approaches to automatically detect the DR can be found [4,6,9]. Traditional methods were based on manually designed features [9]. In recent years, deep learning (DL) methods have achieved better performance and have become the preferred solution for many automatic classification tasks, including DR grading [9]. Unlike traditional methods, DL models allow for automatically optimizing the features in an end-to-end manner [9]. For these reasons, the number of publications related to DR and DL has increased dramatically in the last years [4,9]. However, DR grading is a complex challenge and, consequently, most DR research has generally focused on the binary classification of referable DR. For instance, Gulshan et al. [10] developed a DL algorithm for automated detection of referable DR based on an ensemble of 10 convolutional neural networks (CNNs) with Inception-v3 architecture [11]. The method was trained using 128,175 images and was validated over two datasets. Similarly, other authors proposed an ensemble of two CNNs to detect referable DR while detecting the pathological pixels [12]. In the work of Abràmoff et al. [13], a hybrid system was proposed: several DL models were used to extract DR related features to be integrated into a classic system. Other authors [14] applied a pre-trained fully convolutional network (FCN) to build a weakly-supervised model. They were able to produce patch-level predictions for DR lesions and to detect DR while using image labels only. In the literature, we can also find a few studies aimed at graduating the DR in several stages. González-Gonzalo et al. [15] proposed an iterative approach that obtains a refined localization of abnormalities using a VGG-16 network architecture. In [5], 1665,151 images were used to train the same CNN architecture exposed in [10]. The method included a cascade of thresholds on the output probabilities to obtain the final classification. Concurrently, De la Torre et al. [16] introduced a weighted kappa loss function for multi-class classification proving effectiveness in DR grading. Later, the same authors proposed an interpretable classifier based on a novel pixel-wise score propagation model [17]. Recently, Araujo et al. [2] published a method that provides, together with the DR grading, a medically interpretable explanation and an uncertainty estimation. A novel Gaussian-sampling approach based on a Multiple Instance Learning framework was proposed. All mentioned approaches were built upon a certain CNN architecture that extracted a set of features over the whole image. However, some parts of the image are more relevant than others when determining DR severity. For this reason, attention mechanisms enter the picture allowing learning to focus on the areas of the image most useful for classification and ignoring the less relevant areas [18,19]. Based on this mechanism, Wang et al. [20] proposed a CNN algorithm that mimicked the zoom-in process of a clinician to examine the retinal images. This method could generate attention maps for suspicious regions and predict the DR severity degree based on both the whole image and its high-resolution suspicious patches. However, the approach required the images of both eyes, which is not always possible. Other authors proposed a category attention block for imbalanced data distributions and a global attention block to capture more detailed small lesion information [21]. In the same context, an ensemble of various deep architectures were used together with a convolutional block attention module, which combines spatial and channel attention [22]. Lin et al. [23] proposed a lesion information-based Attention Fusion Network (AFN) that learns weights by reducing the interference of noise on classification for different DR severity degrees. Although their model performed well for referral DR detection, it failed for DR grading. In [24] the Attention-Driven Cascaded Network (ADCNet) was published. Both low-level and high-level features were transmitted through a cascaded architecture and a novel attention module helped capture lesion-aware features without any manual lesion annotations. Finally, Bhati et al. [25] proposed a dual attention network based on Bi-directional Spatial Attention (BSA) module and a Channel-wise Parallel Attention (CPA) module in parallel to capture lesion-specific features from the input fundus image. They managed to capture the small lesions more efficiently than previous methods. Table 1 summarizes the related work for a better understanding.

All the studies mentioned above applied their corresponding CNN architecture to jointly detect any sign of DR at the same time. We hypothesize that this joint detection makes the classification task more difficult and that classifiers are harder to optimize than when a separate detection of the bright and the dark lesions is attempted. Additionally, most previous approaches lack explainability in predictions, which generates mistrust in the medical environment [2]. In this regard, an approach based on attention maps that reveal the pathological regions in images could benefit clinicians and support the prediction of the automatic model.

### 1.3. Contributions

In this paper, we propose a novel CNN architecture based on an innovative attention mechanism for DR grading. The approach is based on two hypotheses. For the first one, we noticed that the independent detection of dark lesions and bright lesions in fundus images used to be

**Table 1**
Related work summary.

| Method | Description | Performance |
|---|---|---|
| Gulshan et al. 2016 | • Binary referable classification only<br>• Training/test: 128,175/9963 images<br>• Ensemble CNN | AUC: 0.991 |
| Abràmoff et al. 2016 | • Binary referable classification only<br>• 1748 images<br>• Hybrid architecture | AUC: 0.980 |
| Quellec et al. 2017 | • Binary referable classification only<br>• Training/test: 28,100/53,576 images<br>• Ensemble CNN | AUC: 0.955 |
| Wang et al. 2017 | • Training/test: 35,126/42,670 images<br>• Attention mechanism | QWK: 0.85 |
| González-Gonzalo et al. 2018 | • Training/test: 28,098/7028 images<br>• Iterative saliency map refinement | QWK: 0.72 |
| De la Torre et al. 2018 | • Training/test: 35,126/53,576 images<br>• Weighted kappa loss function | QWK: 0.72 |
| Krause et al. 2018 | • Training/test: 1665,151/1818 images<br>• Ensemble CNN | Acc: 85.49<br>QWK: 0.84 |
| Lin et al. 2018 | • Training/test: 35,126/42,670 images<br>• Anti-noise Detection and Attention-Based Fusion | QWK: 0.86 |
| Costa et al. 2019 | • Binary referable classification only<br>• Training/test: 768/240 images<br>• Weak supervision | AUC: 0.971 |
| De la Torre et al. 2020 | • Training/test: 75,650/10,000 images<br>• Pixel-wise score propagation model | QWK: 0.80 |
| Araújo et al. 2020 | • Training/test: 35,126/53,576 images<br>• Uncertainty awareness | Acc: 72.36<br>QWK: 0.74 |
| He et al. 2021 | • Training/test: 35,126/53,576 images<br>• Category Attention Block | Acc: 86.18<br>QWK: 0.87 |
| Nirthika et al. 2022 | • Training/test: 35,126/42,670 images<br>• Siamese CNN | Acc: 84.55<br>QWK: 0.86 |
| Yue et al. 2023 | • Training/test: 15,453/3863 images<br>• Cascade architecture and lesion-aware attention | Acc: 72.44<br>QWK: 0.54 |
| Bhati et al. 2024 | • Training/test: 35,126/42,670 images<br>• Parallel attention for discriminative features | QWK: 0.89 |

common before severity grading in traditional methods, where the inspiration of this method comes from [26]. Red lesions and exudates, as main indicators of DR, are clearly distinguishable between them in terms of visual appearance. Therefore, the problem can be divided into two parallel phases: the detection of red lesions and the detection of bright lesions. This approach would allow the classification to be individually maximized in each phase and the result subsequently combined. Hence, an independent classification task for each group of lesions would simplify and optimize the detection. However, lesion-level annotations are hard to obtain and not usually available. Moreover, the severity grading does not require an accurate segmentation of the lesions. Consequently, we believe that the automatic DR grading can be improved by separating the processing of red lesions and bright lesions without the need for lesion-level annotations. As for the second hypothesis, clinicians would greatly benefit from visual explanations to support the automatic model predictions. These visual explanations could be more useful whether separate information is provided for the two main groups of lesions. In order to combine these ideas, we propose a novel DL approach based on separating the attention of the dark structures from the bright structures of the retina, which is the main contribution of this work. This approach allowed us to generate interpretable attention maps, separating red lesions and bright lesions, while exclusively using image-level annotations (corresponding to the DR severity degree). Never before were these lesions detected separately in a DL-based method by means of weak supervision. The generated attention maps provide visual explanations for the predicted DR degree, thus contributing to the field of explainable artificial intelligence.

The proposed approach consists of an end-to-end DL framework which accepts a single fundus image as input. This way, it could work with different capture protocols or even when the image of one of the eyes is not available. The framework includes an image quality assessment stage and additional DL related techniques such as data augmentation, transfer learning and fine-tuning. This set of techniques make up a system capable of analyzing any fundus image with a great generalizability. The method proposed in this research could assist in the diagnosis of DR, alleviating the burden on ophthalmologists and enhancing healthcare for diabetic patients.

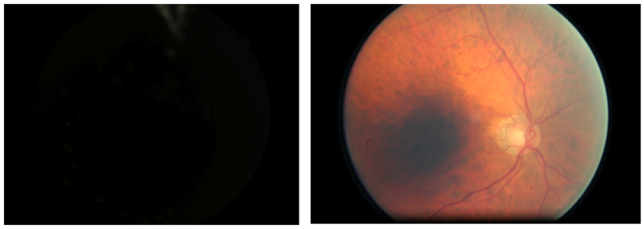
## 2. Materials

In order to perform our experiments, we used the retinal image dataset provided by EyePACS for the Diabetic Retinopathy Detection competition published on Kaggle [27]. This is the largest public database available with 35,126 images meant for training and 53,576 for testing. In this database, a clinician rated the DR severity degree at image-level according to the International Clinical DR Scale [7] in 5 levels: 65,343 images with no DR, 6205 images with Mild NPDR, 13,153 images with Moderate NPDR, 2087 images with Severe NPDR, and 1914 images with proliferative DR. It can be seen that the dataset is highly imbalanced, which has been taken into account in the training procedure. The dataset contains two images per patient (one for each eye). The images in the dataset come from different models and types of cameras, which can affect its visual appearance and the output resolution [27]. Additionally, they may contain artifacts, be out of focus, be underexposed, or be overexposed. This way, this database is greatly representative of real-world scenarios. Both the images and labels may be affected by noise [27]. In fact, we have noticed that some of the images are not suitable for analysis, as shown in Fig. 1. These cases should not be considered for diagnosis and, therefore, were discarded using an image quality assessment system, described in Section 3.2. After discarding the poor-quality retinal images, we randomly selected 10 % of the images from the training set for validation purposes during the development of the method. In this way, we used 18,860 images for training, 2096 images for validation and 32,017 for testing.

## 3. Methods

### 3.1. Overview

For a better understanding of the whole method, the diagram in Fig. 2 is provided. Initially, the input image was evaluated in terms of quality. Only those images with enough quality for analysis underwent the DR grading stage. Then, we applied a minimal preprocessing stage to prepare the image for the CNN architecture. After this and before the deep network, we decomposed the preprocessed image to obtain 2 different versions. In the first version ($I_{bri}$), the bright regions of the fundus image were preserved, while dark regions were removed. Conversely, in the second version ($I_{dark}$), the dark regions in the retina were highlighted, while the bright areas were removed. These two versions of the image, together with the original preprocessed image,



(a) (b)

**Fig. 1.** Some examples of poor-quality fundus images in the dataset. (a) The image is completely black. (b) The area of the macula lacks illumination, which could hide relevant lesions around the fovea.

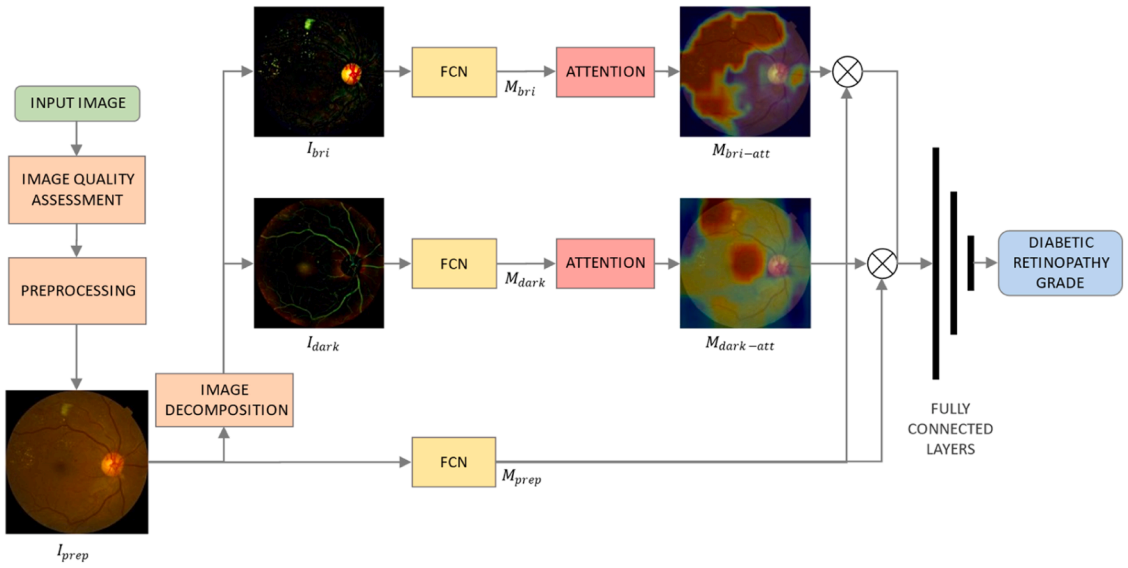



**Fig. 2.** Overview of the proposed method. First, we applied an image quality assessment stage to check if the input image was gradable. Second, a minimal preprocessing step was performed ($I_{prep}$). Then, we decomposed the resulting image to obtain the images $I_{bri}$ and $I_{dark}$. These images, together with $I_{prep}$, were processed using a pre-trained FCN. The resulting feature matrices $M_{bri}$ and $M_{dark}$ underwent an attention mechanism, producing the feature matrices $M_{bri\ -\ att}$ and $M_{dark\ -\ att}$. Next, the extracted features from $I_{prep}$ and $M_{prep}$ were separately multiplied by $M_{bri\ -\ att}$ and $M_{bri\ -\ att}$, and then combined. Finally, a set of 3 fully connected layers performed the final classification.

were the inputs of the CNN architecture. Each of these inputs was processed by a pre-trained FCN, allowing optimal feature extraction. As a result, we obtained the feature matrix $M_{pre}$ from the preprocessed image and the feature matrices $M_{bri}$ or $M_{dark}$ from the bright-object and dark-object image versions, respectively. Next, the matrix $M_{pre}$ was combined, separately, with the matrices $M_{bri}$ and $M_{dark}$ using an attention mechanism. This mechanism selected the relevant elements from the matrices. Finally, the proposed architecture included a set of fully connected layers to classify the image into the different severity degrees based on the complete set of extracted features.

### 3.2. Image quality assessment

The diagnostic capacity of automatic screening systems depends, to a great extent, on the quality of the input image. Unfortunately, not all of the captured fundus images are of sufficient quality for reliable medical analysis [28]. The main factors causing poor quality can be roughly divided into two aspects: inadequate imaging conditions (e.g., insufficient illumination, poor focus or improper operation), and patient related issues (e.g., head or eye movement, pupil dilation, wrong fixation or media opacity) [28]. These problems also appear in the retinal image dataset used in this work (see Fig. 1), since images are representative of a real clinical setting [27]. Therefore, some images are unsuitable for DR grading and need to be previously discarded. For this task, we applied the automatic method presented in [29]. The approach was based on a pretrained CNN model and fine-tuning. After applying this method to the EyePACS database, only the images with sufficient quality for medical analysis were considered for further DR grading. After this stage, 35,729 images were discarded and 52,973 images were used for subsequent analysis. The high number of discarded images is due to the well-known low average quality of the dataset [30–32].

### 3.3. Preprocessing

We applied a minimal preprocessing stage before using the CNN architecture. This way, any fundus image is adapted to our method. Input color images were cropped around the field of view (FOV) by trimming the black external background borders, which makes the retinal diameter equal to the width of the image [2,17]. Then, they were resized to a resolution of 640 × 640 pixels. As proved in [17], the optimal size for DR grading is a retina diameter equal to 640 pixels. Finally, we normalized the pixel values to the interval [−1, 1] for a better training process [33], obtaining the preprocessed image ($I_{prep}$).

### 3.4. Image decomposition to separate the dark and the bright regions

The main purpose of this stage is to separate the dark ($I_{dark}$) and the bright ($I_{bri}$) pixels in the fundus images. These two new images were used as inputs to the CNN. In this way, we were able to separate the attention of dark structures (e.g., red lesions) from bright structures (e.g., hard exudates). In order to perform the image decomposition, we applied the multi-scale algorithm [34]. This approach combines the alternating sequential filtering method and a mean filter in an iterative process for different scales. Fig. 3 shows an example of the images obtained in this stage. In $I_{bri}$, the color difference of the bright pixels with respect to the retinal background was enhanced, while the rest of pixels remained black. In the same way, the color difference between the dark pixels and the background was highlighted in $I_{dark}$, while leaving the rest of the pixels black.

### 3.5. Data augmentation

Deep neural networks are proven to work better when trained with a large amount of data. In order to increase the number of training samples and minimize potential biases, we applied online data augmentation. This technique consists of generating new random, fake samples from the original data to feed the model in every training batch with new images [35]. Accordingly, the number of generated images at the end depends on the number of training epochs. In this work, the fake images were obtained applying the following simple transformations: rotations in the range [−50,+50] degrees, zoom in the scale range [−0.1,+0.1] and horizontal and vertical flips [35].

### 3.6. Feature extraction with transfer learning

In the proposed method, we applied an FCN as a feature extractor for all 3 input images ($I_{prep}$, $I_{bri}$ and $I_{dark}$). The architecture selected for this network was Xception (eXtreme version of Inception) [36]. This

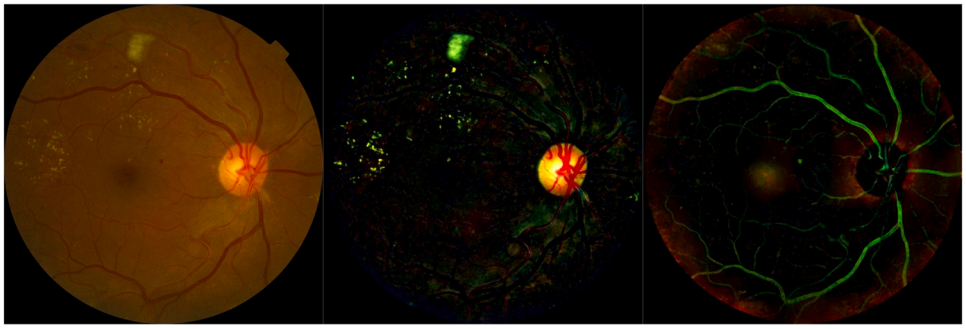



**Fig. 3.** Result of the image decomposition. (a) Original image. (b) Image $I_{bri}$, in which dark pixels were removed and bright regions highlighted. (c) Image $I_{dark}$, in which bright pixels were removed and dark structures of the retina were highlighted.

architecture replaces the Inception modules with depthwise separable convolutions, which reduces computational cost and memory requirements [36]. Additionally, Xception has been proven to outperform other CNN architectures on large image classifications datasets [36].

Training a deep network from scratch could be very slow. Additionally, when the training data is limited, the model could not converge or may achieve poor results. In these situations, the use of transfer learning is a common practice that has proven to work satisfactorily [37]. This technique allows the resolution of a machine learning problem in a particular domain of interest with the knowledge learned from the training data of another domain of interest [37]. In practice, the easiest way to apply transfer learning is the use of pretrained networks. This way, the model is initialized with a set of pretrained weights and then is fine-tuned for the given task. This technique has already been successfully applied for DR grading [14,20]. In this work, we used an FCN pretrained on the images from the project ImageNet [38].

At the end of this stage, the extracted features for $I_{prep}$, $I_{bri}$ and $I_{dark}$ are the matrices $M_{prep}$, $M_{bri}$ and $M_{dark}$, respectively, all of them having dimensions $14 \times 14 \times 1556$. This is the output size of the Xception backbone for the input size of $640 \times 640 \times 3$ pixels.

### 3.7. Attention mechanism

The advantage of using an FCN is that the extracted features preserve the spatial information, which allows interpreting the mentioned matrices as a set of $14 \times 14$ patches with 1556 features each. This way, we can focus on those features associated with the spatial locations of the image (patches) that are more relevant for our classification task. In this case, attention mechanisms produce a weight map which weights the features based on their relevance [19]. In this work, we applied an attention mechanism for the matrices $M_{bri}$ and $M_{dark}$, which were respectively computed from the input images $I_{bri}$ and $I_{dark}$. It should be remembered that the image $I_{bri}$ only contains information about the bright structures of the retina and, in the same way, the image $I_{dark}$ only contains information about the dark structures of the retina. Therefore, this approach allowed us to produce separate attention maps for bright lesions and red lesions. Additionally, theses maps were useful to reduce the impact of unneeded lesion information for identifying different DR grades. This way, the classification task is optimized by focusing on the most relevant areas only.

The architecture of the proposed attention mechanism can be seen in Fig. 4. It takes a feature matrix as input ($14 \times 14 \times 1556$) and outputs an attention map ($14 \times 14 \times 1$). It is composed of a set of four Conv2D layers with kernel ($1 \times 1$), which work as dimension reduction modules [39]. Using this kernel size, each of these modules outputs a matrix with dimensions $14 \times 14 \times N$, where $N$ is the number of filters. The first 3 layers included a ReLU activation function. The last one had a sigmoid activation function, producing an output between 0 and 1 [40], which can be seen as the weight of every patch based on their relevance for the classification task. The closer the value is to 1, the higher the relevance of the patch.

Once both attention maps were computed, they were applied over $M_{prep}$ using the element-wise multiplication on every channel (layer in the last dimension), obtaining the feature matrices $M_{bri\_att}$ and $M_{dark\_att}$. The novel approach of this work allowed us to compute different feature matrices for the bright and the dark elements in the image. Then, these matrices were concatenated on channel dimension (depth direction). Finally, we applied global average pooling to obtain the feature vector that feeds the fully connected layers.

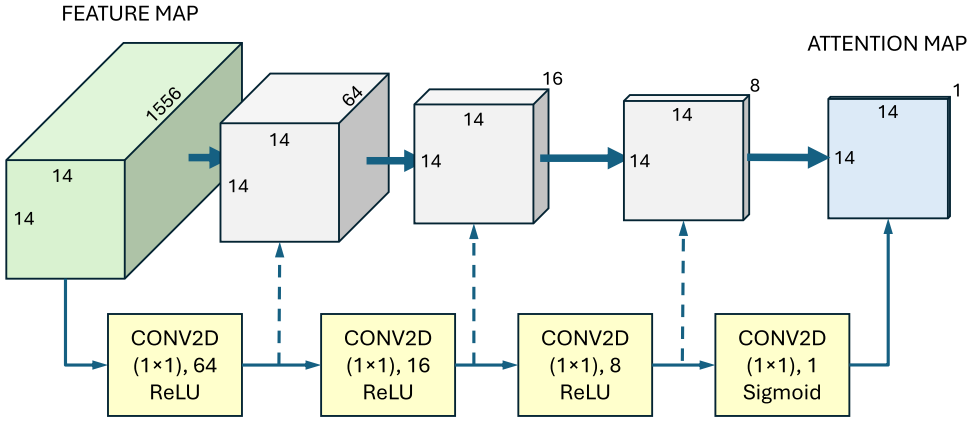



**Fig. 4.** Architecture of the attention mechanism.

### 3.8. Fully-connected layers

The last part of the proposed architecture produced the final DR classification using the combined features obtained with the attention mechanism as inputs. It is composed of 3 fully connected layers. The first 2 layers had 1024 and 512 neurons, respectively, and a ReLU activation function [33]. Additionally, they included a L2 regularization penalty with a factor of 0.005. The last layer had 5 neurons, one for each DR severity degree. The activation function for this layer was softmax, which represents the output probability of every class to be predicted (in our study, each severity degree) [33].

### 3.9. Training procedure

In order to deal with class imbalance, we used the focal loss error function [41]. This approach reshapes the standard cross entropy loss, since it down-weights the loss assigned to well-classified examples [41]. This way, the model focuses on the misclassified examples rather than the ones that it can confidently predict. With data imbalance, the focal loss function focuses on the least represented classes. The focal loss function introduces two hyperparameters. The first one, the focusing parameter γ, controls the strength of the modulating term. The second parameter is the weighting factor $\alpha$ and balances the importance of positive/negative examples. The focal loss function is defined as [41]:

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

where $p_t$ is the estimated probability for the ground truth class. In this work, γ was set to 2 and α was set to 1 after analyzing various experiments using the object detector RetinaNet [41].

The model was trained for 100 epochs using a batch size of 8 images [16]. We used the stochastic gradient descent as the optimization algorithm with a momentum of 0.9 and learning rate of 0.005 [15,20]. To avoid overfitting in advanced epochs, the learning rate was reduced by a factor of 10 when the validation error reached a plateau. Additionally, to deal with the problem of exploding gradients, the L2 norm of the gradient vector was limited to 1 [33].

Experiments were performed on a Workstation with Intel Xeon CPU E5–1620 v4 @ 3.5 GHz × 8, 32GB RAM, 2 × NVIDIA TITAN X (Pascal), using Python 3.6, Keras 2.3 and Tensorflow 2.1.

### 3.10. Ablation studies

In order to evaluate the influence of the proposed attention mechanism on the results, we additionally performed two ablation studies: (1) we built another architecture with no attention mechanism, directly feeding the fully connected layers with the matrix $M_{prep}$; and (2) we modified the attention mechanism such that it was applied to the matrix $M_{prep}$, which includes joint information about the bright and dark pixels. The rest of the hyperparameters were identical in both experiments.

## 4. Results

In this study, we dealt with the DR grading multiclass classification problem. We evaluated the proposed method on the test set of 32,017 images from the EyePACS for the Diabetic Retinopathy Detection dataset [27]. As described in Section 2, the dataset was highly imbalanced. For this reason, the proper evaluation of the performance of the proposed method requires an adequate metric. The Cohen's Kappa [42] is one of the most commonly used statistics to test inter-rate reliability. However, it does not take into account the degree of disagreement. When the categories are ordered, it is preferable to use the Weighted Kappa since it allows disagreements to be weighted differently [43]. This is the case of DR severity degrees, where each class can be seen as an evolution of the previous one. Three matrices are involved in the calculation of the Weighted Kappa: the matrix of observed scores ($O$), the matrix of expected scores based on chance agreement ($E$), and the weight matrix (ω). Each element $O_{i,j}$ is computed by counting the number of samples predicted as the $i$ th class that belong to the $j$-th class. The matrix $E$ is the outer product between the two histogram vectors corresponding to the prediction and true value, normalized such that $E$ and $O$ have the same sum. The matrix ω is the weight penalization. The Weighted Kappa can range from −1 to +1 and is defined as [43]:

$$\kappa = 1 - \frac{\sum_{i,j}\omega_{i,j}O_{i,j}}{\sum_{i,j}\omega_{i,j}E_{i,j}}, \tag{2}$$

where $\omega_{i,j}$, $O_{i,j}$ and $E_{i,j}$ are elements in the weight, observed, and expected matrices, respectively. Quadratic Weighted Kappa (QWK) is the particular case with quadratic weighting [43]:

$$\omega_{i,j} = \frac{(i-j)^2}{(N-1)^2}, \tag{3}$$

where $N$ is the number of classes. This is the most common metric for DR grading [2,5,14,15,20]. For this reason, we computed QWK, allowing us to compare our results with those in other studies. We achieved a QWK=0.78 on the test set, which corresponds to 83.7 % accuracy, as shown in the last row of Table 2. However, when dealing with class imbalance, QWK is dominated by the most representative classes, which is the class 0 (No DR) in our dataset. For this reason, the confusion matrix is also important to evaluate the results. Table 3 shows the confusion matrix obtained with our method.

We conducted two supplementary ablation analyses to assess the impact of the proposed attention mechanism. First, we created an alternative architecture that excludes the attention mechanism and directly supplies the fully connected layers with the $M_{prep}$ matrix. In this experiment, the model failed to converge effectively as a result of the class imbalance. Consequently, every sample was classified as no DR (the most prevalent class) and the detection was incorrect for the rest of the classes. This experiment proved that the attention mechanism is useful to deal with class imbalance. In the second ablation study we applied the attention mechanism directly to the feature matrix $M_{prep}$, which contains combined information regarding both the bright and dark pixels. In this case, we obtained QWK=0.76, lower than the value achieved using the proposed method. This result shows that the separate optimization of bright and dark regions can improve the results of the classification task. Table 2 shows the results of the ablation experiments together with the results of the proposed method.

The computed attention maps are an additional outcome of this work. In Fig. 5, some sample results on images from the test set of the Kaggle DR detection dataset are shown. The original image is shown in the first column (Fig. 5a, d, g), the attention map for red lesions is shown in the second column (Fig. 5b, e, h) and the attention map for bright lesions is shown in the third column (Fig. 5c, f, i). Attention maps are shown overlaid on the original image. Regions with pathological signs are represented with warm colors based on the colormap exposed on the right part of the figure.

## 5. Discussion

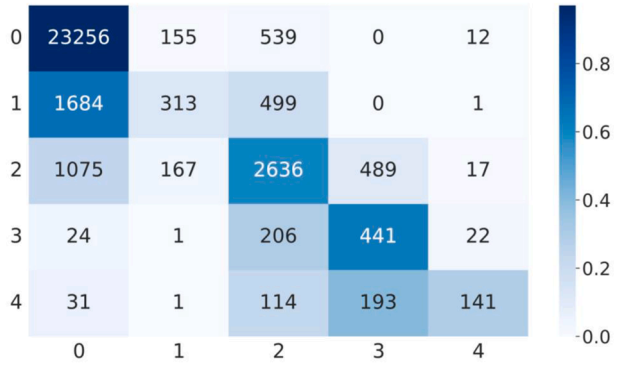
### 5.1. Overview

This work presents a new DL framework for DR grading where the

**Table 2**
Results on the test set of the Kaggle DR detection dataset, including the ablation experiments and the proposed method.

| Method | Accuracy (%) | QWK |
|---|---|---|
| No attention (ablation experiment #1) | 73.79 | 0.42 |
| Joint attention (ablation experiment #2) | 81.93 | 0.76 |
| **Separate attention (proposed method)** | **83.70** | **0.78** |

**Table 3**
Confusion matrix.




proposed attention mechanism is highlighted and was separately applied to the bright and dark pixels of the fundus image. The dataset used in this study contained multiple poor-quality images, which makes them unsuitable for medical analysis. Consequently, an image quality assessment stage was required. For this task, an automatic algorithm was applied, discarding 35,729 out of 88,702 images (40 %).

The number of images discarded by the validated algorithm was very high. On the one hand, this means that the overall quality of the dataset is poor [30]. On the other hand, we can assume that the quality distribution of the dataset is close to real clinical scenarios, which makes it a reliable research material [27]. In this sense, the image quality has also been estimated in previous studies using the EyePACS dataset. In some of these studies, the number of discarded images was also high, reaching 19.50 % [31], 25 % [30] and 40 % [32].

Before using the CNN architecture, an image decomposition process was required to separate the bright and the dark structures of the fundus image. For this task, we applied a multiscale algorithm based on advanced image processing techniques. Despite the great success of DL in computer vision, this study shows that traditional image processing methods could still be very useful. They can provide context for the problem at hand and, when combined with the powerful optimization capacity of deep networks, they can improve results and allow more complex problems to be approached.

### 5.2. Comparison with previous methods

We achieved a QWK of 0.78 on the test set of the Kaggle DR detection dataset. Thus, the proposed method achieves similar QWK to other state-of-the-art methods for DR grading, as shown in Table 4. Results of the studies presented in Table 4 were obtained using the same database. However, the train, validation and test sets could differ among studies and, therefore, the comparisons should be carefully considered. With our approach, we achieved a higher QWK than that obtained by González-Gonzalo et al. [15], De la Torre et al. [16], Araújo et al. [2] and Yue et al. [24]. The later work of De la Torre et al. [17] achieved a higher QWK=0.80 than ours. However, they accomplished the training phase including part of the images meant for testing and, therefore, they tested their method on a reduced subset of 10,000 images. Krause et al. [5] also outperformed our method, reaching QWK=0.84. However, their model was trained on a large, private dataset with more than 1.6 million fundus images. Moreover, their test set was composed of 1818 images, much smaller than ours (32,017). Finally, Wang et al. [20] reported QWK=0.85. However, their method required the annotated labels of the lesions of some images and the fundus images of both eyes, which are not always available. The same drawback stands out for the siamese network proposed in [44], where the images of both eyes were required to capture the patient-level DR features to achieve QWK=0.86. On the contrary, our model exclusively requires image labels from a single eye. Although the method of He et al. achieved QWK=0.87, no interpretable results were provided [21]. Bathi et al., however, provided interpretable activation maps while obtaining the highest QWK=0.89 among the revised studies [25]. Nevertheless, it should be noted that our approach is the only one that provides independent interpretable attention maps for red lesions and bright lesions. Not only does the model provide visual explanations of the predicted DR degree, but it also provides information about the group of lesions that encompass the anomalies found. These visual explanations could greatly assist clinicians in DR grading.

### 5.3. Analysis of results

Analyzing the obtained confusion matrix allows us to extract several conclusions on the results for the proposed method. First, class 0 was detected with high accuracy: only 2.9 % (706 out of 23,962) of those images were over-diagnosed. More importantly, only 0.0005 % (12 out of 23,962) of them were rated class 3 or 4. Conversely, class 1 was easily confused with classes 0 and 2, becoming the most misclassified severity degree. On one hand, this is because some images in this category only present tiny MAs that are hard to detect, as shown in Fig. 6a. On the other hand, several types of red lesions are hardly distinguishable between them, even for clinicians [5]. When class 1 is misclassified as class 0, the implications are not crucial since none of them require medical referral. However, when class 1 is mistaken for class 2, the patient would be unnecessarily referred to the doctor. Regarding class 2, an acceptable detection accuracy of 77.2 % (2636 out of 3414 images) was achieved. However, most of the misclassified images were classified as class 0, which should be taken into account for patient management. Under-diagnosing the moderate NPDR degree could lead to a risky situation for the vision if not detected in upcoming screenings. Images of class 3 were often misclassified as class 2, as the example in Fig. 6b. However, only 7.2 % (50 out of 694) of the severe NPDR images were classified as any of the other degrees. The incorrect detection of the class 3 as class 2 is not crucial since both degrees would involve a manual examination by a specialist. Finally, poor detection accuracy for class 4 was obtained. As shown in the confusion matrix, 40.2 % (193 out of 480) of the class 4 images were diagnosed as class 3. The main reason for this result is that multiple eye fundus images contain photo-coagulation treatment and laser marks, hindering the detection of the characteristic signs of the proliferative DR, such as neovessels and pre-retinal HEs. Fig. 6c exhibits one of these examples. Nevertheless, laser-treated retinas should never reach screening scenarios since they are already under supervision. It should also be noted that the severity degree associated with an image is independent of its quality. For this reason, multiple images with laser marks successfully passed the quality assessment stage.

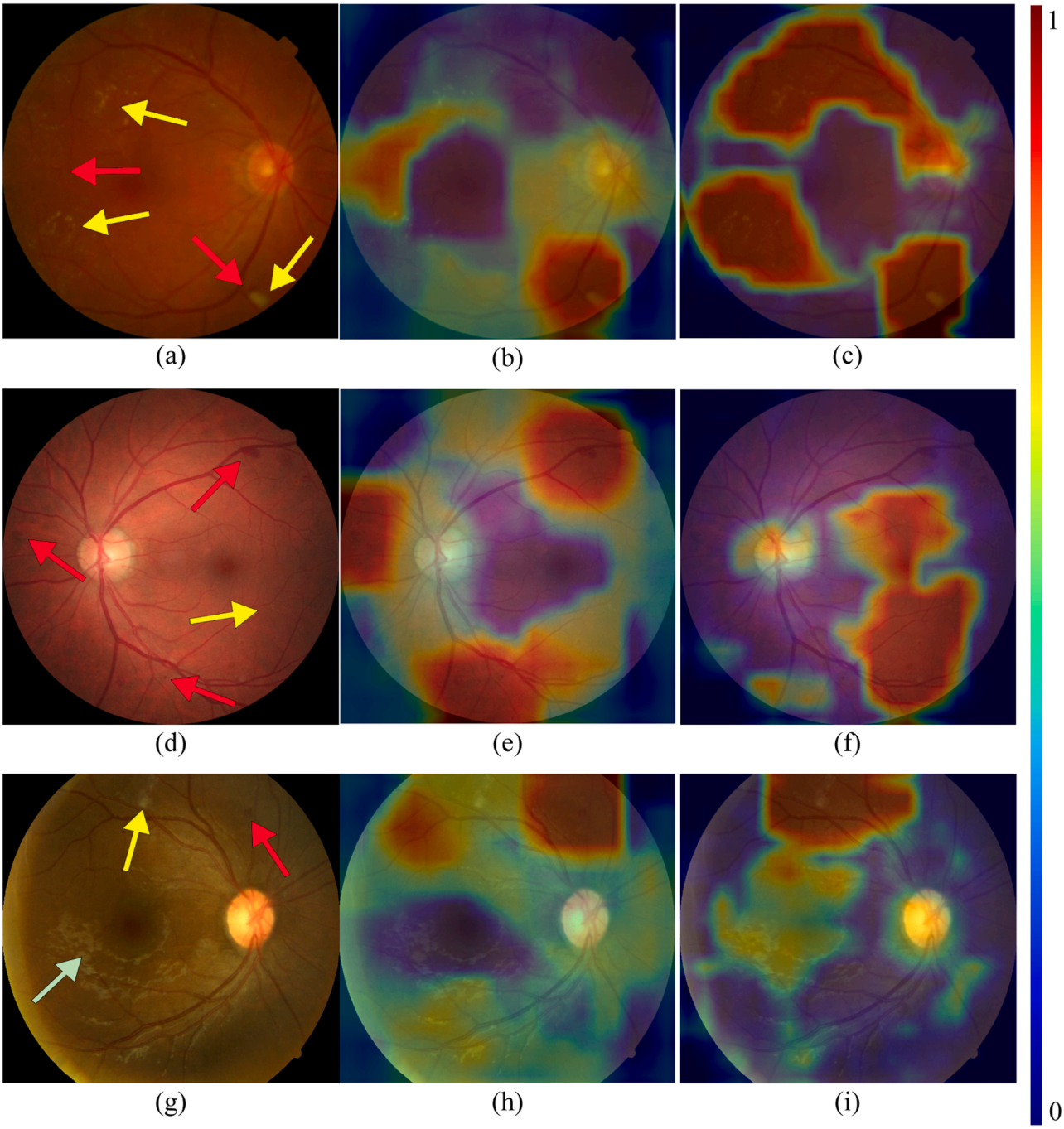



**Fig. 5.** Three sample results on images from the test set of the Kaggle DR detection dataset. First column (a, d, g): original image. Second column (b, e, h): attention map for red lesions. Third column (c, f, i): attention map for bright lesions. The hot regions of the maps represent the areas most likely to be pathological according to the colormap exposed on the right part of the figure.

### *5.4. Attention maps*

Among the outcomes of this work, it is worth analyzing the obtained attention maps, independently generated for red lesions and bright lesions. Never before were these lesions detected separately using image-level annotations. The example in Fig. 5a presents a HE right next to a cotton wool spot on the bottom right of the image and some other HEs (slightly visible) on the left. The attention map in Fig. 5b correctly displays both regions. The same example presents numerous bright lesions which are correctly covered by the attention map in Fig. 5c. The second sample (Fig. 5d) is characterized by a large HE on the top right of the image, a few red lesions to the left of the optic disc, and a clear MA close to a HE located on the bottom center of the image. The attention map in Fig. 5e shows that the mentioned spots can be distinguished. Regarding the bright lesions in this image, the attention map in Fig. 5f seems to highlight various small marks near the macula and below. Finally, the sample in Fig. 5g is interesting since it contains multiple reflective features caused by the nerve fiber layer (mainly concentrated along the widest vessels). These features are very common in retinas from young patients and cannot be considered as abnormalities [26]. On one hand, the attention map in Fig. 5h mainly shows the detection of a HE above the optic disc. On the other hand, the attention map in Fig. 5i highlights the bright lesions on top of the image while successfully showing much less relevance to the mentioned reflective features. In view of this

**Table 4**
Comparison of some methods for DR grading in terms of QWK and accuracy using the Kaggle DR detection dataset.

| Method | Test images | Label usage | Acc. (%) | QWK |
|---|---|---|---|---|
| Wang et al. 2017 | 42,670 | binocular | – | 0.85 |
| González-Gonzalo et al. 2018 | 7028 | monocular | – | 0.72 |
| De la Torre et al. 2018 | 53,576 | monocular | – | 0.72 |
| Krause et al. 2018 | 1818 | monocular | 85.49 | 0.84 |
| De la Torre et al. 2020 | 10,000 | monocular | – | 0.80 |
| Araújo et al. 2020 | 53,576 | monocular | 72.36 | 0.74 |
| He et al. 2021 | 53,576 | monocular | 86.18 | 0.87 |
| Nirthika et al. 2022 | 42,670 | binocular | 84.55 | 0.86 |
| Yue et al. 2023 | 3863 | monocular | 72.44 | 0.54 |
| Bhati et al. 2024 | 42,670 | monocular | – | 0.89 |
| **Proposed method** | 32,017 | monocular | **83.70** | **0.78** |

qualitative analysis, we can state that the generated attention maps are effective in separately detecting red lesions and bright lesions in fundus images. These visual explanations could greatly assist clinicians in DR grading. This way, when the human decision does not match the model prediction, the attention map could be helpful to figure out the reason of the discrepancy. If the model is mistaken, the attention map would show the information that led to error. Conversely, when the human misses out some evidence, the attention map could help rectify the diagnosis.

### 5.5. Limitations and future work

Our study also has some limitations that should be mentioned. The image decomposition algorithm previous to the CNN is time-consuming, requiring approximately 3 s per image. This time would not be a very important problem in a clinical setting, since the time required for image capture is considerably longer. However, it would be desirable to find a faster algorithm for this task. Another limitation is directly related to model performance. As previously mentioned, the proposed approach tends to fail in the classification of classes 1 and 4. Additionally, class 2 tends to be under-diagnosed. This is especially critical when the automatic estimation misclassifies the images as class 0 (no DR), since it would prevent patients with threatened vision from being revised by an ophthalmologist, putting their vision at risk. In future studies, we will try to optimize the detection of referable DR, assuring that no referable case remains undetected. Regarding the obtained attention maps, their low resolution (14×14) limits the precise detection of retinal lesions. Instead, these maps represent a rough estimate of the areas where pathological signs are found. It would be desirable to further delimit these areas by generating higher resolution attention maps. Finally, it would be desirable to train the proposed method with more images and to test it using different databases and capture protocols. Despite the fact that the database used in this study is very representative of the real clinical environment, more examples would be desirable in a deep-learning framework to test the generalization ability of the proposed methodology.

## 6. Conclusions

We propose a novel DL framework for DR grading based on an attention mechanism. Unlike previous methods, our approach performs separate attention for the bright and the dark pixels in the retinal image. On the one hand, this approach allows generating independent attention maps for red and bright lesions, which facilitates visual interpretation from a clinical point of view. On the other hand, dividing the problem makes the model easier to manage and allows to improve model optimization. We used the Kaggle DR detection dataset and achieved results comparable to previously published methods. Our results, including the attention maps and the QWK=0.78, suggest that the proposed method could be a diagnostic aid for the early detection of DR. However, some ethical considerations should be considered. In order to make medical artificial intelligence trustworthy, we must prioritize promoting human health as the primary ethical consideration in its design [45]. Therefore, the proposed framework should be used with screening purposes. The ultimate decision on a patient's treatment must be made by a human grader. As an aiding tool, the automated system would allow diabetic patients to receive better eye care in order to avoid preventable vision loss.

## Authorship responsibility

- The material in this manuscript is original and does not contain any libelous or otherwise unlawful matters.
- The manuscript represents valid work and neither this manuscript nor any other with substantially similar content under my authorship has been published or is being considered for publication elsewhere.
- I have participated sufficiently in the work to take public responsibility for all its content.

## CRediT authorship contribution statement

**Roberto Romero-Oraá:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Writing – original draft. **María Herrero-Tudela:** Conceptualization, Data curation, Writing – review & editing. **María I. López:** Validation, Visualization, Writing – review & editing. **Roberto Hornero:** Conceptualization, Funding acquisition, Project administration, Writing – review & editing. **María García:** Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

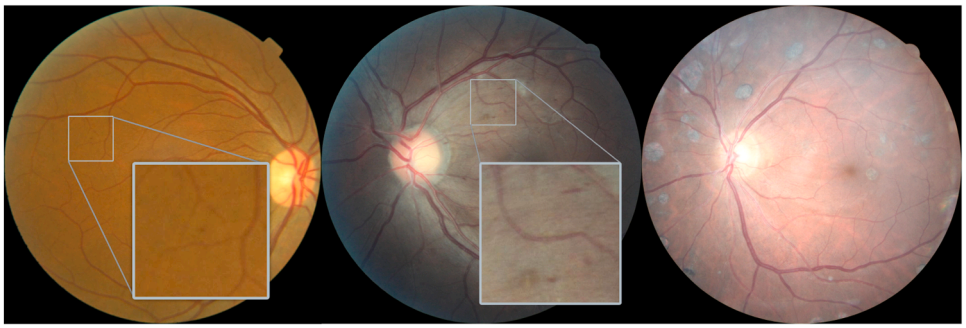



**Fig. 6.** Misclassified examples. (a) Class 1 image with a tiny MA on the left part (see zoomed-in area) and misclassified as class 0. (b) Class 3 image misclassified as class 2 due to the similarity of the lesions (see zoomed-in area). (c) Class 4 image with laser marks and misclassified as R3.

the work reported in this paper.

## Acknowledgements

This research has been developed under the grants TED2021-131913B-I00 and PID2020-115468RB-I00 funded by 'Ministerio de Ciencia e Innovación/Agencia Estatal de Investigación/10.13039/501100011033/', European Regional Development Fund (ERDF) A way of making Europe and European Union NextGenerationEU/PRTR.; and by 'CIBER en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN)' through 'Instituto de Salud Carlos III' co-funded with ERDF funds. M. Herrero Tudela was in receipt of a PIF-UVa grant of the University of Valladolid.

## References

[1] A. Grzybowski, et al., Artificial intelligence for diabetic retinopathy screening: a review, Eye (Basingstoke) 34 (3) (2020) 451–460, https://doi.org/10.1038/s41433-019-0566-0.
[2] T. Araújo, et al., DR|Graduate: uncertainty-aware deep learning-based diabetic retinopathy grading in eye fundus images, Med. Image Anal. 63 (2020) 101715, https://doi.org/10.1016/j.media.2020.101715.
[3] M.M. Islam, H.C. Yang, T.N. Poly, W.S. Jian, Y.C. (Jack) Li, Deep learning algorithms for detection of diabetic retinopathy in retinal fundus photographs: a systematic review and meta-analysis, Comput. Methods Programs Biomed. 191 (2020) 105320, https://doi.org/10.1016/j.cmpb.2020.105320.
[4] S. Stolte, R. Fang, A survey on medical image analysis in diabetic retinopathy, Med. Image Anal. 64 (2020) 101742, https://doi.org/10.1016/j.media.2020.101742.
[5] J. Krause, et al., Grader Variability and the Importance of Reference Standards for Evaluating Machine Learning Models for Diabetic Retinopathy, Ophthalmology. 125 (8) (2018) 1264–1272, https://doi.org/10.1016/j.ophtha.2018.01.034.
[6] M.D. Abramoff, M.K. Garvin, M. Sonka, Retinal imaging and image analysis, IEEe Rev. Biomed. Eng. 3 (2010) 169–208, https://doi.org/10.1109/RBME.2010.2084567.
[7] C.P. Wilkinson, et al., Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales, Ophthalmology. 110 (9) (2003) 1677–1682, https://doi.org/10.1016/S0161-6420(03)00475-5.
[8] M.D. Abràmoff, et al., Automated analysis of retinal images for detection of referable diabetic retinopathy, JAMA Ophthalmol. 131 (3) (2013) 351–357, https://doi.org/10.1001/jamaophthalmol.2013.1743.
[9] T. Li, et al., Applications of deep learning in fundus images: a review, Med. Image Anal. 69 (2021) 101971, https://doi.org/10.1016/j.media.2021.101971. Elsevier B.V.
[10] V. Gulshan, et al., Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs, JAMa 316 (22) (2016) 2402, https://doi.org/10.1001/jama.2016.17216.
[11] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the Inception Architecture for Computer Vision, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826, https://doi.org/10.1109/CVPR.2016.308.
[12] G. Quellec, K. Charrière, Y. Boudi, B. Cochener, M. Lamard, Deep image mining for diabetic retinopathy screening, Med. Image Anal. 39 (2017) 178–193, https://doi.org/10.1016/j.media.2017.04.012.
[13] M.D. Abràmoff, et al., Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning, Invest. Ophthalmol. Vis. Sci. 57 (13) (2016) 5200–5206, https://doi.org/10.1167/iovs.16-19964.
[14] P. Costa, et al., EyeWeS: weakly supervised pre-trained convolutional neural networks for diabetic retinopathy detection, in: Proceedings of the 16th International Conference on Machine Vision Applications, MVA 2019, Institute of Electrical and Electronics Engineers Inc., 2019, pp. 1–6, https://doi.org/10.23919/MVA.2019.8757991.
[15] C. González-Gonzalo, B. Liefers, B. Van Ginneken, C.I. Sánchez, Improving weakly-supervised lesion localization with iterative saliency map refinement, in: International Conference on Medical Imaging with Deep Learning, 2018.
[16] J. de la Torre, D. Puig, A. Valls, Weighted kappa loss function for multi-class classification of ordinal data in deep learning, Pattern. Recognit. Lett. 105 (2018) 144–154, https://doi.org/10.1016/j.patrec.2017.05.018.
[17] J. de la Torre, A. Valls, D. Puig, A deep learning interpretable classifier for diabetic retinopathy disease grading, Neurocomputing. 396 (2020) 465–476, https://doi.org/10.1016/j.neucom.2018.07.102.
[18] A. Vaswani, et al., Attention is all you need. Advances in Neural Information Processing Systems, Neural information processing systems foundation, 2017, pp. 5999–6009.
[19] L.C. Chen, Y. Yang, J. Wang, W. Xu, A.L. Yuille, Attention to Scale: scale-Aware Semantic Image Segmentation, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, 2016, pp. 3640–3649, https://doi.org/10.1109/CVPR.2016.396.
[20] Z. Wang, Y. Yin, J. Shi, W. Fang, H. Li, X. Wang, Zoom-in-net: deep mining lesions for diabetic retinopathy detection, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2017, pp. 267–275, https://doi.org/10.1007/978-3-319-66179-7_31.
[21] A. He, T. Li, N. Li, K. Wang, H. Fu, CABNet: category Attention Block for Imbalanced Diabetic Retinopathy Grading, IEEe Trans. Med. ImAging 40 (1) (2021) 143–153, https://doi.org/10.1109/TMI.2020.3023463.
[22] Z. Ai, X. Huang, Y. Fan, J. Feng, F. Zeng, Y. Lu, DR-IIXRN : detection Algorithm of Diabetic Retinopathy Based on Deep Ensemble Learning and Attention Mechanism, Front. Neuroinform. 15 (2021) 66, https://doi.org/10.3389/FNINF.2021.778552/BIBTEX.
[23] Z. Lin et al., "A Framework for Identifying Diabetic Retinopathy Based on Anti-noise Detection and Attention-Based Fusion," 2018, pp. 74–82. doi:10.1007/978-3-030-00934-2_9.
[24] G. Yue, Y. Li, T. Zhou, X. Zhou, Y. Liu, T. Wang, Attention-Driven Cascaded Network for Diabetic Retinopathy Grading from Fundus Images, Biomed. Signal. Process. Control 80 (2023) 104370, https://doi.org/10.1016/j.bspc.2022.104370.
[25] A. Bhati, N. Gour, P. Khanna, A. Ojha, N. Werghi, An interpretable dual attention network for diabetic retinopathy grading: iDANet, Artif. Intell. Med. (2024) 102782, https://doi.org/10.1016/j.artmed.2024.102782.
[26] R. Romero-Oraá, M. García, J. Oraá-Pérez, M.I. López-Gálvez, R. Hornero, Effective fundus image decomposition for the detection of red lesions and hard exudates to aid in the diagnosis of diabetic retinopathy, Sensors (Switzerland) 20 (22) (2020) 1–17, https://doi.org/10.3390/S20226549.
[27] Kaggle, "Diabetic Retinopathy Detection competition." [Online]. Available: https://www.kaggle.com/c/diabetic-retinopathy-detection/.
[28] J. Lin, L. Yu, Q. Weng, X. Zheng, Retinal image quality assessment for diabetic retinopathy screening: a survey, Multimed. Tools. Appl. (2019), https://doi.org/10.1007/s11042-019-07751-6.
[29] R. Romero Oraá, M. García, J. Oraá-Pérez, I.López María, R. Hornero, Automatic fundus image quality assessment: diagnostic accuracy in clinical practice, Invest. Ophthalmol. Vis. Sci. 61 (2020) 2033.
[30] A. Rakhlin, "Diabetic Retinopathy detection through integration of Deep Learning classification framework," *bioRxiv*, p. 225508, 2018, doi:10.1101/225508.
[31] M. Voets, K. Møllersen, 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 14 (6) (2019), https://doi.org/10.1371/JOURNAL.PONE.0217541.
[32] G.M. Lin, et al., Transforming retinal photographs to entropy images in deep learning to improve automated detection for diabetic retinopathy, J. Ophthalmol. 2018 (2018), https://doi.org/10.1155/2018/2159702.
[33] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. 2016. doi:10.1007/978-981-13-9113-2_16.
[34] R. Romero-Oraá, M. García, J. Oraá-Pérez, M.I. López, R. Hornero, A robust method for the automatic location of the optic disc and the fovea in fundus images, Comput. Methods Programs Biomed. 196 (2020) 105599, https://doi.org/10.1016/j.cmpb.2020.105599.
[35] L. Perez and J. Wang, "The Effectiveness of Data Augmentation in Image Classification using Deep Learning,"2017.
[36] F. Chollet, Xception: deep learning with depthwise separable convolutions, in: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Institute of Electrical and Electronics Engineers Inc., 2017, pp. 1800–1807, https://doi.org/10.1109/CVPR.2017.195.
[37] S.J. Pan, Q. Yang, A Survey on Transfer Learning, IEEe Trans. Knowl. Data Eng. 22 (10) (2010) 1345–1359, https://doi.org/10.1109/TKDE.2009.191.
[38] Jia Deng, Wei Dong, R. Socher, Li-Jia Li, Kai Li, Li Fei-Fei, ImageNet: a large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255, https://doi.org/10.1109/cvprw.2009.5206848.
[39] C. Szegedy, et al., Going deeper with convolutions, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9, https://doi.org/10.1109/CVPR.2015.7298594.
[40] C. Bishop, Neural Networks for Pattern Recognition, 1st ed., 1995, Oxford University Press, New York, 1995.
[41] T.Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollar, Focal loss for dense object detection, IEEe Trans. Pattern. Anal. Mach. Intell. 42 (2) (2020) 318–327, https://doi.org/10.1109/TPAMI.2018.2858826.
[42] J. Cohen, A Coefficient of Agreement for Nominal Scales, Educ. Psychol. Meas. 20 (1) (1960) 37–46, https://doi.org/10.1177/001316446002000104.
[43] J. Cohen, Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit, Psychol. Bull. 70 (4) (1968) 213–220, https://doi.org/10.1037/h0026256.
[44] R. Nirthika, S. Manivannan, A. Ramanan, Siamese network based fine grained classification for Diabetic Retinopathy grading, Biomed. Signal. Process. Control 78 (2022) 103874, https://doi.org/10.1016/J.BSPC.2022.103874.
[45] J. Zhang, Z. ming Zhang, Ethics and governance of trustworthy medical artificial intelligence, BMC. Med. Inform. Decis. Mak. 23 (1) (2023), https://doi.org/10.1186/S12911-023-02103-9.