

Article

# Diabetic Retinopathy Diagnosis Based on RA-EfficientNet

San-Li Yi, Xue-Lian Yang \*, Tian-Wei Wang, Fu-Rong She, Xin Xiong and Jian-Feng He \*

School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China; 20120024@kust.edu.cn (S.-L.Y.); 20192204062@stu.kust.edu.cn (T.-W.W.); 20202104016@stu.kust.edu.cn (F.-R.S.); xiongxin@kust.edu.cn (X.X.)

\* Correspondence: 20192304004@stu.kust.edu.cn (X.-L.Y.); jfenghe@kust.edu.cn (J.-F.H.)

**Abstract:** The early detection and grade diagnosis of diabetic retinopathy (DR) are very important for the avoidance of blindness, and using deep learning methods to automatically diagnose DR has attracted great attention. However, the small amount of DR data limits its application. To automatically learn the disease's features and detect DR more accurately, we constructed a DR grade diagnostic model. To realize the model, the authors performed the following steps: firstly, we preprocess the DR images to solve the existing problems in an APTOS 2019 dataset, such as size difference, information redundancy and the data imbalance. Secondly, to extract more valid image features, a new network named RA-EfficientNet is proposed, in which a residual attention (RA) block is added to EfficientNet to extract more features and to solve the problem of small differences between lesions. EfficientNet has been previously trained on the ImageNet dataset, based on transfer learning technology, to overcome the small sample size problem of DR. Lastly, based on the extracted features, two classifiers are designed, one is a 2-grade classifier and the other a 5-grade classifier. The 2-grade classifier can diagnose DR, and the 5-grade classifier provides 5 grades of diagnosis for DR, as follows: 0 for No DR, 1 for mild DR, 2 for moderate, 3 for severe and 4 for proliferative DR. Experiments show that our proposed RA-EfficientNet can achieve better performance, with an accuracy value of 98.36% and a kappa score of 96.72% in a 2-grade classification and an accuracy value of 93.55% and a kappa score of 91.93% in a 5-grade classification. The results indicate that the proposed model effectively improves DR detection efficiency and resolves the existing limitation of manual feature extraction.

**Keywords:** EfficientNet; transfer learning; residual attention block; retinal image; diabetic retinopathy



**Citation:** Yi, S.-L.; Yang, X.-L.; Wang, T.-W.; She, F.-R.; Xiong, X.; He, J.-F. Diabetic Retinopathy Diagnosis Based on RA-EfficientNet. *Appl. Sci.* **2021**, *11*, 11035. <https://doi.org/10.3390/app112211035>

Academic Editor: Cecilia Di Ruberto

Received: 17 September 2021

Accepted: 16 November 2021

Published: 22 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

According to WHO statistics, the number of adults with diabetes in the world reached 463 million in 2019. This number is expected to increase significantly in the future, reaching 700 million by 2045. Diabetic retinopathy (DR) presents one of the most serious complications of diabetes. Visual impairment is irreversible in diabetic retinopathy, and presents different pathological features at different stages, eventually causing eye damage and leading to blindness. Therefore, the early screening and diagnosis of DR are conducive to timely and effective treatment of diabetic patients.

Diabetic retinopathy is the most evident symptom of diabetes, which is characterized by microaneurysms, exudates, new blood vessel formation, hemorrhage, etc. Generally, DR can be divided into two stages, non-proliferative and proliferative retinopathy. The non-proliferative stage can be further classified as mild, moderate or severe. The mild stage is the early stage with small bleeding spots or small microhemangiomas. The moderate stage is subsequent to the mild stage, during which some yellowish-white punctate hard exudates may be examined. The severe stage is the last stage of non-proliferative retinopathy, accompanied by white, cotton-like, soft exudate. During the second DR stage, proliferative retinopathy, retinal damage will stimulate new blood vessel proliferation, which will further cause massive bleeding in the retina and vitreous body, leading to severe loss of

vision or even complete blindness. Since the characteristics of diabetic retinopathy are complex and diverse, finding the corresponding features of the disease is a challenge.

At present, in clinical practice, doctors diagnose DR using fundus photograph, which is still a time-consuming artificial method. However, the increasing number of DR patients and the scarcity of high-level ophthalmology experts will likely result in a missed diagnosis, misdiagnosis alongside other problems. Computer aided diagnosis (CAD) does not present the problem of manual diagnosis, which greatly reduces the workload and the time spent by doctors in diagnosing diseases, and provides high accuracy. Recently, deep learning has led to great progress and made valuable research and application contributions in the field of CAD analysis. A Convolution Neural Network (CNN) is most effective in the field of computer vision for its excellent performance in image classification tasks. In recent years, scholars have proposed multiple CNN algorithms, such as VGGNet [1], GoogLeNet [2] and ResNet [3], however, as a routine operation in these works, they usually only scale one of the three dimensions—depth, width and resolution, which often leads to poor accuracy and efficiency. EfficientNet [4] demonstrates better performance by uniformly scaling depth, width and resolution, and this algorithm provides a new research direction for the subsequent development of CNN. The development of CNN has achieved revolutionized progress in medical areas, such as retinal vascular segmentation [5], glaucoma screening [6], cancer subtype classification [7] and so on. The application of these CNN architectures is well suited for medical classification tasks, nevertheless, the problem of insufficient medical data has emerged, which makes the model's training challenging. Transfer learning [8] is a technique which, to ensure that the network is well trained, transfers information from the pre-trained dataset, which includes a huge number of images, to a new dataset. It is an excellent tool for the task of making the network more efficient and stable based on insufficient data.

Deep learning has become a research hotspot in the medical field, and its development has effectively promoted the progress of DR research. Gargeya et al. [9] used ResNet and a decision tree classifier to distinguish between sick and healthy images, and the AUC for the Messidor dataset is up to 0.94. Chetoui et al. [10] used EfficientNet combined with transfer learning to detect referable diabetic retinopathy (RDR) and vision-threatening DR (VTDR), which obtained satisfying results of up to 0.98 AUC both for the APTOS 2019 dataset and EyePACS dataset. Rao et al. [11] employed pre-trained ResNet50 to achieve 96.59% accuracy for an APTOS 2019 dataset for the binary classification task, that is to detect No-DR or DR. The above classification models performed well, however they are of binary classification, meaning that it is unable to make an in-depth classification of specific diseases. Therefore, it is necessary to further study the grade classification of DR. Shanthi et al. [12], based on the Messidor dataset, designed a neural network to automatically classify normal images, for stage 1, stage 2, and stage 3 of DR, with an accuracy up to 96% for each grade. This algorithm realizes four classifications of DR, that can better reflect the severity of DR than the use of two classifications. In recent years, the five classifications of DR have attracted greater attention because they are able to better reflect the DR and its severity levels. Dondeti et al. [13], based on the APTOS 2019 dataset, combined the pre-training model NASNET with the T-SNE space to extract deep features, achieving an accuracy rate of 77.90%. Bodapati et al. [14] proposed a composite deep neural network of Xception and VGG16 with gated attention mechanism to automatically diagnose DR. The accuracy of this model on the APTOS 2019 dataset was 82.54%. Majumder et al. [15] combined the Xception model with a regression model to classify the five stages of DR, achieving an accuracy of 82%, 86% for the EyePACS and APTOS 2019 datasets respectively. Patel et al. [16] used the pre-trained MobileNetV2 to classify the DR of the APTOS 2019 dataset, and obtained an accuracy of 91%. With the development of deep learning, the ability of a network to extract DR features has been enhanced, and the accuracy has been improved to some extent. However, for clinical diagnosis, the accuracy of a DR severity classification still deserves further improvement.

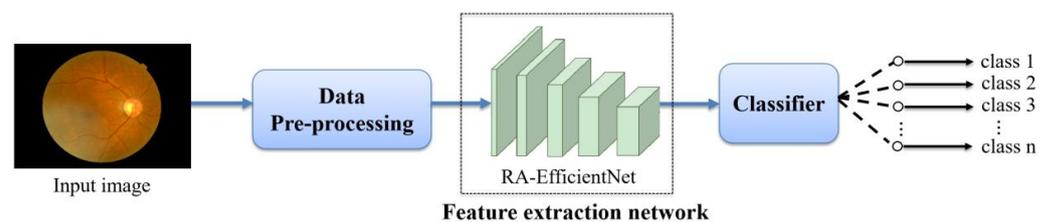
From the above statements, we can see that, firstly, in the diagnosis of DR, deep learning methods have received more attention for their improved performance; secondly, the grade classification for different severity levels of DR is easier for doctors to interpret; thirdly, the deep learning networks combined with transfer learning technology can provide more accurate results, especially with regard to the latest algorithms. Therefore, a simple and efficient network is required for a more accurate and effective diagnosis of DR of different severities.

Based on the above analysis, we constructed a DR grade diagnostic model in this paper. The steps performed for the construction of this model are summarized as follows: firstly, to reduce the inclusion of redundant information and to make the APTOS 2019 dataset more suitable for the diagnosis model, pre-processing steps are adopted. Secondly, to extract the features of DR images more accurately and efficiently, we propose the creation of a new network named RA-EfficientNet for the feature extractor of the model, based on the combination of EfficientNet and the residual attention block i.e., RA block. Finally, according to classification tasks, we designed two classifiers, one of which is a 2-grade classifier used for the identification of DR, and the other is a 5-grade classifier for the diagnosis of DR with respect to severity levels.

The rest of the paper is organized as follows: Section 2 introduces the workflow of our proposed DR diagnostic model, including data pre-processing, the structure of RA-EfficientNet and the classifiers for two classification tasks. In Section 3, the experimental process and the evaluation of the performance of different networks are illustrated. Finally, the work of the paper is summarized in Section 4.

## 2. Methodology

The DR diagnostic model of this work is mainly constructed using the following three steps: the pre-processing of input data, the network of feature extraction and classifiers. The pipeline of the DR diagnostic model is shown in Figure 1.



**Figure 1.** The pipeline of the DR diagnostic model.

### 2.1. Dataset

In this study, the classification experiments are conducted based on the APTOS 2019 dataset, downloaded from the kaggle website [17]. The dataset is provided by Aravind Eye Hospital in India, and includes 3662 images of a high resolution with each sample attached, diagnosed by highly trained doctors. According to the diagnosis, these images have been divided into 5 grades from 0 to 4 to present the severity level of diabetic retinopathy, where grade 0 includes 1805 normal images, indicating no diabetic retinopathy; grade 1, includes 370 images, indicating mild non-proliferative diabetic retinopathy; grade 2 includes 999 images, indicating moderate non-proliferative; grade 3, which includes 193 images, showing severe non-proliferative diabetic retinopathy, and grade 4 with 295 images of proliferative diabetic retinopathy.

It is evident that the distribution of the 5 grades is severely uneven, which will be further discussed below. Furthermore, Figure 2 shows the retinal images with grades ranging from 0 to 4.

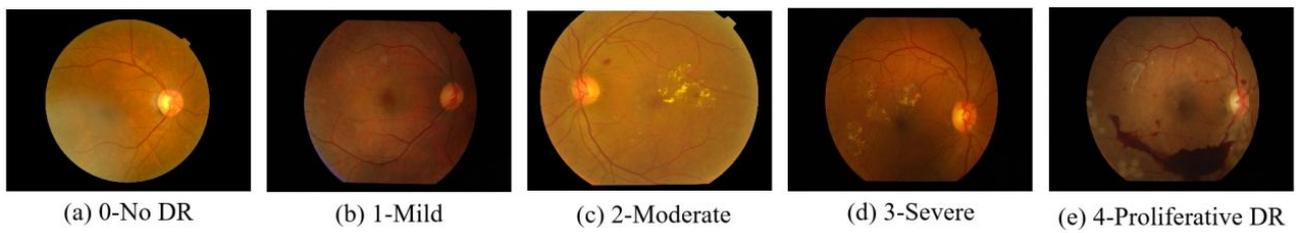


Figure 2. Image samples based on different severity.

### 2.2. Data Pre-Processing

To reduce the network’s over-fitting and enhance its learning ability, the dataset is pre-processed, which mainly includes four sections:

As is shown in Figure 2, the black border of the fundus images is large, containing no information regarding the lesion area. To reduce useless information in the images, the black border is cropped, and all the images are resized to  $1024 \times 1024$ .

As previously mentioned, the distribution of the dataset is severely uneven, which is shown in Figure 3. To manage this problem, we use data amplification techniques such as rotation, flip, brightness and contrast adjustment to form a data set with each grade containing 2000 samples.

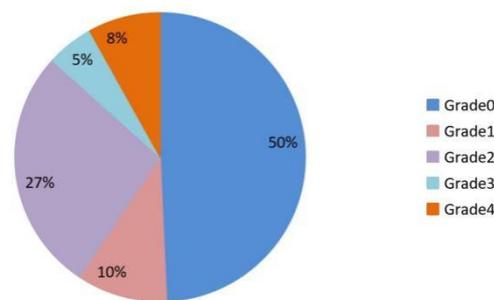


Figure 3. Grade distribution of APTOS 2019.

In order to facilitate convergence during network training, the range of original images is scaled between  $[0, 1]$ .

The dataset is divided into training sets and testing sets according to the ratio of 8:2, whereby 80% of the data is used for training and the remaining 20% for testing.

### 2.3. Feature Extraction Network

This section presents the structure of a feature extraction network. Firstly, the image data is input to the EfficientNet, which is pre-trained on the dataset of ImageNet by using transfer learning technology. Secondly, following the EfficientNet, a new function module named RA block is added, forming RA-EfficientNet. Lastly, the output of RA-EfficientNet is sent to the classifiers. The structure of the feature extraction network is shown in Figure 4.

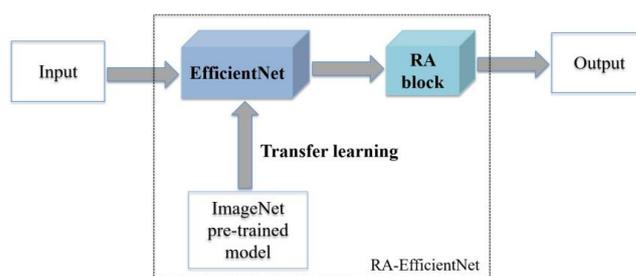


Figure 4. The structure of feature extraction network.

### 2.3.1. Transfer Learning

Given that sufficient public datasets for the diagnosis of DR do not exist, it is difficult to obtain a satisfactory result based on deep learning technology. To solve this problem, transfer learning technology is adopted in our model, and its steps of implementation are as follows: (1) Pre-training the network. To obtain the pre-trained network, the EfficientNet is pre-trained on ImageNet, which is currently the largest image recognition dataset in the world, with 1.2 million images of 1000 categories. Then the diabetes data is loaded into the pre-trained EfficientNet. (2) Fine-tuning. The final step of the feature extraction backbone is a fine-tuning stage, at which we build a new fully connected layer and use Adam as optimizer with 0.001 learning rate.

### 2.3.2. EfficientNet

EfficientNet, an efficient and robust network developed by scaling three parameters i.e., depth, width, and resolution, has been adopted in the model. The network has different versions, ranging from B0 to B7, and in this paper we choose EfficientNetB0 as the feature extractor. In this network, the size of the input image is set as  $224 \times 224$ , and the key building block of the network is mobile inverted bottleneck convolution, MBConv, which is an addition of the squeeze-and-excitation network (SENet) [18] structured on the basis of an inverted residual block, introduced in MobileNetV2 [19].

Figure 5 is the schematic representation of EfficientNetB0. The network has 16 MBConv blocks with the kernel of each MBConv block set as  $3 \times 3$  or  $5 \times 5$ . From the figure we can see that the input data, i.e., the fundus image is first placed in the network and then processed through a  $3 \times 3$  Conv2D layer, 16 MBConv layers and a  $1 \times 1$  Conv2D layer in sequence, and finally, the output of the network is sent to the layers of the next operation, which is the RA block.

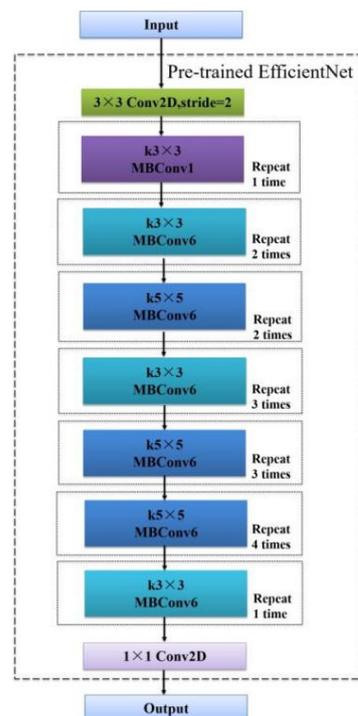


Figure 5. The Schematic representation of EfficientNet.

In EfficientNet, a new compound scaling method is proposed—using a compound coefficient  $\phi$  to uniformly scale the network width, depth, and resolution, so as to improve the accuracy of the model when the model parameters and calculation amount are maxi-

mized. The maximum accuracy of the model is denoted as, and the specific formula is as Equation (1).

$$N(d, w, r) = \odot \hat{F}_i^{d \times \hat{L}_i} \left( X_{[r \times \hat{H}_i, r \times \hat{W}_i, w \times \hat{C}_i]} \right) \quad (1)$$

$$i = 1, 2, \dots, s$$

In Equation (1),  $N$  represents the classification network,  $\odot$  represents the convolution operation,  $i$  represents the number of convolutional layers,  $\hat{F}_j$ ,  $\hat{L}_i$ ,  $\hat{H}_i$ ,  $\hat{W}_i$ ,  $\hat{C}_i$  are predefined parameters in the baseline network. Meanwhile,  $w$ ,  $d$  and  $r$ , are coefficients for scaling the network width, depth and resolution, and their calculations are as follows:

$$\text{depth: } d = \alpha^\varnothing \quad (2)$$

$$\text{width: } w = \beta^\varnothing \quad (3)$$

$$\text{resolution: } r = \gamma^\varnothing \quad (4)$$

$$\alpha \geq 1, \beta \geq 1, \gamma \geq 1$$

In order to obtain the three dimensional parameters that satisfy the Equation (1), the composite parameter  $\varnothing$  is used to optimize the depth, width and resolution of the network. First, we fix  $\varnothing = 1$ , and then the best values that satisfy the Equations (1)–(4) are found for  $\alpha$ ,  $\beta$ ,  $\gamma$  through a grid search. More specifically, we find that the best value for EfficientNetB0 is  $\alpha = 1.2$ ,  $\beta = 1.1$ ,  $\gamma = 1.15$ , under a constraint of  $\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$ . Then, we fix  $\alpha$ ,  $\beta$ ,  $\gamma$  as constants and scale up the baseline network with different  $\varnothing$  using Equations (2)–(4), to obtain EfficientNetB1 to B7.

### 2.3.3. RA Block

For the five classification tasks of DR, features of small differences such as the microaneurysms and exudation are critical for DR image classification. In the feature extraction, the attention mechanism can be added to allow the network to independently select the area that requires attention and can be learned, so the type of lesion can be assessed. Inspired by ResNet [3] and GCNet [20], a new function module named RA block that can further extract DR images' features, with small differences, is designed, and is shown in Figure 6.

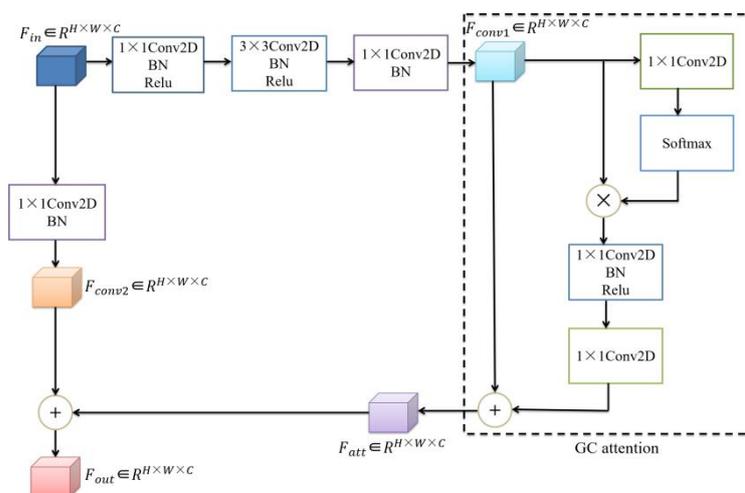


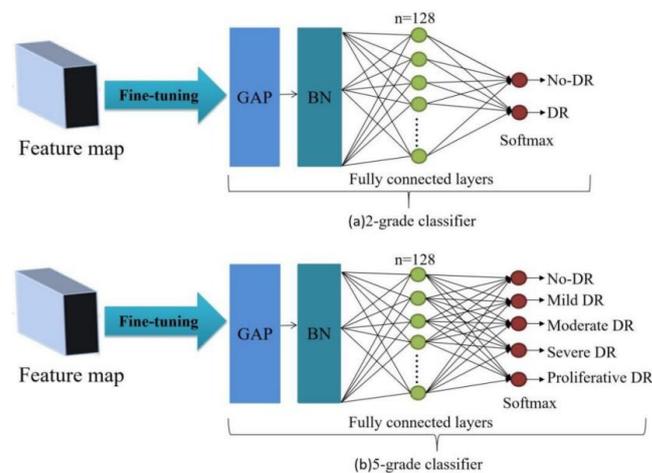
Figure 6. The structure of the RA block.

In Figure 6, the EfficientNet's output  $F_{in} \in R^{H \times W \times C}$  is the input of this block, which contains high-level semantic features of the fundus images, where  $H$ ,  $W$  and  $C$  denote height, width and the number of channels of the feature maps respectively. Firstly,  $F_{in}$  is placed into two different convolution operation paths, and two kinds of features are

obtained,  $F_{conv1} \in R^{H \times W \times C}$  and  $F_{conv2} \in R^{H \times W \times C}$ .  $F_{conv1}$  is obtained by three convolution operations, which increase the cross-channel information interaction and reduces the necessary levels of computation. Then, it is output to the GC attention module. The GC attention module gains channel-wise attention feature maps  $F_{att} \in R^{H \times W \times C}$ , which can be used to highlight the main lesion information and suppress less useful information. Finally, we add  $F_{att}$  with  $F_{conv2}$  and obtain the output feature maps  $F_{out} \in R^{H \times W \times C}$ .

#### 2.4. Classifiers

There are two classifiers for our tasks, a 2-grade and a 5-grade classifier. The 2-grade classifier is used for DR identification tasks, and a 5-grade classifier is used for a severity level prediction of DR. The two classifiers are shown in Figure 7, and the detailed explanation is as below:



**Figure 7.** The proposed 2-grade and 5-grade classifiers for classification.

The feature map of the fundus images is extracted by RA-EfficientNet, then the fine-tuning of the hyper-parameters is performed to build new fully connected layers, that are then matched with the new classifiers. The new and fully connected layers consist of a global average pooling (GAP) layer, BN layer and Softmax layer. The GAP layer reduces the spatial resolution to a single set of data and the nodes for the filters of the RA block is set as 128. The BN layer can enhance the network convergence and the last Softmax layer is used for different classification tasks of fundus images.

### 3. Experiments

In this section, experiments are detailed to present the efficiency of RA-EfficientNet, which is the feature extraction network of our model. The work conducted in the experimental phase is based on the Tensorflow framework.

#### 3.1. Evaluation Indicators

To prove the effectiveness of our DR diagnose model, we use different state-of-the-art CNN networks, such as InceptionResNetV2 [21], MobileNetV2 [19] and Xception [22] for comparison. All these networks are combined with transfer learning respectively. Moreover, to prove the advancement of our model in a 5 classification task, the results of our model are compared with those of the models proposed in Section 1.

In order to present the results of the above diagnostic models, the following evaluation methods are applied: firstly, we adopt evaluation metrics such as accuracy, precision, and an F1 and kappa score for comparison. Secondly, the normalized confusion matrix, a matrix containing the proportion of samples of any class labelled as any of the possible outputs, is provided. Thirdly, the receiver operating characteristic (ROC) curve, as well as the area

under the ROC (AUC) are computed to present the performance of the model classifying task. These metrics are defined as follows:

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

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

$$\text{Kappa} = \frac{\text{ObservedAccuracy} - \text{ExpectedAccuracy}}{1 - \text{ExpectedAccuracy}} \quad (8)$$

### 3.2. Experimental Results and Discussion

The experiment has two different tasks: task 1 is a 2-grade classification i.e., the identification of DR and task 2 is a 5-grade classification i.e., grading the severity levels of DR. In order to train the network better, the super parameters of the experiment are set as the loss function of sparse `_categorical_crossentropy`, for which the batch size is 32, and the number of epochs is 60. The results of these tasks are presented, and an analysis of the results is detailed below.

#### 3.2.1. Task 1: 2-Grade Classification

The goal of task 1 is to identify whether an image presents DR or No-DR. This is a binary classification task, requiring samples in a dataset with two types of data, DR images and No-DR images. According to this requirement, the APTOS 2019 dataset is divided into two classes, with 1857 DR images in one class and 1805 No-DR images in the other. Setting the ratio of the training set to the testing set as 8:2, i.e., 2929 images are used for training and 733 images for testing. Based on the above setup, our diagnostic model and comparison models are used for this task, and the results are presented below.

EfficientNet is comprised of eight versions, from B0 to B7. Table 1 summarizes the default image input size, the number of parameters and defines the accuracy for EfficientNet models in a binary classification.

**Table 1.** Comparison of different EfficientNet models.

Model	Input Size	Parameter	Accuracy (%)
EfficientNetB0	224 × 224	4,057,253	97.95
EfficientNetB1	240 × 240	6,582,914	97.54
EfficientNetB2	260 × 260	7,777,012	96.04
EfficientNetB3	300 × 300	10,792,746	95.90
EfficientNetB4	380 × 380	17,684,570	94.13
EfficientNetB5	456 × 456	28,525,810	95.22
EfficientNetB6	528 × 528	40,973,969	98.22
EfficientNetB7	600 × 600	64,113,049	98.36

Table 1 presents the performance levels for all eight of the EfficientNet models over image input size, the number of parameters and the accuracy, we can see that EfficientNetB7 achieves the highest accuracy, of 98.36% for a 2-grade classification task, but the input size is too large, and the parameter is the highest. EfficientNetB0 achieves an accuracy of 97.95%, with a small input size and lower parameter. Given the problem of hardware resources, we chose EfficientNetB0 as the feature extractor with a 224 × 224 input size.

Table 2 presents the comparison in the binary classification, displaying the evaluation metrics belonging to different CNN networks. From the table, we can see that EfficientNet performs better than other models with an accuracy, precision, F1, and a kappa score of 97.95%, 97.96%, 97.95%, and 95.90% respectively.

**Table 2.** Comparison of different CNN networks for Task 1.

Model	Accuracy (%)	Precision (%)	F1 (%)	Kappa (%)
InceptionResNetV2	97.54	97.55	97.54	95.08
Xception	97.81	97.85	97.81	95.63
MobileNetV2	97.68	97.68	97.68	95.36
EfficientNet	<b>97.95</b>	<b>97.96</b>	<b>97.95</b>	<b>95.90</b>

A comparison of the different CNN networks with an RA block for the binary classification is listed in Table 3. We can see from Table 3 that, as RA block can strengthen the attention of the lesion areas, the networks with an RA block perform better, presenting higher values of the evaluation indicators than the evaluation indicators in Table 2. Our network, i.e., RA-EfficientNet, also listed in the table, obtains the best results. Compared with EfficientNet in Table 2, RA-EfficientNet helps us improve the binary classification performance by nearly 4.1%.

**Table 3.** Comparison of different CNN networks with RA block for Task 1.

Model	Accuracy (%)	Precision (%)	F1 (%)	Kappa (%)
InceptionResNetV2 + RA	97.95	97.95	97.95	95.90
Xception + RA	98.22	98.23	98.22	96.45
MobileNetV2 + RA	98.09	98.11	98.09	96.18
RA-EfficientNet	<b>98.36</b>	<b>98.37</b>	<b>98.36</b>	<b>96.72</b>

Table 4 presents the comparison of parameters of different networks for task 1. Compared with other pre-trained networks, the parameter of EfficientNet is relatively small. Although the parameter of MobileNetV2 is the smallest overall, it is not as accurate as EfficientNet. Furthermore, RA-EfficientNet performs better than the rest, with its parameter only 5.2% higher than that of EfficientNet, which means the addition of computer calculation is acceptable.

**Table 4.** Comparison of parameters for different CNN networks for Task 1.

Model	Parameter
InceptionResNetV2	54,354,954
Xception	20,873,770
MobileNetV2	2,265,666
EfficientNet	4,057,253
RA-EfficientNet	<b>4,272,070</b>

The confusion matrix measures the classification accuracy of a classifier. Each column is a predicted category, and each row is the true category of data. The diagonal line indicates that its prediction is the same as the true value, and the more accurate the predictions are, the darker the color. Figure 8 provides the confusion matrix of EfficientNet and RA-EfficientNet for the 2-grade classification. Although there are 733 images available for testing, we are able to observe that the total number predicted to be No-DR or DR is 716, and the misclassification number is 17 in EfficientNet. When trained on RA-EfficientNet, we observe that the total number of correctly predicted images is 722, demonstrating the advantages of our diagnostic model.

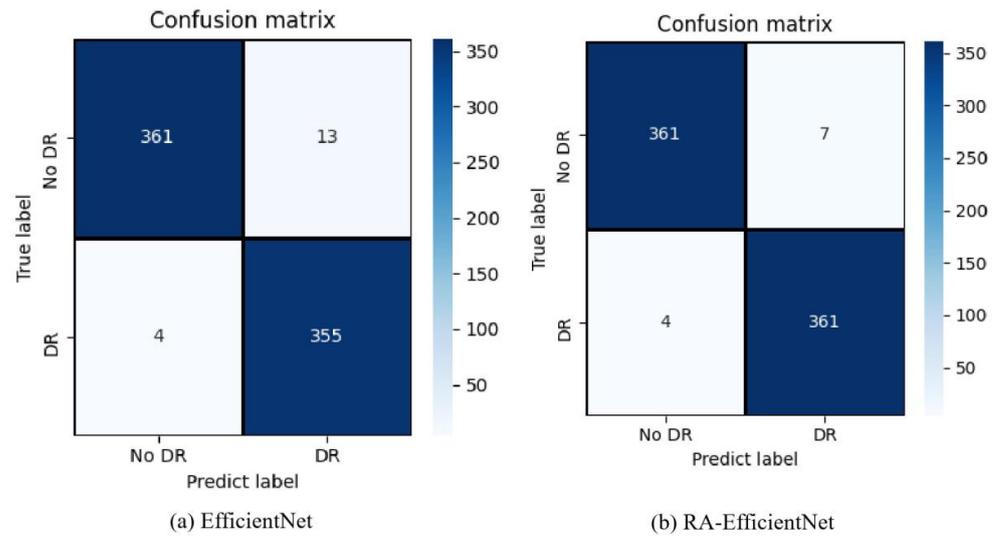


Figure 8. Confusion matrix of EfficientNet and RA-EfficientNet for task 1.

To better understand the representative of our RA-EfficientNet, we compare it with existing methods in the literature [11,14]. In Table 5, we can see that the proposed method achieves 98.36% accuracy, which performs significantly better than existing methods.

Table 5. Comparison with existing works for task 1 on APTOS 2019 dataset.

Method	Accuracy (%)	Precision (%)	F1 (%)
Rao et al. [11]	96.59	97.00	96.59
Bodapati et al. [14]	97.82	98.00	98.00
RA-EfficientNet	<b>98.36</b>	<b>98.37</b>	<b>98.36</b>

### 3.2.2. Task 2: 5-Grade Classification

Task 2 is a series of experiments for a 5-grade classification, which can make doctors' treatment more convenient and improve them by identifying the severity of DR. The APTOS 2019 dataset already has five grades, which is required for the task. The proportion in the training set and the testing set is the same as for task 1. Based on the 5-grades dataset, our diagnostic model and comparison models are conducted, and the results are presented below.

Table 6 lists the comparison of the same networks as task 1 on severity grades. The results show that EfficientNet obtains the highest values compared with other networks in accuracy, precision, F1 and kappa score, for which the value are 91.00%, 91.03%, 91.01%, and 88.74%, respectively.

Table 6. Comparison of different CNN networks for Task 2.

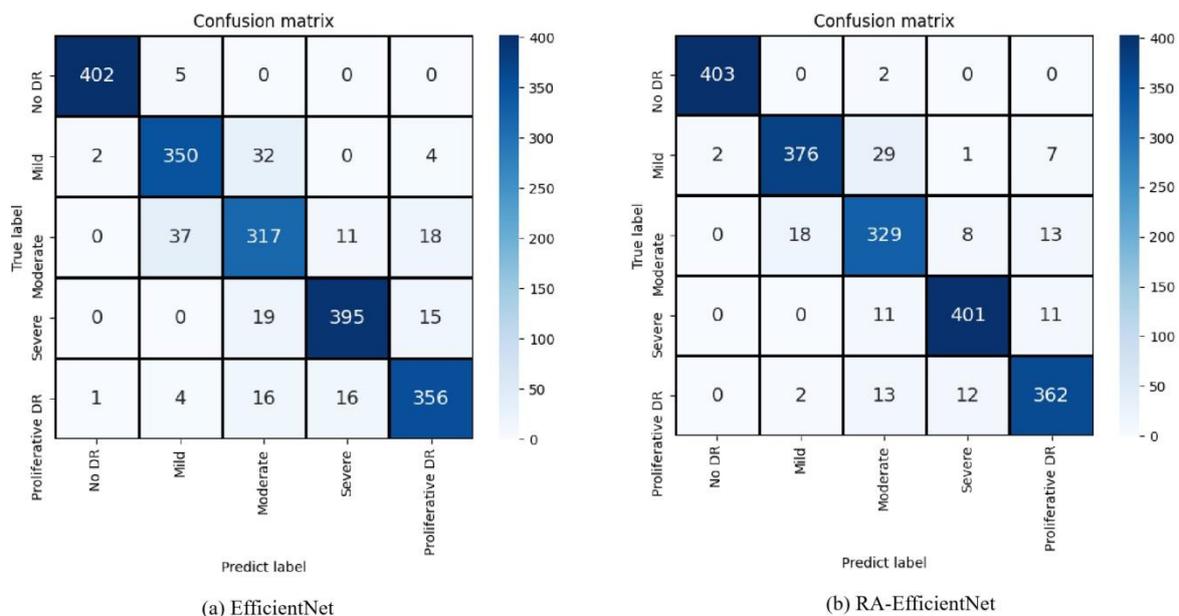
Model	Accuracy (%)	Precision (%)	F1 (%)	Kappa (%)
InceptionResNetV2	85.55	85.65	85.56	81.93
Xception	90.80	90.73	90.75	88.49
MobileNetV2	90.55	90.70	90.57	88.18
EfficientNet	<b>91.00</b>	<b>91.03</b>	<b>91.01</b>	<b>88.74</b>

Table 7 is the comparison of different CNN networks with RA block for task 2. From this table, the same conclusion as in Table 3 can be drawn.

**Table 7.** Comparison of different CNN networks and RA block for Task 2.

Model	Accuracy (%)	Precision (%)	F1 (%)	Kappa (%)
InceptionResNetV2 + RA	91.55	91.66	91.57	89.43
Xception + RA	93.00	93.06	93.02	91.24
MobileNetV2 + RA	92.60	92.75	92.65	90.74
RA-EfficientNet	<b>93.55</b>	<b>93.62</b>	<b>93.57</b>	<b>91.93</b>

Figure 9 is the confusion matrix of EfficientNet and RA-EfficientNet for task 2. From Figure 9a, we can see that the total number classified as No-DR, mild, moderate, severe and proliferative DR using EfficientNet, is 402, 350, 317, 395, and 356, respectively. For the moderate class, the rate of correct classification is lower than that in other severity levels, because the features of retinal images are difficult to identify at this severity level. From Figure 9b, we can see that the addition of RA significantly improves the classification results in the moderate level, and strengthens the accuracy for other levels.



**Figure 9.** Confusion matrix of EfficientNet and RA-EfficientNet for task 2.

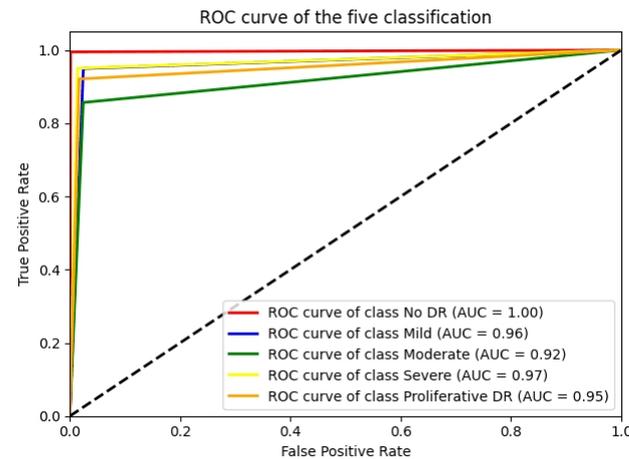
The ROC curve of our RA-EfficientNet model is shown in Figure 10. The ROC curve and AUC value represent how close the prediction is to a perfect classification, which is presented at the top left corner of the ROC coordinate. The AUC value shows the area under the ROC curve. The closer the value is to 1, the better the model performs. The figure shows the AUC value of 100%, 96.00%, 97.00% and 95.00%, representing No-DR, mild, severe and proliferative DR classes, respectively. Since a morphological variation of the fundus images, for the moderate grades, affects the recognition of pathological structures, the AUC of moderate DR is lower than the others with 92.00%.

Table 8 presents the comparison between our 5-grade diagnostic model and the five existing models proposed in Section 1, all of which are based on APTOS 2019 dataset.

**Table 8.** Comparison with existing works for DR severity grades on APTOS 2019 dataset.

Method	Accuracy (%)	Precision (%)	F1 (%)
Dondeti et al. [13]	77.90	76.00	75.00
Bodapati et al. [14]	82.54	82.00	82.00
Majumder et al. [15]	86.00	77.00	73.00
Patel et al. [16]	91.00	NA	NA
RA-EfficientNet	<b>93.55</b>	<b>93.62</b>	<b>93.57</b>

NA: represents not applicable.

**Figure 10.** ROC Curves for severity grades of Diabetic Retinopathy.

From Table 8, we can see that the RA-EfficientNet, used in this paper, achieves the highest performance accuracy for DR severity grades predictions. The model's accuracy improves by 2.55% as compared with the most advanced model recently proposed by Patel et al. [16].

To further prove the robustness of our RA-EfficientNet, we train and test the proposed RA-EfficientNet on the EyePACS dataset [23], from which we selected a total of 11,756 images. Table 9 shows a comparison of our RA-EfficientNet model, including three recent studies by Harihanth et al. [24], Majumder et al. [15] and He et al. [25] based on the EyePACS dataset. The results clearly illustrate that RA-EfficientNet performs better for the classification of the five stages of DR.

**Table 9.** Comparison with existing works for DR severity grades on EyePACS dataset.

Method	Accuracy (%)	Precision (%)	F1 (%)
Harihanth et al. [24]	81.85	70.00	56.00
Majumder et al. [15]	82.00	69.00	66.00
He et al. [25]	86.18	NA	NA
RA-EfficientNet	<b>89.29</b>	<b>89.92</b>	<b>89.29</b>

NA: represents not applicable.

#### 4. Conclusions

In this paper, we focus on 2-grade classification and 5-grade classification, which can be used for doctors' diagnosis of DR, of which a 5-grade classification provides more accurate grading information. We conduct experimental studies on many current deep learning networks using transfer learning technology, including InceptionResNetV2, MobileNetV2, Xception, and EfficientNet, among which EfficientNet achieves the best results because of the balance of the depth, width and resolution, with a 97.95% and 91.00% accuracy for a 2-grade and 5-grade classification task, respectively. To achieve more efficient results, we adopted an RA block modeled on pre-trained CNN networks, and all the models with an RA block performed better than the original networks. After a small increase in the

parameters, the evaluation index of the two tasks reached 98.36% and 93.55% accuracy, respectively, using RA-EfficientNet, which shows that our RA block better distinguishes between the lesion features of DR images. The results demonstrate that our diagnostic model provides the best performance and a better evaluation value for the APTOS 2019 dataset, compared to existing DR classification methods. Meanwhile, we verify the robustness of the proposed RA-Efficientnet algorithm by training and evaluating an EyePACS dataset of fundus images, through which we achieved good performance.

The advantages of this paper are as follows: (1) Two types of classification are realized in the model, proving its feasibility both for the diagnosis of DR and its severity levels; (2) Designs provided for the classification of DR, and the new network RA-EfficientNet performs better than other networks in comparison, with a higher accuracy and requires less computer-based calculations; (3) Combined with transfer learning, which can overcome the problem of a small sample size of DR, RA-EfficientNet has obtained satisfactory results. In the near future, we will further optimize the method to improve the accuracy of DR detection and try to develop a more powerful DR diagnosis model to assist doctors in clinical examinations.

**Author Contributions:** Conceptualization, S.-L.Y. and X.-L.Y.; methodology, X.-L.Y.; software, T.-W.W.; validation, F.-R.S., X.X. and J.-F.H.; formal analysis, S.-L.Y.; resources, X.-L.Y.; data curation, T.-W.W. and F.-R.S.; writing—original draft preparation, X.-L.Y.; writing—review and editing, S.-L.Y.; supervision, X.X.; project administration, J.-F.H.; funding acquisition, X.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the National Natural Science Foundation of China (No. 82060329); Yunnan Provincial Department of Education Project (No. 2020J0052).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Publicly available datasets were analyzed in this study. This data can be found here: [<https://www.kaggle.com/c/diabetic-retinopathy-detection>] accessed on 15 November 2021, [<https://www.kaggle.com/c/aptos2019-blindness-detection/data>] accessed on 15 November 2021.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
2. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
3. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
4. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, PMLR, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
5. Wang, S.; Yin, Y.; Cao, G.; Wei, B.; Zheng, Y.; Yang, G. Hierarchical retinal blood vessel segmentation based on feature and ensemble learning. *Neurocomputing* **2015**, *149*, 708–717. [[CrossRef](#)]
6. Raghavendra, U.; Fujita, H.; Bhandary, S.V.; Gudigar, A.; Tan, J.H.; Acharya, U.R. Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images. *Inf. Sci.* **2018**, *441*, 41–49. [[CrossRef](#)]
7. Hashimoto, N.; Fukushima, D.; Koga, R.; Takagi, Y.; Ko, K.; Kohno, K.; Nakaguro, M.; Nakamura, S.; Hontani, H.; Takeuchi, I. Multi-scale Domain-adversarial Multiple-instance CNN for Cancer Subtype Classification with Unannotated Histopathological Images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 3852–3861.
8. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2009**, *22*, 1345–1359. [[CrossRef](#)]
9. Gargeya, R.; Leng, T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology* **2017**, *124*, 962–969. [[CrossRef](#)] [[PubMed](#)]
10. Chetoui, M.; Akhloufi, M.A. Explainable Diabetic Retinopathy using EfficientNET. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montréal, QC, Canada, 20–24 July 2020; IEEE: Piscataway, NJ, USA; pp. 1966–1969.

11. Rao, M.; Zhu, M.; Wang, T. Conversion and Implementation of State-of-the-Art Deep Learning Algorithms for the Classification of Diabetic Retinopathy. *arXiv* **2020**, arXiv:2010.11692.
12. Shanthi, T.; Sabeenian, R.S. Modified Alexnet architecture for classification of diabetic retinopathy images. *Comput. Electr. Eng.* **2019**, *76*, 56–64. [[CrossRef](#)]
13. Dondeti, V.; Bodapati, J.D.; Shareef, S.N.; Veeranjanyulu, N. Deep Convolution Features in Non-linear Embedding Space for Fundus Image Classification. *Rev. d'Intell. Artif.* **2020**, *34*, 307–313. [[CrossRef](#)]
14. Bodapati, J.D.; Shaik, N.S.; Naralasetti, V. Composite deep neural network with gated-attention mechanism for diabetic retinopathy severity classification. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *12*, 9825–9839. [[CrossRef](#)]
15. Majumder, S.; Kehtarnavaz, N. Multitasking Deep Learning Model for Detection of Five Stages of Diabetic Retinopathy. *arXiv* **2021**, arXiv:2103.04207. [[CrossRef](#)]
16. Patel, R.; Chaware, A. Transfer Learning with Fine-Tuned MobileNetV2 for Diabetic Retinopathy. In Proceedings of the 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 5–7 June 2020; IEEE: Piscataway, NJ, USA; pp. 1–4.
17. Available online: <https://www.kaggle.com/c/aptos2019-blindness-detection/data> (accessed on 21 May 2021).
18. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
19. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
20. Cao, Y.; Xu, J.; Lin, S.; Wei, F.; Hu, H. Gcnet: Non-local networks meet squeeze-excitation networks and beyond. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Korea, 27–28 October 2019.
21. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the Thirty-first AAAI conference on artificial intelligence, San Francisco, CA, USA, 4–9 February 2017.
22. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
23. Available online: <https://www.kaggle.com/c/diabetic-retinopathy-detection> (accessed on 15 April 2021).
24. Harihanth, K.; Karthikeyan, B. Diabetic Retinopathy Detection using ensemble deep Learning and Individual Channel Training. In Proceedings of the 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Coimbatore, India, 3–5 December 2020; IEEE: Piscataway, NJ, USA; pp. 1042–1049.
25. He, A.; Li, T.; Li, N.; Wang, K.; Fu, H. CABNet: Category attention block for imbalanced diabetic retinopathy grading. *IEEE Trans. Med Imaging* **2020**, *40*, 143–153. [[CrossRef](#)] [[PubMed](#)]