

Article

Age-Related Macular Degeneration Detection in Retinal Fundus Images by a Deep Convolutional Neural Network

Andrés García-Floriano ^{1,*}  and Elías Ventura-Molina ² ¹ Instituto Politécnico Nacional, Escuela Superior de Cómputo, Mexico City 07738, Mexico² Instituto Politécnico Nacional, Centro de Innovación y Desarrollo Tecnológico en Cómputo, Mexico City 07700, Mexico; eventuram@ipn.mx

* Correspondence: agarciafl@ipn.mx; Tel.: +52-555-729-6000 (ext. 52519)

Featured Application: The Deep Neural Network model employed in this work will help build a system for the pre-diagnosis of retinopathies that can lead to blindness. The main intention of this type of system is to support the work performed by ophthalmology specialists.

Abstract: Computer-based pre-diagnosis of diseases through medical imaging is a task worked on for many years. The so-called fundus images stand out since they do not have uniform illumination and are highly sensitive to noise. One of the diseases that can be pre-diagnosed through fundus images is age-related macular degeneration, which initially manifests as the appearance of lesions called drusen. Several ways of pre-diagnosing macular degeneration have been proposed, methods based entirely on the segmentation of drusen with prior image processing have been designed and applied, and methods based on image pre-processing and subsequent conversion to feature vectors, or patterns, to be classified by a Machine-Learning model have also been developed. Finally, in recent years, the use of Deep-Learning models, particularly Convolutional Networks, has been proposed and used in classification problems where the data are only images. The latter has allowed the so-called transfer learning, which consists of using the learning achieved in the solution of one problem to solve another. In this paper, we propose the use of transfer learning through the Xception Deep Convolutional Neural Network to detect age-related macular degeneration in fundus images. The performance of the Xception model was compared against six other state-of-the-art models with a dataset created from images available in public and private datasets, which were divided into training/validation and test; with the training/validation set, the training was made using 10-fold cross-validation. The results show that the Xception neural network obtained a validation accuracy that surpasses other models, such as the VGG-16 or VGG-19 networks, and had an accuracy higher than 80% in the test set. We consider that the contributions of this work include the use of a Convolutional Neural Network model for the detection of age-related macular degeneration through the classification of fundus images in those affected by AMD (drusen) and the images of healthy patients. The performance of this model is compared against other methods featured in the state-of-the-art approaches, and the best model is tested on a test set outside the training and validation set.

Keywords: age-related macular degeneration; deep learning; transfer learning; fundus images

MSC: 68T07



Citation: García-Floriano, A.; Ventura-Molina, E. Age-Related Macular Degeneration Detection in Retinal Fundus Images by a Deep Convolutional Neural Network. *Mathematics* **2024**, *12*, 1445. <https://doi.org/10.3390/math12101445>

Academic Editor: Jakub Nalepa

Received: 22 March 2024

Revised: 26 April 2024

Accepted: 3 May 2024

Published: 8 May 2024



Copyright: © 2024 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

Age-related macular degeneration (AMD) is a chronic disease that can cause irreversible damage to the retina in people over 50 [1] and is currently considered the leading cause of blindness in the U.S. population over 50 years of age [2].

The appearance of affections called drusen denotes this disease. It can be more serious if these lesions appear in the area known as the macula lutea, which is considered the major

element of human vision. Therefore, one way to pre-diagnose this disease is by looking for these conditions in human retinal images [3]. Figure 1 illustrates healthy and AMD-affected retinas; note that the three fundamental structures of the retina are the optic disc (the shiny disc that can be observed in the images), the macula lutea (the brown spot that appears near the optic disc), and the retinal vascular network.

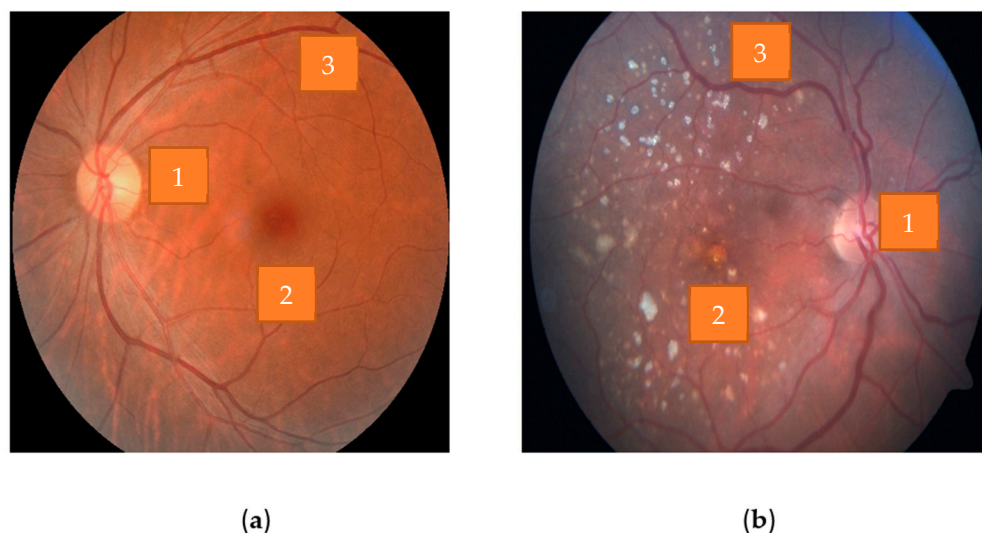


Figure 1. (a) Healthy human retina, where it is possible to see all its structural elements. (b) An AMD-affected retina, where elements like the macula are occluded by drusen [4]. The most relevant structural elements of the retina are 1. papilla or optic disc, 2. macula lutea, and 3. vascular network. Reproduced with permission from García-Florian, Sistema Integral de Análisis para la Prevención de la Ceguera; published by IPN, 2011.

All the techniques for automated analysis of retinal images depend on the type of the analyzed images. There are invasive methods like that proposed by Iwama et al. [3], who performed a pre-diagnosis through Optical Coherence Tomography (OCT) images combined with color and shape criteria. This proposal was evaluated on a dataset comprising 2034 images, obtaining an error of just over 3%. There are non-invasive methods for the pre-diagnosis of AMD. The so-called fundus images or ophthalmoscopic retinal images that are characterized by being obtained using a fundus camera [4]. However, they are difficult to analyze due to the presence of noise and non-uniform illumination [1]. These images are also characterized by presenting in color structural elements of the retina, such as the optic papilla or optic disc, the macula lutea, and the vascular network.

1.1. Literature Review

Regarding fundus images, many approaches have been proposed for their analysis. Mittal and Kumari [1] detected the presence of AMD by drusen segmentation in retinal images; for this purpose, they used a homomorphic filter and subsequently applied a Gaussian filter on the green grayscale channel of the retinal image. According to the authors, this method has an average accuracy of 96.17% in the ARIA [5] and STARE [6] public datasets, although the results were based on the segmentation of artifacts related to AMD.

The use of conventional image processing techniques has been successfully used in various retinopathies detection methods. Mvoulana, Kahouri and Akil [7] developed a pre-diagnostic method for glaucoma by using shape and area criteria regarding the optic disc of the eye. The authors tested their proposal on 10 public datasets, obtaining up to 98% performance; however, the techniques for glaucoma detection are based on the criteria of brightness, size, and shape of objects and are not compatible with AMD detection.

The accurate detection of AMD can depend on the results of segmentation operations; however, the computational cost of these operations [8] is high. With the rise of Graphics Processing Units (GPUs) and Deep-Learning models, segmentation methods based on Convolutional Neural Networks (CNNs) have emerged. Ronnenberger, Fischer and Brox [9] developed the U-Net, which seeks to solve two fundamental problems: the segmentation of objects in medical images and augmentation techniques to deal with the limited availability of images in medical image datasets. Twenty-three convolutional layers characterized the U-Net architecture. The contraction followed the typical behavior of Convolutional Neural Network models. The authors proved that by employing data augmentation, it was possible to solve the problem of having few images and provide the neural network with invariance to rotations, deformations, or variations in the gray levels. This model was tested against other image segmentation approaches by employing the ISBI cell-tracking challenge 2015 dataset, obtaining quite competitive results. Like other proposals concerning the state-of-the-art methods, the results can only be verified with a segmented dataset.

Based on U-Net, some authors proposed methods for the pre-diagnosis of diseases by analyzing medical images. In the work of Rundo et al. [10], a study was made in which three CNN architectures were compared in segmenting the central gland (CG) and the peripheral zone (PZ) of the prostate, with the aim of early detection of prostate cancer. This method was based on Magnetic Resonance Imaging (MRI). Two datasets were chosen: one composed of 193 images and the other sample of 503 images of 19 patients. The images were pre-processed, cropped, background filtered, segmented using some CNN model, and post-processed using some mathematical morphology techniques. Regarding the CNN model, the authors performed tests with SegNet, pix2pix, and U-Net, where the best results were obtained by U-Net; however, this method only works on grayscale images.

In recent years, some research works have proposed an alternative approach based on image processing techniques as pre-processing and subsequently feature extractors to form patterns or data vectors that were later classified using Machine-Learning models.

Koh et al. [11] developed a method for drusen detection by CLAHE: Contrast Limited Adaptive Histogram Equalization. The resulting images when using this method were subsequently processed with the 2D Continuous Wavelet Transform; the generated features were weighted with the Particle Swarm Optimization (PSO) metaheuristic. The dataset generated by the authors contemplates 404 images or patterns that are not affected by any disease, 381 affected by macular degeneration, 195 by diabetic retinopathy, and 506 by glaucoma. The pattern classification stage was performed with the Random Forest model, with 92.48% of accuracy, using 10-fold cross-validation as the validation method. Unlike the use of Deep-Learning models, this method required that the images are first processed to create a pattern dataset.

Another work based on the Machine-Learning approach were proposed by Vijayan et al. [12], who detected conditions related to diabetic retinopathy. In this work, they extracted 60 features from a Gabor filter. These features were used to build a database of 35,126 patterns. Classifiers such as the Random Forest, One Rule, IBK and J48 were tested with this dataset and validated with 10-fold cross-validation. The best performance was obtained with Random Forest since it obtained an accuracy of 70.15% and an ROC value of 0.862. It is noteworthy that the accuracy seems to be low regarding other works.

In recent years, the use of Deep Learning has gained special relevance [13]. Images can be classified without pre-processing operations or filters. This kind of classification can be achieved by designing Deep Neural Networks. Using Convolutional Neural Networks (CNNs) implies considering concepts such as convolution or pooling and incorporating new activation functions such as the Rectified Linear Unit (or ReLU).

Raghavendra et al. [14] developed a CNN-based tool for glaucoma detection in fundus images. This neural network can process pixels with three color channels. This network achieved up to 98% validation accuracy on a dataset composed of 589 normal cases and 837 glaucoma cases. The authors divided the dataset into 70% training and 30% testing.

This architecture was designed for glaucoma detection, where the classification criteria are different from the AMD criteria since it depends in larger bright objects.

Diabetic retinopathy is another target for CNN-based pre-diagnostic tools. Liu et al. [15] developed a CNN network and also adopted some ideas from classifier ensembles. The base architecture of this model was named WP-CNN. This architecture performed the classification under a voting scheme, achieving an accuracy of 94.23% in the STARE database [16], which is better than other models such as ResNet-50 and DenseNet-121. Unlike our work, that work did not consider the Xception model as an option for classification and was based on only one public dataset.

CNNs have been used as the basis for the detection of the hepatocellular carcinoma (HCC), the second most lethal tumor and the fourth leading cause of cancer-related mortality worldwide. Peng et al. proposed a method for preoperative prediction of the response of patients with intermediate-stage HCC by trans arterial chemoembolization (TACE). With this method, the authors took a dataset composed of Computed Tomography (CT) images of 789 patients from 3 different hospitals and adjusted the weights and parameters of the ResNet50 model. According to their results, the Deep Neural Network model achieved an accuracy of 84.3% and area under the ROC values of above 0.9 [17].

Another application for cancer detection was proposed by Sun et al., who developed a new multi-modal point-of-care system capable of predicting the response of the prediction of the preoperative trans arterial chemoembolization efficacy. According to the authors, the accuracy of this proposal is about 98% with the validation test and a cross-entropy loss of about 0.4. After the pre-processing phase, the authors used a model called GhostNet, which is intended to be used to generate more features maps by using a few parameters [18].

Regarding macular degeneration, some works based on Deep Learning or CNN models have been proposed. Das et al. [19] developed a cloud and Internet of Things (IoT)-based system that aims to provide an accurate pre-diagnosis of the presence of macular degeneration. In this system, the Res-Net Convolutional Network of 152 layers was used. The authors collected a total of 130,000 images, organized into 4 classes and divided as follows: 60% training, 20% validation, and 20% testing. With these data, the authors obtained an accuracy of 97.49% and an AUC of 0.97. Although the success of this model required a high-performance computing platform to work, in our research, we dealt with a smaller dataset and limited hardware resources.

Gour and Khana [20], developed a method whose aim was to pre-diagnose macular degeneration, glaucoma, and diabetic or hypertensive retinopathy in fundus images. This method, like our work, was based on the use of transfer learning, where a set of previously trained modes are used for classification. The chosen models were the Res-Net, Inception V3, Mobile-Net, and VGG-16 networks. This proposal was validated on the dataset named ODIR, of which 1744 were used as the test set. According to the authors, the best model was the VGG-16 network with 89.06%. Like the work of Das et al., the authors worked with the Inception V3 model, which has some similarities with the Xception model.

Even though the results obtained have been promising, some authors consider that fundus images are not adequate for the pre-diagnosis of retinopathies. Alsaih et al. [21] proposed the use of Convolutional Neural Networks in OCT images. In their proposal, the authors relied on the use of the 8s and VGG-16 networks and in a complementary way with U-Net, Seg-Net, and Deeokavv3+ [22], and they tested their proposal on the OPTIMA [23] and RETOUCH [24] datasets, obtaining a performance of 0.92 concerning Dice's similarity coefficient. However, we consider working with non-invasive methods.

From the literature review, we can summarize that the detection of retinopathies can be performed by image-processing operations and filters, by combining image-processing methods and Machine-Learning Models and using Deep Learning, and by creating dedicated architectures or using pre trained architectures and only adjusting a minimum set of weights (transfer learning).

In this research, we decided to venture into the application of a Convolutional Neural Network architecture for the classification, and hence, detection of AMD. After reviewing

the architectures that have been employed, developed from scratch or by transfer learning, we decided to use the transfer learning method with a slightly modified version of the Xception model since, as will be seen later, it can be adapted to work with datasets comprised of a few images, and it can run in mid- or low-end GPUs.

1.2. Theoretical Background

1.2.1. Deep Learning

It is known that Neural Networks are models that seek to approximate a function f^* , which represents a pattern classifier where a pattern x is mapped to a class y [25].

The term Deep Learning should be understood as a Machine-Learning technique that uses a Deep Neural Network, which is nothing more than a multilayer Neural Network containing two or more hidden layers made up of simple processing elements. Deep Learning has successfully attacked some problems where the backpropagation algorithm did not work properly, particularly in Computer Vision or Natural Language-Processing tasks [14]. Particularly with Deep Learning, three important problems were solved: the vanishing gradient in which the error of the output layer of the neural network does not reach the nodes closest to the input layer using backpropagation so that the training is not performed correctly. This problem was solved by using the activation function Rectified Linear Unit (ReLU) [25].

1.2.2. Convolutional Neural Network (CNN)

The base architecture of the CNN was developed during the 1980s and 1990s. However, it was not very relevant because it was impractical to apply in industry, particularly in analyzing complex images. However, by 2012, the idea was revived and has now achieved outstanding results. It is important to mention that a CNN is not just a Deep Neural Network with many hidden layers; its operation is based on mimicking how the visual cortex of the brain processes and recognizes images [13,22].

Essentially, the image recognition task is a classification problem. It will seek to solve problems such as distinguishing whether there are people or animals in an image, recognizing handwritten digits, and recognizing disease-related conditions. For this task to be successful, the CNN relies on the traditional concept of a multiclass classification Neural Network. If one wanted to perform this recognition with only a traditional Neural Network and the pixels of the images, the results would be quite deficient, and it would most likely be necessary to apply pre-processing techniques to the images. One of the great strengths of CNNs is that they do not require pre-processing or feature extraction methods to be designed; in the CNN training process, the extraction of relevant features will be performed automatically.

CNNs provide better image recognition when the feature extraction layers of the neural network are deeper, but care must be taken not to fall into overfitting or the vanishing gradient. Regarding feature extraction in CNNs, it is important to mention that it is carried through stacks of convolutional and pooling layer pairs. As its name shows, the convolution layer applies a filter on the images. Subsequently, its dimension is reduced by applying a pooling layer; therefore, unlike traditional Neural Networks, the convolution and pooling layers operate in a 2D plane. In summary, a CNN comprises a set of connected layers in which image feature extraction is performed by convolution and pooling operations (the image convolution filters are built in the training process and are used to extract relevant features of each class of image) and classification by a multiclass classification Neural Network [21].

The structure of a CNN is presented in Figure 2.

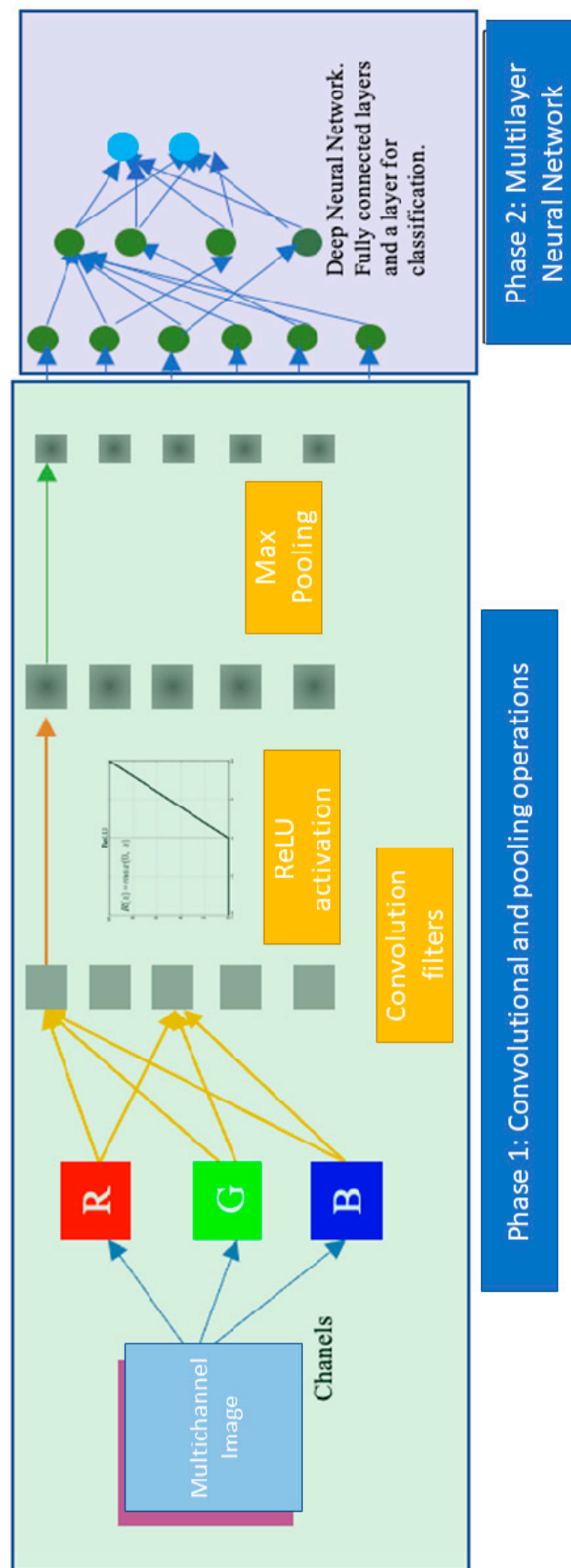


Figure 2. Deep Convolutional Network.

As mentioned above, CNNs make intensive use of convolution and pooling operations. The convolution layers generate new images that are called feature maps. The features that differentiate the different classes in which the images are organized are stressed in these maps. Another important fact is that no connection weights or product sums are handled in this layer; only filters are used to process images, and they are called convolutional filters. It should also be noted that one or more filters can form the convolutional layers, and for each filter, an output feature map will be generated [22,26].

It is important to mention that, besides convolution and pooling, CNNs employ the SoftMax activation function; in multilayer perceptron-based neural networks, the usual activation function was the sigmoid. However, this works adequately for the weighted sum of inputs, but it is not intended to receive the output of other output nodes, while the SoftMax function considers the weighted sum of the inputs and the values of the output nodes, and it also fits multi-class classification problems [22,27].

1.2.3. Transfer Learning

This name identifies a widespread approach that takes advantage of the ability of models to leverage the knowledge gained from solving other classification problems in a new task. However, this learning is achieved by fine-tuning the model using images from a different domain or problem than the one it originally learned. This learning will provide better results if the images to be classified are like the images on which the model originally learned. However, it has been shown that transfer learning can obtain superior performance when using a model with randomly initialized weights [28,29].

Transfer learning guides the initialization of the weights for a new classification task. This guidance leads to two ways of doing transfer learning, either to adjust all the previously initialized weights during the training process or to freeze the original values of the weights of the first layers and only update the weights of the last layers, known as partial adaptation. The choice of the modality will depend on the size of the training set of the target classification problem [29,30].

2. Materials and Methods

2.1. Proposed Solution

In this section, we will discuss the key features of the Deep Neural Network model we chose to classify images into healthy and age-related macular degeneration-affected images.

2.1.1. Xception

The term refers to “Extreme Inception”, which corresponds to another Deep Neural Network architecture called Inception. This model is characterized by replacing the inception modules with depthwise convolutions. Google scientists developed these models [31]. With Xception, a depth-separable convolution is used, which improves the results obtained by Google with the Inception v3 model [32].

In this network model, the depth-separable convolution is made by performing a pointwise convolution, followed by a depth wise convolution. This change of order in the operations concerning the original models leads to the following differences: very similar results are obtained regarding the results obtained in the inception modules and, on the other hand, the non-linearity that existed after the first operation in the inception modules is eliminated. In this 71-layer model (we choose the number of layers of the original model since our efforts to simplify it did not lead to good results), a pointwise convolution is first performed, followed by a depthwise convolution. In addition to the order of operations, this model differs from the Inception network because the intermediate non-linearity between the layers associated with the ReLU function no longer exists because of the modified depthwise separable convolution. The complete architecture of the Xception Deep Neural Network is presented in Figure 3.

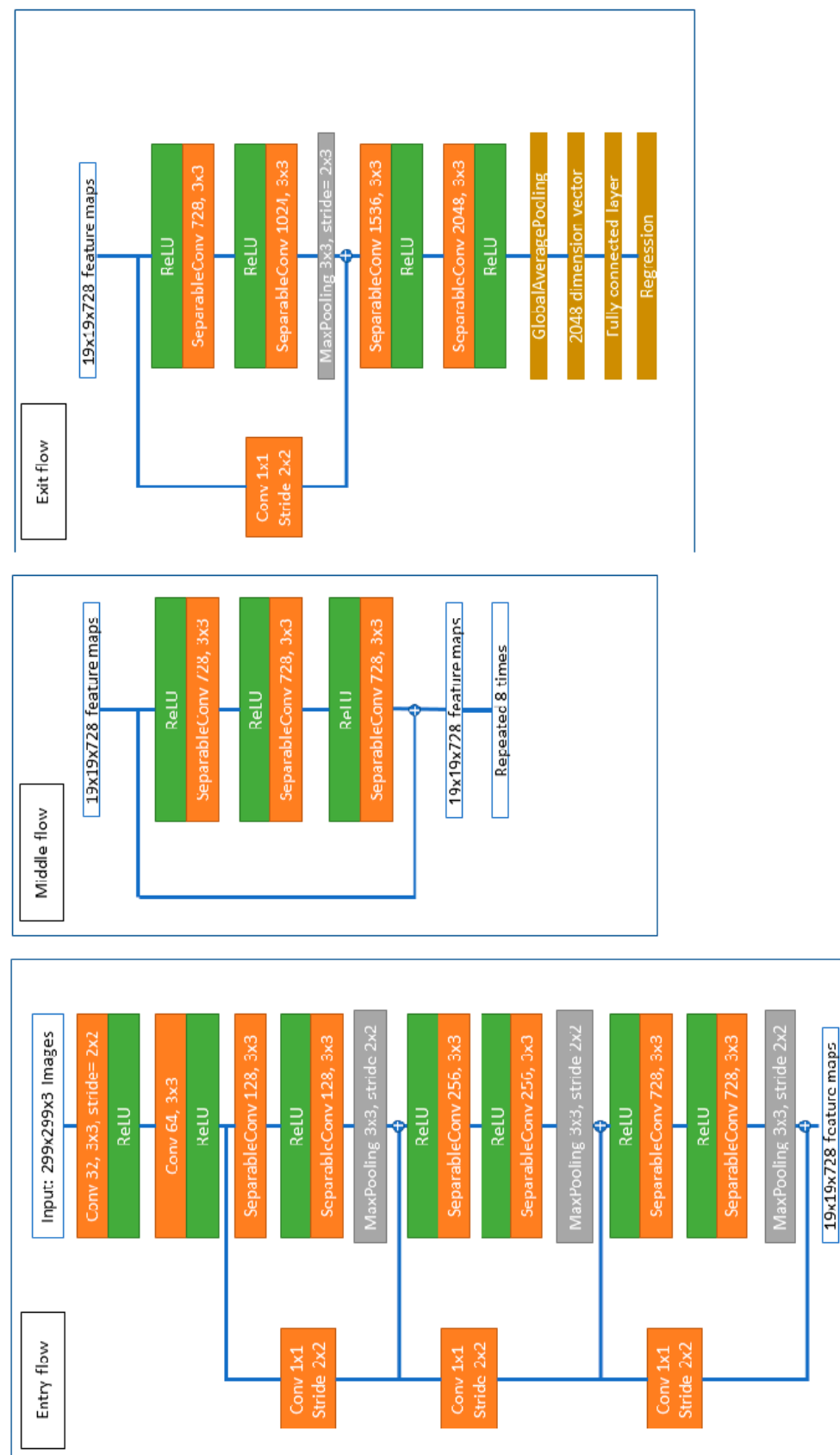


Figure 3. Xception model architecture.

Derived from this architecture, this model, which takes as input 3-channel images, has 22,855,952 parameters adjusted in 71 depth layers. In this work, we propose a series of slight changes to the architecture of the network. Some connections and operations of convolutions 2×1 are removed (only in the levels closer to the fully connected layers).

2.1.2. Parameters for the Xception Neural Network and the Other Models

In this work, we used transfer learning to take advantage of the previously trained weights of a Neural Network model, so we will only adjust the parameters of the fully connected layers of the CNN models. We selected the following parameters, as shown in Table 1.

Table 1. Parameter configuration for the experiments.

Parameter	Value
Learning algorithm	SGDM (Stochastic Gradient Descent with Momentum)
Learning ratio	0.001
Minibatch size	10
Max number of epochs	30
Validation Frequency	50
Folds for K-fold cross-validation	10

It is important to mention that the previous parameters were employed to set up the further processing layers (the connected layer), as we are using the transfer learning method to use the Neural Network model with our data.

3. Results

3.1. Experimental Set-Up

In this section, we describe the experiments prepared to test the performance of the Xception model in the prepared datasets. The idea is that all the images are grouped in two folders named “positive” and “negative”; the names refer to the presence or absence of AMD in the images. Therefore, we are dealing with a bi-class classification problem in which the classes are the images affected by AMD and images of healthy cases with no AMD presence.

3.1.1. Datasets for this Work

This first dataset, called Optical Disease Recognition, which is available on the Kaggle website [33], is composed of ophthalmoscopic color images of the human retina. Experts who assigned one or more of the following diagnoses analyzed these images: normal, diabetic retinopathy, glaucoma, cataract, age-related macular degeneration, hypertensive retinopathy, pathologic myopia, as well as other abnormalities in the human retina. This dataset is heterogeneous since the fundus images were captured in China and with cameras of various makes and models, so the images have heterogeneous characteristics.

The second dataset was created as part of a series of research projects (SAMRH) directed by an expert from the Centro de Investigación en Computación of Mexican Instituto Politécnico Nacional, particularly in the thesis of García-Floriano [4]; a set of 150 images of healthy patients and patients affected by various retinopathies was generated. This set comprises images provided by an ophthalmology specialist and images taken from public repositories such as DRIVE [34].

For the projects mentioned above, the images were used to test the segmentation methods for conditions related to diseases such as diabetic retinopathy, hypertensive retinopathy, macular degeneration, glaucoma, and retinitis pigmentosa; subsequently, they were used to generate feature vectors from invariant moments [35,36]. Unlike what was performed in the works mentioned earlier, only the images were resized to fit the chosen Deep-Learning model in this research.

3.1.2. Dataset Generated for this Work

A dataset was created comprising 180 images from the Ocular Disease Recognition dataset of Kaggle and 70 images from the dataset of the SAMRH-related projects. Of the total of 250 images, 128 images correspond to people affected by age-related macular degeneration and 122 images correspond to those not affected by AMD. The information from this dataset is summarized in Table 2.

Table 2. Training and validation dataset set-up.

Parameter	Value
Overall number of images	250
Images belonging to the Ocular Disease Recognition Dataset.	180
Images belonging to the SAMRH project.	70
Overall number of images of patients affected with AMD.	128
Overall number of images of healthy patients.	122

We used some data augmentation techniques to strengthen the training of the model, particularly the invariance of the rotations, scaling, and translations. Therefore, in training, we considered versions of the training set images with rotation (0° to 180°), scaling (factor up to 1.5), and translation (0 to 200 pixels).

3.1.3. Implementation

The software used in the experiments was developed with the MATLAB R2022 software [37] on a computer with a Core Xeon processor, 96 GB of RAM, and an NVIDIA Quadro P2000 GPU. The operating system of this computer is Microsoft Windows 11 Pro version 23H2.

3.1.4. Cross-Validation Strategy for Training and Validation

Due to the difference in the size of the datasets used to train the Xception model and the dataset created for this work, it was decided to perform the training by employing a cross-validation process. Thus, the experimental process was carried out with two independent datasets: a training and validation dataset composed of 250 images and a test dataset composed of 22 images; image selection was performed randomly. To avoid the problem of overfitting, we employed a cross-validation strategy that avoids bias when training the Xception model. A cross-validation method randomly rearranges the patterns, divides the dataset into several folds or segments, and assigns those segments to the training and validation sets [38]. To avoid bias or overfitting in this work, we chose to apply 10-fold cross-validation to the whole training set so that from the training set, 90% of the data will be randomly taken for model learning, while the remaining 10% will be left for validation; this validation was used to evaluate the model performance and to adjust some of its parameters [29]. Therefore, the process of training was performed with an iterative process, which considered 225 images per training and 25 images per validation; after the training was finished, and the parameters of the connected layers of the model were adjusted, and we performed the testing phase on the set of 22 images that were not considered in the training/validation process.

3.1.5. Performance Metrics for This Work

In medical environments, confusion matrices and performance metrics are generated with four essential values: True Positives, True Negatives, False Positives, and False Negatives. From these elementary values obtained in the confusion matrix, we have the following performance measures.

$$\text{Diagnostic effectiveness or Accuracy (ACC)} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Diagnostic Odds Ratio (DOR)} = \frac{\frac{TP}{FN}}{\frac{FP}{TN}} = \frac{(TN)(TP)}{(FN)(FP)} \quad (2)$$

$$\text{Sensitivity, Recall or True Positive Rate (TPR)} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{Miss rate or False Negative Rate (FNR)} = \frac{FN}{TP + FN} \quad (4)$$

$$\text{Specificity or True Negative Rate (TNR)} = \frac{TN}{TN + FP} \quad (5)$$

$$\text{False Positive Rate (FPR)} = \frac{FP}{TN + FP} \quad (6)$$

$$\text{Precision or Positive Predicted Value (PPV)} = \frac{TP}{TP + FP} \quad (7)$$

$$\text{F1 score} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (8)$$

3.1.6. Deep Neural Network Models for Comparison

In this research, we chose the following Deep Neural Network models to compare with our proposal.

- (1) AlexNet [39]. This model was presented by Krizhevsky et al. at the ImageNet Large-Scale Visual Recognition Challenge 2012. The AlexNet model is composed of eight layers: five convolutional layers and three fully connected layers. The strength of this model is based on the nonlinear properties of the ReLU function, using GPUs, pooling layers overlapping, data augmentation, and a dropout technique. This model represented a first attempt to classify datasets composed of many images.
- (2) SqueezeNet [40]. The model presented by Iandola et al. has also been trained on the ImageNet dataset. The idea of this model is based on working with smaller Neural Networks to reduce communication requirements in cloud servers and adapt them to dedicated hardware. To achieve this goal, this model applies strategies such as using 1×1 filters instead of 3×3 filters. Squeeze layers are also used to decrease the number of input channels and the size of feature maps in the last layers of Neural Network processing. According to its authors, it is a Convolutional Network architecture that performs better than AlexNet, with a ratio up to 50 times lower in terms of the parameters.
- (3) Inception V3 [32]. This architecture was developed by Google and presented in the article by Szegedy (2016). This architecture evolves the concept presented in Inception V1 and Inception V2, which includes batch normalization. In this architecture, the concept of factorization is added, which seeks to reduce the number of connections and parameters without affecting the model's performance.
- (4) GoogLeNet [41,42]. This model, presented by Szegedy et al., defines an architecture that comprises the so-called inception modules. An inception module consists of a few convolution kernels (1×1 , 3×3 , and 5×5), limiting the model parameter number and complexity. The GoogLeNet network consists of 27 layers, of which 9 correspond to inception modules; these modules are heavy to perform feature detection at different scales through convolutions with additional filters. One of the significant advantages of this model is that they prevent overfitting and the vanishing descent gradient through a block called an auxiliary classifier. The extra classifier comprises an average pool layer, a convolutional layer, two fully connected layers, a dropout layer, and a layer with the SoftMax activation function.
- (5) Visual Geometry Group (VGG) [43,44]. This type of network stands out in the state-of-the-art examples of CNNs because it is the basis of quite popular frameworks in terms of object detection such as R-CNN or Single Shot Detection. VGG Neural Networks seek to optimize image classification by varying the size of the filters in the

first convolutional layer and by changing the depth of the Neural Network (that is why the VGG16 and VGG19 models are available). To avoid high consumption of computational resources, it implements strategies such as using 3×3 filters, ReLU activation functions, or at the end of the classification, employ t3 layers fully connected, of which 2 have 4096 neurons each, and the last have as many neurons as classes have the classification problem [45].

- (6) Xception [31]. This model represents a new CNN architecture that adds new layers of the inception type. These new layers are generated from depth convolutional layers, followed by a convolution operation. A key feature is that the convolutions are separable by depth. This model receives images of 300×300 pixels, and after all the processing layers, feature maps of $10 \times 10 \times 2048$ features are obtained [46]. Therefore, unlike the other models that have been considered in this work, Xception bases its operation on the operations of pointwise and depthwise convolution. It is also important to note that Xception has several processing layers, while other models seek to reduce or variably handle the number of processing layers.

3.2. Experimental Results of the Compared Models

To test the performance of the Xception model, we performed a series of experiments, considering the cross-validation strategy in Section 3.1.4 for the training set (which is split into training and validation).

Our proposed model has the following behavior regarding the training and validation process. In Figures 4 and 5, we have charts of the accuracy and loss of the process.

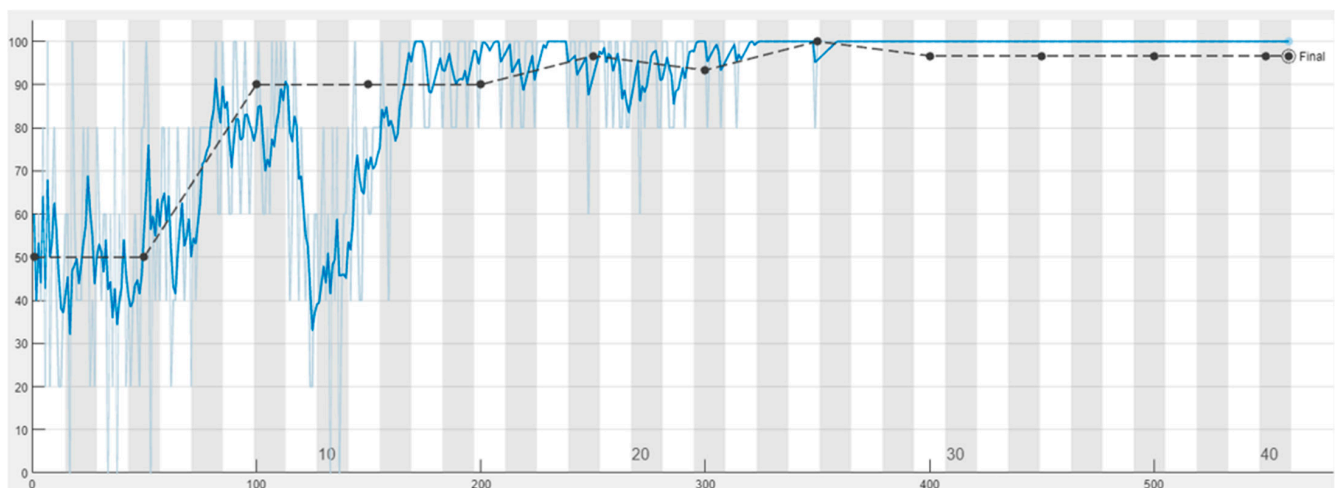


Figure 4. Plot of the accuracy (x-axis: iteration, y-axis: accuracy) of the training process of the modified model (the black dotted line corresponds to validation).

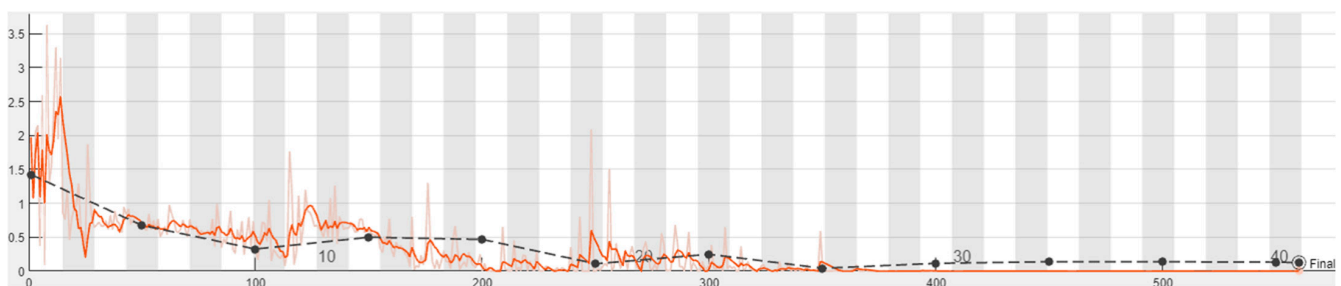


Figure 5. Plot of the loss (x-axis iteration, y-axis loss) of the modified model (the black dotted line corresponds to the validation step).

There are a few differences between the original version, the Xception original architecture, and the version modified in this work. The training and loss charts of the Xception model can be seen in Figures 6 and 7.

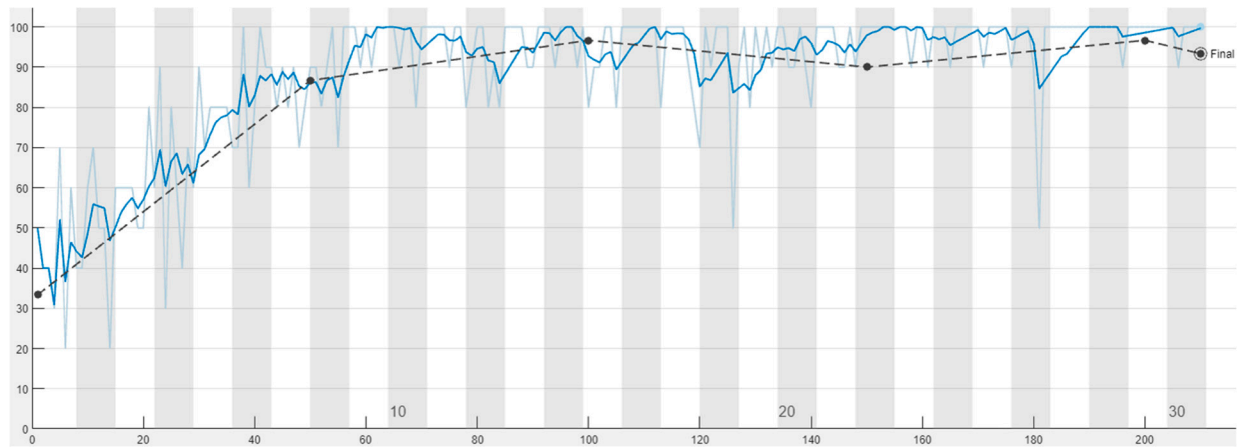


Figure 6. Plot of the accuracy (x-axis: iteration, y-axis: accuracy) of the original model (the black dotted line corresponds to the validation step).

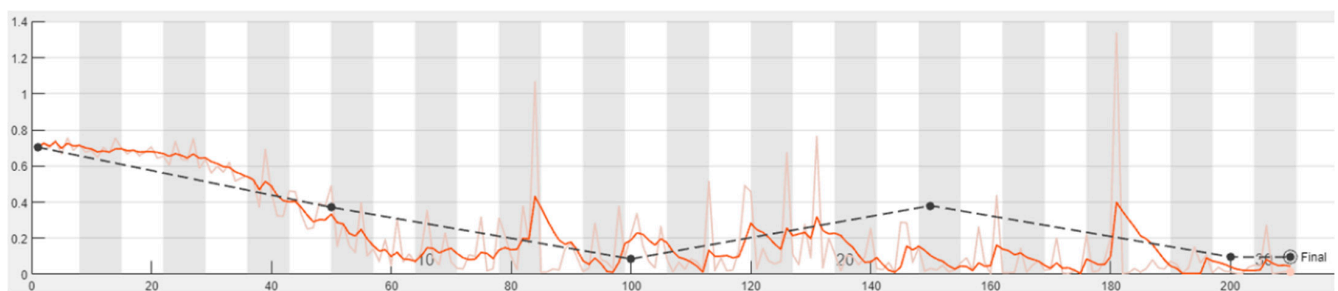


Figure 7. Plot of the loss function (x-axis iteration, y-axis loss) of the original model (the black dotted line corresponds to the validation step).

Both pairs of charts show a similar behavior, although our proposal seems to take more time to converge regarding the accuracy and the loss in the training phase.

Regarding the efficiency of computational resources, we show in Table 3 the employed resources.

Table 3. Computational resources consumption.

Resource	Consumption
GPU memory	1 GB
GPU % of activity	50%
CPU % of activity	30%
Total cores employed	12/12
RAM required by the MATLAB tool	4 GB

Once the Neural Network was configured with the parameters in Table 2 and the dataset was prepared (taken the cross-validation proposal in Section 3.1.4), we performed a comparison between the proposed architecture and the Deep Neural Network models, as shown in Section 3.1.6, which are transfer-learning trained.

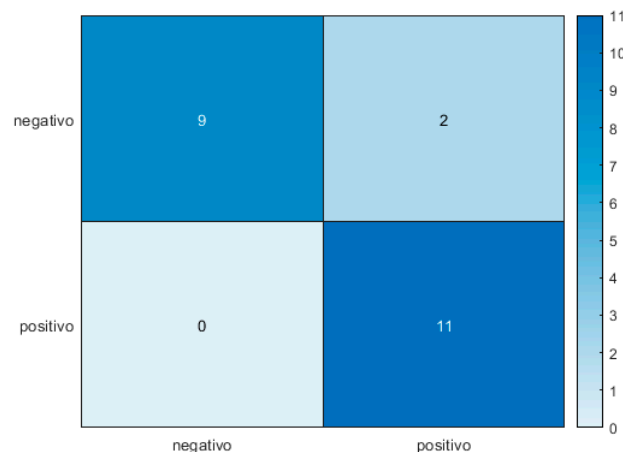
These models were tested with the dataset presented in Section 3.1.1, and the results are presented in Table 4.

Table 4. Accuracy results for the compared models.

Model	Training Accuracy	Validation Accuracy	Validation Loss	Validation Error (95% Confidence Interval)	Execution Time
AlexNet	0.65	0.459	0.6	0.541 ± 0.29	2980 s
SqueezeNet	0.7	0.566	1.0	0.434 ± 0.26	3160 s
InceptionV3	0.95	0.90	0.3	0.10 ± 0.17	7840 s
GoogLeNet	0.9	0.87	0.2	0.13 ± 0.15	3050 s
VGG16	0.7	0.62	1.0	0.38 ± 0.26	3900 s
VGG19	0.7	0.61	1.0	0.39 ± 0.26	3330 s
Xception original	0.91	0.9	0.27	0.08 ± 0.07	5860 s
Xception mod	0.95	0.92	0.25	0.08 ± 0.07	5730 s

As can be seen, the Xception model obtained the best results in terms of classification accuracy, was slightly better than InceptionV3 and was quite better than models like AlexNet; however, regarding the execution time, the Xception model took the second longest time to complete the training and validation of the fundus images but offered the best classification accuracy.

From comparing the validation accuracy, we have determined that the Xception model obtained the best results, so we have tested this model with the test set we created from the designed dataset. This set comprises 22 images, of which 11 belong to the class of images of patients with AMD (positive for) and 11 belong to the class of images of patients who are not ill with AMD (negative). In Figure 8, we present the confusion matrix obtained.

**Figure 8.** Confusion matrix with the original Xception model.

The results of the different performance metrics for the Xception model are shown in Table 5.

Table 5. Results of the Xception models with the test set.

Performance Measure	Xception Model
ACC	0.91
DOR	Infinite
TPR	1
FNR	0
TNR	0.81
FPR	0.18
PPV	0.84
AUC of ROC	0.91
F1 Score	0.91

4. Discussion

According to the results obtained, the Xception model can detect all the positive cases of AMD, which implies that this Neural Network learned the distinctive features of AMD. However, the presence of false positives indicates that some of the images that do not present symptoms of the disease are classified as positive for AMD, which may be due to the presence of light objects or excess brightness in these images.

Because we obtained zero false negatives in the experiment performed with the test set, the DOR value is infinite and unfortunately cannot be considered a valid outcome measure. If we also had zero false positives, the test result would be perfect.

It is essential to mention that among the various existing Convolutional Neural Network architectures, Xception offers better performance than models such as VGG, ResNet, or Inception, mainly in terms of the error or loss function. That is primarily due to the use of the depthwise separable convolution and residual connections.

Although we tried to provide a modified version of the Xception Neural Network (by making adjustments to the connections of the different layers), we could not make significant changes since we found that affecting the number of layers or connections can significantly affect the performance of the Xception Neural Network.

In this work, we used transfer learning, which is essentially based on the use of a Convolutional Neural Network architecture previously trained and validated with a dataset that is not necessarily related to our problem of interest. In this work, we employed the Xception model for the pre-diagnosis of age-related macular degeneration by classifying fundus images as healthy or with signs of AMD. After a series of experiments, we found that the deep network called Xception provided the best results with a dataset created for this research work. It is worth noting that some modifications were even made to the architecture without a major impact on the results. Therefore, we consider that the main contribution of this work is the use of transfer learning, specifically with the Xception model in the detection of AMD in fundus image datasets, where the number of images is not very large.

5. Conclusions

The automated pre-diagnosis of various retinopathies or, as in this work, of age-related macular degeneration is an open research problem for which many approaches and alternative solutions have been proposed. However, so far, no optimal model or method for its solution has been obtained.

In recent years, the term Deep Learning has gained relevance, which refers to the use of Deep Neural Networks, which are essentially Neural Networks with two or more hidden layers. Among the great variety of models that have emerged within this concept, a series of networks based on an operation typically used in digital image processing stand out: convolution. The so-called convolutional networks have been proposed and successfully used in classification and object detection tasks, minimizing the number of pre-processing operations usually applied to images before transforming them to introduce them into a classification model or using them directly.

This work constitutes a first approach to using Deep-Learning models for the pre-diagnosis of macular degeneration. We successfully applied transfer learning, with the Xception model, to classify fundus images in positive or negative cases of AMD; the classification has been made based on the presence of drusen in fundus images with the presence of AMD.

As future work, we consider it essential to expand the size of the database to include a larger sample of images. On the other hand, we consider that the number of models included in the experiments could be expanded since new architectures are constantly emerging. We need to improve the quality of the analysis to avoid the detection of bright objects as drusen in fundus images. Finally, we should consider our approach to classifying fundus retinal images into healthy and those affected by macular degeneration and from the results classify the grade of severity of AMD or the presence of other diseases.

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

Funding: This research received no external funding.

Data Availability Statement: The dataset used to test the proposed method is available upon request to the email agarciafl@ipn.mx. Some images were taken from the Ocular Disease Recognition, which is stored on Kaggle website <https://www.kaggle.com/andrewmvd/ocular-disease-recognition-odir5k> (accessed on 6 November 2023).

Acknowledgments: The authors are grateful for the support received from the Mexican Instituto Politécnico Nacional (IPN), through Centro de Innovación y Desarrollo Tecnológico en Cómputo (CIDETEC), Escuela Superior de Cómputo (ESCOM), and the Secretaría de Investigación y Posgrado (SIP), for their support with the computing and communications infrastructure. We also thank the Consejo Nacional de Ciencia y Tecnología (CONACYT) of the Government of Mexico for the support received through the Sistema Nacional de Investigadores (SNI) stimulus.

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

References

- Mittal, D.; Kumari, K. Automated detection, and segmentation of drusen in retinal fundus images. *Comput. Electr. Eng.* **2015**, *47*, 82–95. [\[CrossRef\]](#)
- Boyd, K. What Is Macular Degeneration? Available online: <https://www.aoa.org/eye-health/diseases/amd-macular-degeneration> (accessed on 8 November 2023).
- Iwama, D.; Hangai, M.; Ooto, S.; Sakamoto, A.; Nakanishi, H.; Fujimura, T.; Domalpally, A.; Danis, R.P.; Yoshimura, N. Automated assessment of drusen using three-dimensional spectral-domain optical coherence tomography. *Investig. Ophthalmol. Vis. Sci.* **2012**, *53*, 1576–1583. [\[CrossRef\]](#) [\[PubMed\]](#)
- García Floriano, A. *Sistema Integral de Análisis Para la Prevención de Ceguera*; Instituto Politécnico Nacional, Centro de Investigación en Computación: Mexico City, Mexico, 2011.
- Farnel, D.J. Automated Retinal Image Analysis (ARIA) Data Set. Available online: http://www.damianjffarnell.com/?page_id=276 (accessed on 14 July 2021).
- Goldbaum, M. The STARE Project. Available online: <https://cecas.clemson.edu/~ahoover/stare/> (accessed on 14 July 2021).
- Mvoulana, A.; Kachouri, R.; Akil, M. Fully automated method for glaucoma screening using robust optic nerve head detection and unsupervised segmentation-based cup-to-disc ratio computation in retinal fundus images. *Comput. Med. Imaging Graph.* **2019**, *77*, 101643. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*, 3rd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2007; ISBN 013168728X.
- Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2015; Volume 9351, pp. 234–241.
- Rundo, L.; Han, C.; Zhang, J.; Hataya, R.; Nagano, Y.; Militello, C.; Ferretti, C.; Nobile, M.S.; Tangherloni, A.; Gilardi, M.C.; et al. CNN-based Prostate Zonal Segmentation on T2-weighted MR Images: A Cross-dataset Study. *Smart Innov. Syst. Technol.* **2019**, *151*, 269–280.
- Koh, J.E.W.; Acharya, U.R.; Hagiwara, Y.; Raghavendra, U.; Tan, J.H.; Sree, S.V.; Bhandary, S.V.; Rao, A.K.; Sivaprasad, S.; Chua, K.C.; et al. Diagnosis of retinal health in digital fundus images using continuous wavelet transform (CWT) and entropies. *Comput. Biol. Med.* **2017**, *84*, 89–97. [\[CrossRef\]](#) [\[PubMed\]](#)
- Vijayan, T.; Sangeetha, M.; Kumaravel, A.; Karthik, B. Gabor filter and machine learning based diabetic retinopathy analysis and detection. *Microprocess. Microsyst.* **2020**, 103353. [\[CrossRef\]](#)
- LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- Liu, Y.P.; Li, Z.; Xu, C.; Li, J.; Liang, R. Referable diabetic retinopathy identification from eye fundus images with weighted path for convolutional neural network. *Artif. Intell. Med.* **2019**, *99*, 101694. [\[CrossRef\]](#)
- Hoover, A.; Goldbaum, M. Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels. *IEEE Trans. Med. Imaging* **2003**, *22*, 951–958. [\[CrossRef\]](#)
- Peng, J.; Kang, S.; Ning, Z.; Deng, H.; Shen, J.; Xu, Y.; Zhang, J.; Zhao, W.; Li, X.; Gong, W.; et al. Residual convolutional neural network for predicting response of transarterial chemoembolization in hepatocellular carcinoma from CT imaging. *Eur. Radiol.* **2020**, *30*, 413–424. [\[CrossRef\]](#) [\[PubMed\]](#)

18. Sun, Z.; Shi, Z.; Xin, Y.; Zhao, S.; Jiang, H.; Wang, D.; Zhang, L.; Wang, Z.; Dai, Y.; Jiang, H. Artificial Intelligent Multi-Modal Point-of-Care System for Predicting Response of Transarterial Chemoembolization in Hepatocellular Carcinoma. *Front. Bioeng. Biotechnol.* **2021**, *9*, 761548. [CrossRef] [PubMed]
19. Das, A.; Rad, P.; Choo, K.K.R.; Nouhi, B.; Lish, J.; Martel, J. Distributed machine learning cloud teleophthalmology IoT for predicting AMD disease progression. *Futur. Gener. Comput. Syst.* **2019**, *93*, 486–498. [CrossRef]
20. Gour, N.; Khanna, P. Multi-class multi-label ophthalmological disease detection using transfer learning based convolutional neural network. *Biomed. Signal Process. Control* **2020**, *66*, 102329. [CrossRef]
21. Alsaih, K.; Yusoff, M.Z.; Tang, T.B.; Faye, I.; Mériaudeau, F. Deep learning architectures analysis for age-related macular degeneration segmentation on optical coherence tomography scans. *Comput. Methods Programs Biomed.* **2020**, *195*, 105566. [CrossRef] [PubMed]
22. Kim, P. *MATLAB Deep Learning*; Apress: New York, NY, USA, 2017.
23. Hrvoje, B. Optima Challenges. Available online: <https://optima.meduniwien.ac.at/research/challenges/> (accessed on 14 July 2021).
24. Bogunovic, H.; Venhuizen, F.; Klimscha, S.; Apostolopoulos, S.; Bab-Hadiashar, A.; Bagci, U.; Beg, M.F.; Bekalo, L.; Chen, Q.; Ciller, C.; et al. RETOUCH: The Retinal OCT Fluid Detection and Segmentation Benchmark and Challenge. *IEEE Trans. Med. Imaging* **2019**, *38*, 1858–1874. [CrossRef] [PubMed]
25. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
26. Tan, J.H.; Bhandary, S.V.; Sivaprasad, S.; Hagiwara, Y.; Bagchi, A.; Raghavendra, U.; Krishna Rao, A.; Raju, B.; Shetty, N.S.; Gertych, A.; et al. Age-related Macular Degeneration detection using deep convolutional neural network. *Futur. Gener. Comput. Syst.* **2018**, *87*, 127–135. [CrossRef]
27. Sunija, A.P.; Kar, S.; Gayathri, S.; Gopi, V.P.; Palanisamy, P. OctNET: A Lightweight CNN for Retinal Disease Classification from Optical Coherence Tomography Images. *Comput. Methods Programs Biomed.* **2020**, *200*, 105877. [CrossRef] [PubMed]
28. Chu, X. Deep Learning. In *Encyclopedia of Big Data Technologies*; Springer International Publishing: Cham, Switzerland, 2019; pp. 639–648.
29. Marín, R.; Chang, V. Impact of transfer learning for human sperm segmentation using deep learning. *Comput. Biol. Med.* **2021**, *136*, 104687. [CrossRef]
30. Tian, Y.; Fu, S. A descriptive framework for the field of deep learning applications in medical images. *Knowl.-Based Syst.* **2020**, *210*, 106445. [CrossRef]
31. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
32. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; IEEE Computer Society: Washington, DC, USA, 2016; Volume 2016, pp. 2818–2826.
33. Larxel, U. Ocular Disease Recognition | Kaggle. Available online: <https://www.kaggle.com/andrewmvd/ocular-disease-recognition-odir5k> (accessed on 14 July 2021).
34. van Ginneken, B.; Kerkstra, S.; Meakin, J. Introduction—DRIVE—Grand Challenge. Available online: <https://drive.grand-challenge.org/> (accessed on 14 July 2021).
35. García-Florian, A.; Ferreira-Santiago, Á.; Camacho-Nieto, O.; Yáñez-Márquez, C. A machine learning approach to medical image classification: Detecting age-related macular degeneration in fundus images. *Comput. Electr. Eng.* **2017**, *75*, 218–229. [CrossRef]
36. Garcia Florian, A.; Yanez Marquez, C.; Camacho Nieto, O. Detection of Age-Related Macular Degeneration in Fundus Images by an Associative Classifier. *IEEE Lat. Am. Trans.* **2018**, *16*, 933–939. [CrossRef]
37. Mathworks MATLAB—El Lenguaje del Cálculo Técnico—MATLAB & Simulink. Available online: <https://la.mathworks.com/products/matlab.html> (accessed on 14 July 2021).
38. Wong, T.-T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognit.* **2015**, *48*, 2839–2846. [CrossRef]
39. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the 26th Annual Conference on Neural Information Processing Systems 2012, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
40. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. *arXiv* **2016**, arXiv:1602.07360.
41. 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 AAAI’17: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 4278–4284.
42. Tang, P.; Wang, H.; Kwong, S. G-MS2F: GoogLeNet based multi-stage feature fusion of deep CNN for scene recognition. *Neurocomputing* **2017**, *225*, 188–197. [CrossRef]
43. 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 (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.

44. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015), San Diego, CA, USA, 7–9 May 2015.
45. Wei, J.; Ibrahim, Y.; Qian, S.; Wang, H.; Liu, G.; Yu, Q.; Qian, R.; Shi, J. Analyzing the impact of soft errors in VGG networks implemented on GPUs. *Microelectron. Reliab.* **2020**, *110*, 113648. [[CrossRef](#)]
46. Rahimzadeh, M.; Attar, A. A new modified deep convolutional neural network for detecting covid-19 from X-ray images. *arXiv* **2020**, arXiv:2004.08052.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.