Acta Polytechnica **64**(1):1–11, 2024
https://doi.org/10.14311/AP.2024.64.0001

Published by the Czech Technical University in Prague

THE CLASSIFICATION OF EYE DISEASES FROM FUNDUS IMAGES BASED ON CNN AND PRETRAINED MODELS

Samir Benbakreti[a,*], Soumia benbakreti[b], Umut Ozkaya[c]

[a] *National High School of Telecommunications and ICT (ENSTTIC), Department of specialty, Oran 31 000, Algeria*

[b] *University of Djillali Liabes, Laboratory of Mathematic, BP 89, Sidi Bel Abbes 22000, Algeria*

[c] *Konya Technical University, Engineering and Natural Science Faculty, Electrical and Electronics Engineering, Konya, Turkey*

[*] corresponding author: samir.benbakreti@ensttic.dz

Abstract. Visual impairment affects more than a billion people worldwide due to insufficient care or inadequate vision screening. Computer-aided diagnosis using deep neural networks is a promising approach, it can analyse and process retinal fundus images, providing valuable reference data for doctors in clinical diagnosis or screening. This study aims to achieve an accurate classification of fundus images, including images of healthy patients as well as those with diabetic retinopathy, cataracts, and glaucoma, using a convolutional neural network (CNN) architecture and several pretrained models (AlexNet, GoogleNet, ResNet18, ResNet50, YOLOv3, and VGG 19). To enhance the training process, a mirror effect technique was applied to augment the volume of data. The experimental study resulted in very satisfactory outcomes, with the GoogleNet model paired with the SGDM optimiser achieving the highest accuracy (92.7 %).

Keywords: Eye diseases classification, retinal fundus images, deep learning, pretrained models, SGDM.

1. Introduction

One of the most crucial senses for humans is the vision, which allows people to comprehend their surroundings and communicate information, such as knowledge and memories. The tear film barrier, corneal barrier, conjunctival barrier, sclera barrier, and blood-ocular barriers are only a few of the numerous barriers that make up the intricate eye anatomy [1]. These barriers can shield the eye from foreign objects, but they also restrict the amount of medication that can reach the eye. More and more people worldwide have developed ocular diseases in recent years, which can harm eye health and even result in blindness. The World Health Organization (WHO) estimates that 2.2 billion people worldwide are blind or have some form of vision loss, and at least 1 billion of them have a vision impairment that might have been treated [2]. A survey found that 36 million of the 7.33 billion people on the planet were blind in 2015. This number is predicted to increase to 115 million people by 2050 [3]. Generally, uncorrected refractive errors are the predominant cause of severe and moderate visual impairments; in middle- and low-income nations, cataracts continue to be the leading cause of blindness. For example, over 110 million people globally were affected by cataracts in 2010 and by 2050, the number of cataracts is estimated to double [4–6]. Age, diabetes, smoking, low socioeconomic condition, ocular trauma, steroid use, exposure to ultraviolet-B radiation and hereditary factors, are the risk factors for the development of the cataracts [7]. The second most common cause of blindness worldwide is glaucoma. By 2020, the WHO predicts that glaucoma will affect about 80 million people, and its prevalence is anticipated to rise. More than 4.5 million people have been estimated to have essential glaucoma, accounting for more than 12 % of all visual impairment globally [8]. More than 347 million people worldwide have diabetes, and by 2030 it will be the seventh most common cause of death. Diabetes puts a person at risk for retinal abnormalities known as diabetic retinopathy (DR), which is the third leading cause of blindness worldwide. In millimeters of mercury, the ocular pressure is calculated. The usual range for ocular pressure is 12–22 mm Hg, and anything higher than 22 mm Hg is regarded as abnormally high. The presence of diabetic retinopathy in the eye is due to the rising intraocular pressure. Diabetic retinopathy cannot be cured, however, treatment can halt its progression, this highlights the significance of early disease diagnosis. The WHO estimates that DR accounts for 4.8 % of 37 million cases of visual impairment brought on by eye illnesses globally [9]. For example, Algeria had 1800 ophthalmologists for its 40 million citizens in 2017 [10]; number has to rise by 1 000 more ophthalmologists to meet worldwide standards. With 80 specialists graduating from the university each year, this target will not be reached for another 20 years. Therefore, the contribution of artificial intelligence tools to help medical personnel to diagnose eye diseases becomes inevitable. Medical professionals can diagnose diseases earlier with the aid of deep learning (DL), an important component of artificial intelligence (AI). The findings of medical test

results and domain knowledge are combined by a vast variety of AI-based disease detection and classification systems [11, 12]. Recent research has shown how AI can be used to classify lung tuberculosis, identify benign or malignant melanoma, diagnose COVID 19 using a chest X-ray, and detect the progression of retinal illness [13–16]. Ophthalmic disorders are typically not life-threatening, but they can have a major impact on the patient's quality of life as they advance over time. Ophthalmological tools are used to conduct physical examinations, and complete interpretation is employed to make diagnoses. In order to make predictions, any machine-based solution must concurrently take into account observations, symptoms, and results from standardised tests. Implementing DL in eye disease screening and turning this research into clinical screening still presents a significant challenge. The objective of this paper is to provide a tool to medical personnel to help them quickly diagnose eye diseases in patients. Through an analysis of the patient's fundus mages, deep learning and computer vision algorithms can be used to recognise three different diseases (DR, cataracts and glaucoma). This will reduce the workload of already overworked doctors and facilitate the treatment of eye diseases diagnosed at an early stage. The structure of the present document is as follows: The introduction is presented in Section 1, and research work on the early identification of various eye diseases is discussed in Section 2. The dataset and the proposed model are described in Sections 3 and 4. Section 5 details the experiments performed on the different architectures and the conclusions that were reached.

2. Related work

The authors of this paper [17], suggest an automatic classification approach that takes into account two scenarios: mild multi-class diabetic retinopathy (DED) and multi-class DED. The model has been tested on many datasets. Two of ImageNet's top pre-trained convolutional neural network (CNN) models were used in the experiment. Various performance-improving approaches, such as fine-tuning, optimisation, and contrast enhancement, were also applied. On the VGG16 model, the maximum accuracy for multi-class classification was 88.3 %, and for mild multi-class classification, it was 85.95 %. For a multiclass classification system using spectral domain optical coherence tomography (SD-OCT), a new automated convolutional neural network (CNN) model has been suggested in this research [18]. Along with normal cases, the approach divides retinal disease into five categories: age-related macular degeneration (AMD), choroidal neovascularization (CNV), drusen and diabetic macular edema (DME). Globally, the proposed CNN architecture with a softmax classifier has produced excellent results. To illustrate how a new deep learning system can be used to identify corneal abnormalities, a new hierarchical deep learning network representing various degrees of eye disorders, generated from a predetermined hierarchical taxonomy of eye diseases was developed by [19]. A family of multi-task, multi-label learning classifiers comprise this system. The proposed technique was directly trained end-to-end using a retrospective dataset of 5325 ocular surface images. With 510 newly enrolled outpatients who had infectious keratitis, non-infectious keratitis, dystrophy, or corneal degeneration as well as a corneal tumor, the algorithm's performance was then compared with that of ten ophthalmologists. The confusion matrices showed similarities in the misclassifications made by the computer and human experts. The main objective of the work in [20] was to create and assess a deep learning model to assess the severity of AMD at the patient level using the AREDS simplified severity scale using fundus images of both eyes. The images were taken from one of the largest datasets currently available, the AREDS dataset, which has approximately 60,000 retinal scans in it. The model simulates the human grading procedure by first identifying certain risk indicators (drusen and pigment abnormalities) in each eye. An ophthalmologist can examine the patient and see a clear result since it combines the numbers from two eyes to give the patient an AMD score that perfectly matches the clinical decision-making process. Alam et al. [21] used a supervised machine learning to achieve a multitasking OCTA classification. At first, it was required to distinguish between healthy eyes and eyes with disorders, between many eye diseases, and between the severity of each eye condition. The foveal avascular area (FAZ), blood vascular caliber (BVC), vascular perimeter index (VPI), blood vessel density (BVD), and FAZ contour irregularity (FAZ-CI) were all quantitative OCTA characteristics that were entirely automatically extracted from the OCTA images. Sensitive OCTA features and the best feature combinations for multitasking categorisation were found using a backward phase-out method. The supervised machine learning classifier was validated using diabetic retinopathy (DR) and sickle cell retinopathy (SCR) for the proof-of-concept demonstration.

In order to automatically categorise the severity levels of nuclear cataract (NC), the authors in [22] suggest a new convolutional neural network (CNN) framework called Adaptive Feature Squeeze Network (AFSNet). The model is constructed utilising an adaptive feature compression module, which includes a compression block and an update operation in global adaptive commonality, to dynamically compress the local feature representations and update the relative importance of the global feature representations. The AS-OCT clinical picture dataset is used, and the outcomes show the technique's efficacy.

A new convolutional and recurrent hybrid neural network (CRNN) for cataract classification based on fundus images is presented in this research [23]. The suggested CRNN keeps the long-term and short-term spatial correlation between patches by combining the

benefits of the convolutional neural network and recurrent neural network. We use AlexNet, GoogLeNet, ResNet, and VGGNet in conjunction with transfer learning to extract multilevel feature representation and assess how well these models classify cataracts. With an average accuracy of 0.97 % for classifying cataracts into four classifications, the data show that the proposed method performs better than state-of-the-art methods, showing its potential for use in diagnosing other retinal diseases.

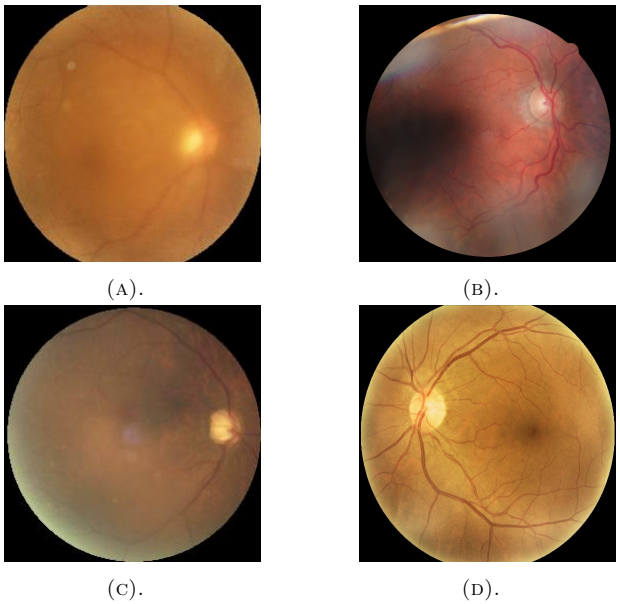
A study [24] suggests a method using the Deep Belief Network (DBN) and Elephant Husbandry Optimisation (EHO) algorithm to detect glaucoma from fundus images. The image is first pre-processed, then the optic disc (OD) and optic section (OC) are segmented, and finally structural, intensity, and texture features are extracted. The ReliefF algorithm is then used to choose the majority of the distinguishing characteristics before sending them to the DBN for glaucoma or normal classification. The EHO method enhances the DBN's parameters in order to increase its classification rate. The model was evaluated using 7 280 images from public and private datasets, and it was able to classify the images with an accuracy of up to 98.5 %.

Because each person's disc optic and optic cup have distinct features (form, size, etc.), the application of cup to disc (CDR) and disc damage probability scale (DDLS) in the detection of Glaucoma is extremely difficult. Prananda et al. [25] suggested a different method to detect glaucoma disease by examining the damage to the retinal nerve fiber layer to get around this issue (RNFL). The pretreatment procedure (removal of optic discs and blood vessels) and the glaucoma categorization procedure are the two steps of the suggested method. The author applied nine deep learning architectures for the latter. On the ORIGA dataset, the highest rate was found to be 92.88 % with an AUC of 89.34 %.

A machine learning method (MLS) to detect AMD using retinal fundus images is put forth in this paper [26]. The several MLS phases that were applied in the proposed works include feature extraction and selection using a statistical test, implementation and validation of the classifier, and image threshold dependent on Shannon entropy and the bat algorithm (BA+ SE). Based on image properties like the grey level co-occurrence matrix and entropies, the retinal image dataset is divided into two different labels, AMD and Non-AMD. A comparison of the effectiveness of various classifiers, including Naive-Bayes (NB), Decision Tree (DT), K-Nearest Neighbors, Random Forest (RF), and Support-Vector-Linear Kernel Machine (SVM), reveals that SVM has a greater classification **accuracy (>93 %) than the other classifiers.** The idea of [27] was to use fundus autofluorescence (FAF) images to automatically classify different inherited retinal disorders (IRDs) using a deep learning system. FAF images of individuals with retinitis pigmentosa (RP), Best's disease (BD), and Stargardt's disease (STGD), as well as a healthy similar group, were used to train a neural network multilayer deep convolution (CNN) based on 389 FAF images. Adam optimiser and data augmentation techniques were used for training. The results of testing the generated classifiers revealed an overall accuracy of 0.95.

3. Description of the dataset and data augmentation

The dataset contains images of the retina affected by diabetic retinopathy, cataract, and glaucoma pathologies; in addition to healthy patients, each class has roughly 1 000 images. The images represent Right and Left eye fundus photographs showing abnormalities due to various ophthalmological diseases (see Figure 1). The dataset employed integrates images from diverse reference datasets, encompassing IDRID, Ocular Recognition, HRF, Retinal Dataset and DRIVE. The objective in merging samples from each dataset is to enrich data diversity, allowing the model to effectively generalise across various scenarios in classification tasks, adapt to diverse conditions, and ensure a balanced representation of classes, thus mitigating training data imbalance. The dataset is available for free online on kaggle: (`https://www.kaggle.com/datasets/\gunavenkatdoddi/eye-diseases-classification`).



(A). (B). (C). (D).

Figure 1. (A) Cataract; (B) Diabetic retinopathy; (C) Glaucoma; (D) Normal eye fundus.

Details about the images from the used datasets are shown in Table 1. Samples of the fundus images are shown in Figure 1.

The training was carried out using Matlab 2021 installed on a workstation with Windows 10 Pro, a

Diseases	Number of images
Cataract	1 038
Diabetic retinopathy	1 098
Glaucoma	1 007
Normal eye	1 074
Total images	4 217

TABLE 1. Description of the dataset.

64-bit operating system, 24 GB of Random Access Memory (RAM), Intel(R) Xeon(R) CPU E5-2620 v3 @ 2.40 GHz and Graphics Processing Unit (GPU). 20 % of the datasets is used for testing, while 80 % is used for training.

Additionally, we used a reflection-based data augmentation technique (Mirror effect) for the sake of preventing overfitting. In fact, by using the reflection effect and keeping the labels, we increased the quantity of images intended for training. The two procedures are as follows:

- Vertical mirror image: The original x and y coordinates of every pixel in the image are used to flip each image vertically. Equation (1) illustrates the new coordinates of each pixel after the reflection over the X-axis as x' and y':

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix}. \tag{1}$$

- Horizontal mirror image: Where x and y represent the original coordinates of each pixel in the image, each image is horizontally inverted. Then, as stated in Equation (2), x' and y' are each pixel's new coordinates post reflection over the Y-axis:

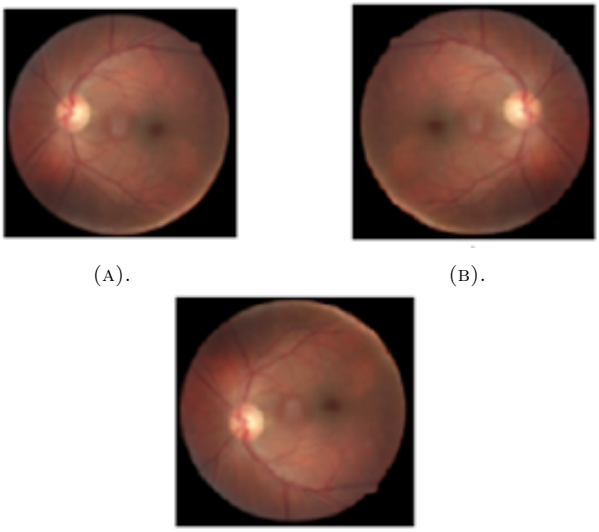
$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix}. \tag{2}$$

Figure 2 highlights the result of using the mirror effect (horizontal and vertical) on an original image of a retinopathic diabetes.

The data augmentation technique is exclusively implemented on the training subset of the dataset. This precaution is taken to prevent potential bias in the accurate evaluation of the model's performance when applied to the entire dataset, including the test set. This ensures that augmented data are encountered only during the training phase and not during testing. Consequently, the training set is increased from 3 373 to 10 119 images, i.e. we triple its size. The model becomes more robust as the training set grows.

4. Proposed model

In this study, we examined the various methods for identifying eye diseases from fundus photographs (as mentionned in Figure 3). After that, we looked at the CNN architecture, which is based on Hubel and Wisiel's study on the cat's visual cortex [28], as well as a variety of pre-trained models, including AlexNet, ResNet 18, ResNet 50, GoogleNet, VGG 19, and YOLOv3.



(A). (B). (C).

FIGURE 2. Data augmentation reflection (DAR) results: (A) Original image; (B) Horizontal mirror image; (C) Vertical mirror image.

In general, all neural networks need calculations, and the image classification in particular has to comprehend several essential aspects of the image, such as its size, channel, and processing capabilities. This shows a broad application of deep learning neural networks for classifying fundus images. Additionally, it describes the important outcomes produced by these neural approaches. The feature extraction stage, which is very laborious since it must be done by hand, is one that deep learning approaches allow us to avoid in addition to processing a very high quantity of images. Due to their capacity to collect features and learn to differentiate between various classes, convolutional neural networks (CNN) are the most often employed DL technology in the healthcare industry. The pre-trained model has a specific design and offers a more straightforward method to accurately and speedily retrain neural networks on chosen datasets (all the pretrained models architectures are detailed in Section 5.2).

5. Results

5.1. The proposed CNN

As our CNN's structure demonstrates, there are several layers represented. As shown in Table 2, the CNN network is composed of six convolution layers, with Conv 1 consisting of 32 feature maps, C2 consisting of 64 feature maps, and C3 consisting of 128 feature maps as its input. Six pooling layers are used, with Layers S1, S2, ..., and S6 being subsampling layers with the same amount of feature maps as their previous convolution. Following the pre-processing layers,

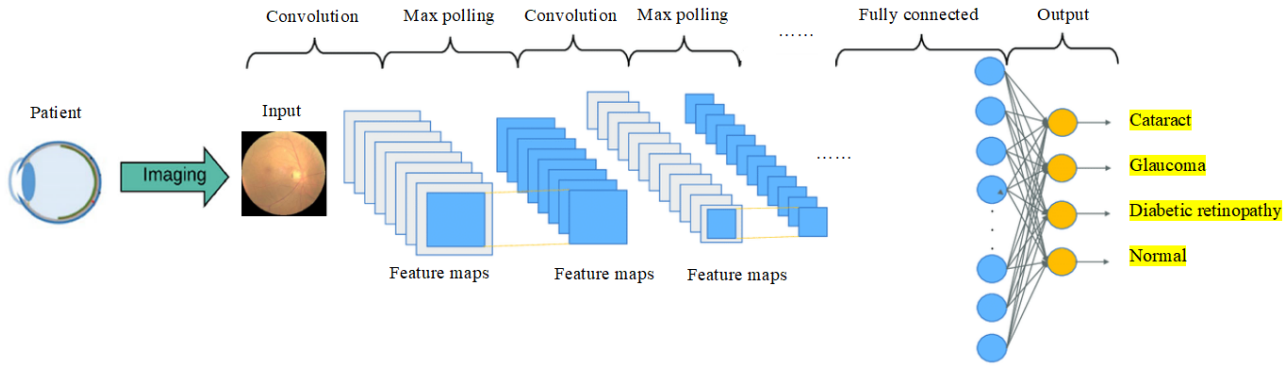


FIGURE 3. The proposed model with different algorithms.

there are four completely connected (FC) layers with four neurons each, which connect to all of the maps of the previous pooling layer. The output layer is then a fully connected layer as well. In Table 2, the CNN architecture is illustrated.

Name	Type	Size
Input layer	Input data	512*512
Conv1	Convolution + Relu	32*32*8
S1	Max pooling	3.2
Conv2	Convolution + Relu	64*64*3
S2	Max pooling	3.2
Conv3	Convolution + Relu	128*128*5
S3	Max pooling	3.2
Conv4	Convolution + Relu	256*256*5
S4	Max pooling	3.2
Conv5	Convolution + Relu	512*512*5
S5	Max pooling	3.2
Conv6	Convolution + Relu	1 024*1 024*5
S6	Max pooling	3.2
Fc	Fully connected	1 Fc(4)

TABLE 2. The architecture of the CNN model.

All of the hyperparameters used in this model, including the number of epochs employed, the mini batch size (64), the learning rate of 0.001, and the frequency validation of 20, are defined in the choices side, where the trained parameters are contained. The accuracy of this model was tested using the given CNN, which was trained using various parameters. In Table 3, the CNN model's training parameters are listed.

Parameter	Value
Initial learning rate	0.001
Optimiser	SGDM/Adam/RMSProp
Momentum	0.9
Max epoch	5 to 20
Mini batch size	64
Validation frequency	20

TABLE 3. The training parameters of the CNN model.

We take into account several optimisers when training the models:

- SGDM: the stochastic gradient descent with momentum Solver [29].
- Adam: adaptive moment estimation [30].
- RMSProp: Root Mean Square Propagation [31].
- Momentum is the contribution to the current iteration of the stochastic gradient descent with momentum of the parameter update stage of the previous iteration.
- Mini Batch size: The stochastic gradient descent technique uses a portion of the training data for each iteration to evaluate the gradient and update the parameters. At each iteration, a distinct subset, known as a mini-batch, is used to assess the gradient of the loss function and update the weights.
- The Validation Frequency value indicates how many iterations between evaluations of validation metrics.

To evaluate the performance of the trained models, we used the accuracy parameter. Accuracy is termed as the ratio of correctly classified images over the total number of images. The following Formula (3) is used:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \qquad (3)$$

As indicated in Formula (3), accuracy is calculated by dividing the sum of the true positive (TP) and true negative (TN) classes by the sum of the true positive, true negative, false positive (FP), and false negative (FN) classes.

We can conclude from Table 4 that the suggested approach based on the CNN produced satisfactory results. After 20 epochs, the best recognition rate was 87.4 %. The results are not enhanced by adding more epochs; rather, the opposite has happened. However, we think that these results can still be improved. To this end, we will use a variety of pretrained models in the sections that follow. It is anticipated that these models will aid in the better diagnosis and assessment of eye illnesses. So, we observe that while increasing the number of epochs, the execution time increases without improving the accuracy. This is the reason

The CNN model				
	5 epochs	10 epochs	15 epochs	20 epochs
Accuracy	82.46 %	77.7 %	85.2 %	87.4 %
GPU execution time	67 m 22 s	133 m 44 s	198 m 20 s	253 m 30 s

Table 4. The results of the CNN model (Image input 256*256).

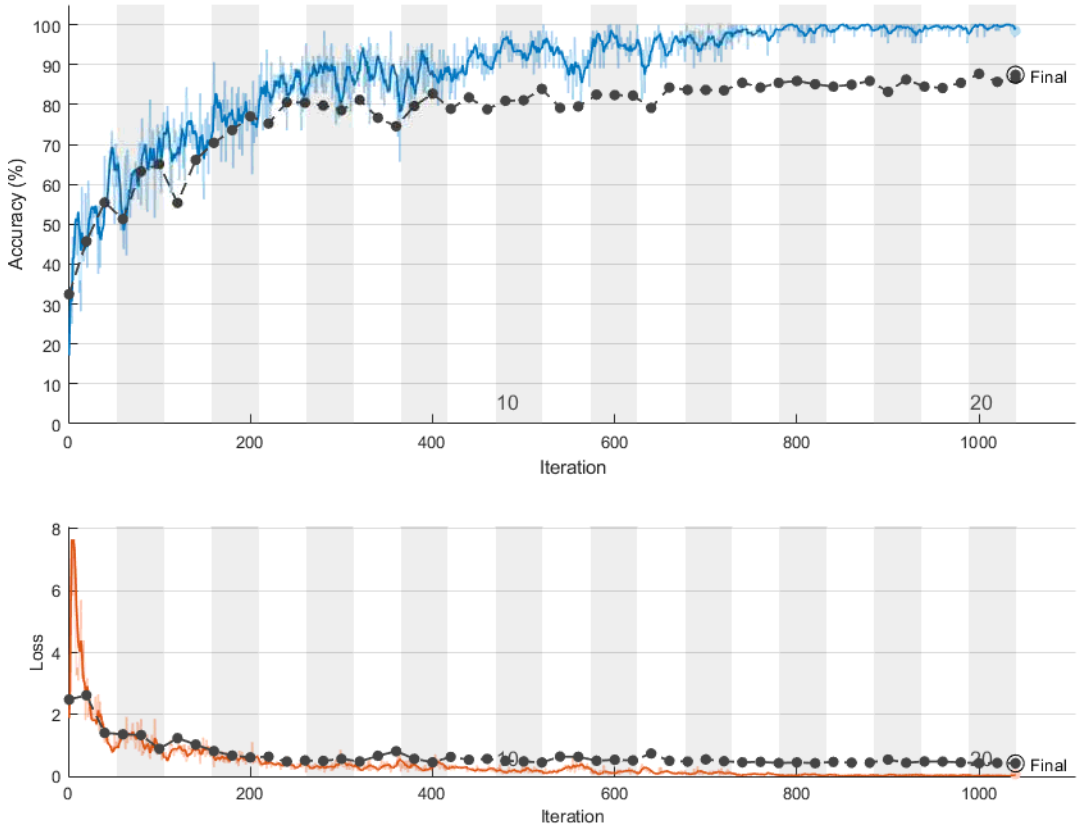


Figure 4. The results of CNN model after 20 epochs.

why we did not go beyond 20 epochs. Figures 4 and 5 mention the accuracy of CNN, the loss function of the same model, and the confusion matrix, respectively.

5.2. The proposed pretrained models

5.2.1. YOLO v3

YOLOv3 employs Darknet-53, a brand-new network inspired by ResNet, to extract features. The YOLOv3 architecture has 3 prediction heads, 53 convolutions, and skip connections similar to ResNet. Each one compresses the image differently in terms of space. And by clustering, a total of nine preceding bounding boxes of various sizes can be generated. The YOLOv3 architecture is used exactly as described in [32]. The original fundus images were scaled and transformed to RGB pictures in order to fit each pretrained model's input image size. The learning rate was maintained throughout training for the training options, validation data and training data were both shuffled prior to each network validation.

5.2.2. AlexNet

AlexNet is an eight layer convolutional neural network [33] made up of five convolutional layers with convolutional filter sizes of three by three and two by two for the maximum pooling operation. Fully connected layers are the final three layers.

AlexNet uses ReLU as the activation function rather than the conventional sigmoid and tanh functions, which is different from previous neural networks. ReLU is an unsaturated activation function that not only significantly accelerates the model's training time, but also more effectively addresses the gradient disappearance and gradient explosion issues, making it simple to train a deeper network. The AlexNet model's standard input size is 227*227*3.

5.2.3. ResNet 50

A ResNet deep residual learning network introduces the residual block notion [34]. Through residual blocks, the output of the second block is linked to the input of the first block. The residual block can learn about the residual function using this method without inflating the parameters. A convolutional layer, 48

FIGURE 5. The confusion matrix of CNN model after 20 epochs.

residual blocks, and a classifier layer with eleven and thirty-three tiny filters make up the 50-layer residual block known as ResNet 50.

5.2.4. GOOGLENET (INCEPTION V3)

GoogleNet contains 50 layers and is a convolutional neural network [35]. The algorithm, named "Going deeper with convolutions," was developed and trained by Google. The development of the Inception module, which consists of a series of 1-by-1 convolutional layers/blocks used for dimensionality reduction and feature aggregation, is the key component of GoogleNet/Inception architecture. This model had 9 inception modules and a total of 22 layers. Up to 1 000 objects can be classified using the pretrained version of Inceptionv3 with the ImageNet dataset [36] weights. This network's image input was 299×299 pixels in size.

5.2.5. VGG 19

A convolutional neural network with a depth of 19 layers is called the VGG. Karen Simonyan and Andrew Zisserman created and trained it at the Visual Geometry Group at Oxford University in 2014 [35], demonstrating that using tiny filters of size 3*3 in each convolutional layer across the network can improve performance. The fundamental tenet of VGG architecture is that several small filters can simplify design and replicate results more faithfully than bigger filters. Using more than a million images from the ImageNet dataset, the VGG 19 network is trained. Of course, you can import the model with training weights from ImageNet. The network was trained on colorful images with a resolution of 224×224 pixels.

5.2.6. RESNET 18

A convolutional neural network with 18 layers in depth is called ResNet 18. Deep Residual Learning for Image Recognition, as it is known, was developed and trained by Microsoft in 2015 [34]. To address the issue of vanishing gradient that may affect the weightage change in neural networks, ResNet architectures introduced the use of residual layers and skip connections. This made training easier and allowed neural networks to get much deeper with greater performance. More than a million images from the ImageNet dataset were used to train this model. The network was trained on coloured images with a resolution of 224×224 pixels and can categorise up to 1 000 objects.

Basic identical parameters of the six networks AlexNet, GoogleNet, ResNet 18 ... and YOLOv3 are mentioned in the Table 5.

Parameter	Value
Initial learning rate	0.001
Optimiser	SGDM/Adam/RMSProp
Max epoch	5
Mini batch size	20
Activation function	Softmax
Validation frequency	20

TABLE 5. The training parameters of the pretrained models.

We have, therefore, summarised our contributions as follows.

- We use the CNN architecture to evaluate the performance.

- We suggested the use of pretrained models, such as AlexNet, ResNet 18, ResNet 50, GoogleNet, VGG 19, and YOLOv3, to detect glaucoma, diabetic retinopathy, cataract and a normal patient using fundus images and improve the diagnosis.
- We trained the models separately to discriminate between four categories.

In addition to the accuracy parameters, we estimated the GPU execution time for each model. The results obtained are shown in Table 6.

The most popular method for evaluating model performance based on true positives (TP), true negatives (TN), false positives (FP), and false negatives is the confusion matrix (FN).

Figure 6 represent the confusion matrix of the GoogleNet and ResNet 18, which gave the best results in terms of accuracy with the SGDM optimiser.

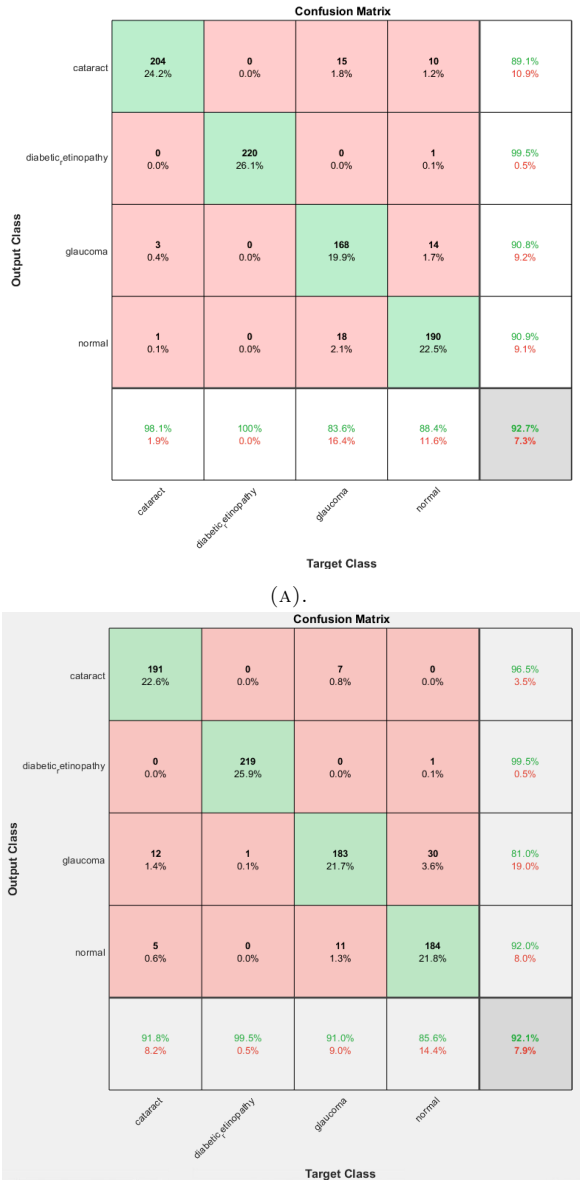
The Figure 7 represent the confusion matrix of the ResNet 18 and VGG 19, which gave the best results in terms of accuracy with the Adam optimiser.

Figure 8 represents the confusion matrix of the ResNet 18 and VGG 19 , which gave the best results in terms of accuracy with the RMSProp optimiser.

We were able to accurately classify the images in our dataset thanks to experiments using pre-trained models and the classic CNN. The accuracy rate for CNN is 88.7 %. The best accuracy was achieved for the SGDM optimiser, 92.7 % and 92.1 % with GoogleNet and ResNet 18, respectively, when the same dataset and hyper-parameters were employed. The best accuracy for the Adam optimiser was achieved with ResNet 18 (88.3 %) and VGG 19 (87.1 %). The best accuracy was obtained with the identical architectures (ResNet 18 and VGG19), 79.9 % and 83.3 %, respectively, when the last optimiser (RMSProp) was used.

The improved classification rate with ResNet 18 can be linked to the network's usage of methods to reduce over-fitting in its model. The first method involved artificially enlarging the dataset with the aid of a label-preserving transformation. This involved extracting random patches (224×224 for ResNet 18) and training the network on them while varying the intensities of the RGB channels in training images. The result was generating image translations and horizontal reflections. The second strategy to lessen over-fitting was "dropout", which involves removing neurons that do not contribute to the forward pass or back-propagation. This reduces the complicated neuronal co-adaptations and forces the model to learn stronger features.

Table 6 shows that almost all of the pretrained models used are outperformed by the SGDM optimiser. This efficiency can be attributed to the adaptive approaches' propensity to quickly converge to sharper minima whereas the SGD with momentum appears to find flatter minima than Adam. Sharper minima



(A).

(B).

FIGURE 6. Confusion matrix for the best models with SGDM optimiser: (A). GoogleNet; (B). ResNet 18.

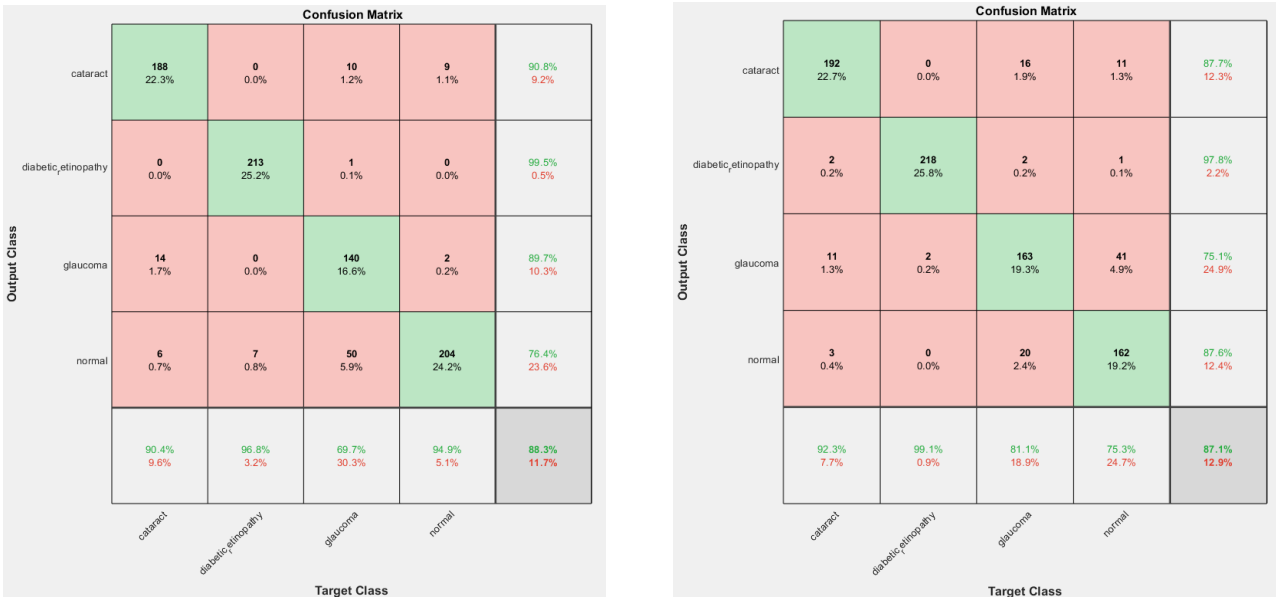
do not generalise as well as flat ones.

By compressing the input image while preserving significant details/information, GoogleNet increases performance. Utilising Inception units allows us to create deeper networks by reducing the amount of parameters in a network. Additionally, it is evidently a lot faster than the VGG 19, from this comparison.

In the case of AlexNet, the GPU execution time is ideal, but this architecture only works well with the SGDM optimiser. Some pretrained models converge significantly faster than CNN, and this requires our attention. The time saved is enormous; for illustration, CNN obtained 82.46 % with 67 minutes and 22 seconds after 5 epochs. AlexNet provides a superior performance for the same epochs, 88.9 % in 16m 56s (It provides a better result in only a one-fourth of the time). However, in order to make a meaningful com-

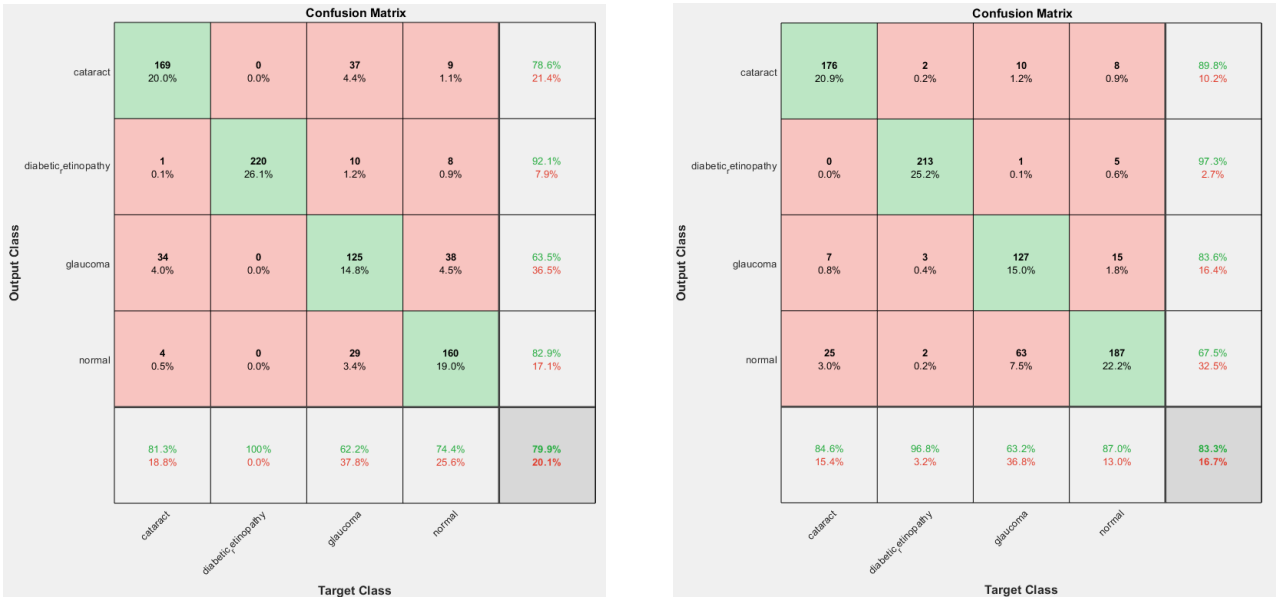
		The pretrained models					
		Alex net	**Google Net**	**ResNet 18**	**ResNet 50**	**VGG 19**	**YOLO v3**
SGDM	**Accuracy**	88.9 %	92.7 %	92.1 %	57.8 %	**87.9 %**	68.07 %
	GPU execution time	16m 56s	42m 9s	35m 42s	9m 43s	**282m 32s**	183m 13s
ADAM	**Accuracy**	40.3 %	86.3 %	88.3 %	68.4 %	**87.1 %**	51.42 %
	GPU execution time	19m 11s	42m 24s	36m 8s	9m 39s	**283m 1s**	184m 47s
RMSProp	**Accuracy**	60.7 %	69.3 %	79.9 %	62.4 %	**83.3** %	55.09 %
	GPU execution time	18m 56s	42m 59s	36m 35s	9m 28s	**280m 58s**	184m 9s

TABLE 6. The results of the pretrained models.



(A). (B).

FIGURE 7. Confusion matrix for the best models with Adam optimiser: (A) ResNet 18; (B) VGG 19.



(A). (B).

FIGURE 8. Confusion matrix for the best models with RMSProp optimiser: (A) ResNet 18; (B) VGG 19.

parison, we have kept the same training parameters for all the pretrained models. We can also remark that these findings can be improved.

6. Conclusions

The goal of this research was to construct a convolutional neural network (CNN) model to classify three different eye diseases using fundus images. In addition to CNN, pre-trained models, such as AlexNet, GoogleNet, ResNet 18, and VGG 19, were used in this study. After training on 80 % of the dataset and testing on 20 %, CNN produced a classification result for the four classes with an accuracy of 87.4 % using 6 convolution layers and 20 epochs. We also tried to use the pre-trained models as a result of the subpar performance. These were applied to the same dataset and produced better results for GoogleNet, ResNet 18, VGG 19, and AlexNet with the SGDM optimiser (92.7 %, 92.1 %, 87.9 %, and 88.9 %, respectively), and 88.3 % , 87.1 % for ResNet 18 and VGG19, respectively, with the Adam optimiser. We can conclude that the RMSProp optimiser does not yield satisfactory results.

In addition to selecting better optimisation techniques, much research is being done to suggest architectures that initially create smoother loss functions, a field of research that we are planning to explore in future work.

References

[1] P. R. Velagaleti, M. H. Buonarati. Challenges and strategies in drug residue measurement (bioanalysis) of ocular tissues. In B. C. Gilger (ed.), *Ocular Pharmacology and Toxicology*, pp. 33–52. Humana Press, Totowa, NJ, 2013. https://doi.org/10.1007/7653_2013_6

[2] World Health Organisation. [2023-02-15]. https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

[3] R. R. A. Bourne, S. R. Flaxman, T. Braithwaite, et al. Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. *The Lancet Global Health* **5**(9):e888–e897, 2017. https://doi.org/10.1016/S2214-109X(17)30293-0

[4] National Eye Institute. Statistics and data: Cataracts. 2018. [2023-02-15]. https://nei.nih.gov/eyedata/cataract

[5] D. Pascolini, S. P. Mariotti. Global estimates of visual impairment: 2010. *British Journal of Ophthalmology* **96**(5):614–618, 2012. https://doi.org/10.1136/BJOPHTHALMOL-2011-300539

[6] World Health Organisation. [2023-02-15]. https://www.who.int/news/item/08-10-2019-who-launches-first-world-report-on-vision

[7] P. Storey, B. Munoz, D. Friedman, S. West. Racial differences in lens opacity incidence and progression: The Salisbury Eye Evaluation (SEE) Study. *Investigative Ophthalmology & Visual Science* **54**(4):3010–3018, 2013. https://arvojournals.org/arvo/content_public/journal/iovs/933467/i1552-5783-54-4-3010.pdf https://doi.org/10.1167/iovs.12-11412

[8] S. Kumar, S. Pathak, B. Kumar. Automated detection of eye related diseases using digital image processing. In *Handbook of Multimedia Information Security: techniques and applications*, pp. 513–544. 2019.

[9] S. Júnior, D. Welfer. Automatic detection of microaneurysms and hemorrhages in color eye fundus images. *International Journal of Computer Science and Information Technology* **5**(5):21–37, 2013. https://doi.org/10.5121/ijcsit.2013.5502

[10] Santé et sciences en Algérie. [2023-01-29]. https://www.aps.dz/sante-science-technologie/66828-la-sao-appelle-a-moderniser-les-plateaux-techniques-des-chu

[11] U. Farooq, N. Y. Sattar. Improved automatic localization of optic disc in retinal fundus using image enhancement techniques and svm. In *2015 IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, pp. 532–537. 2015. https://doi.org/10.1109/ICCSCE.2015.7482242

[12] G. An, K. Omodaka, S. Tsuda, et al. Comparison of machine-learning classification models for glaucoma management. *Journal of Healthcare Engineering* **2018**:6874765, 2018. https://doi.org/10.1155/2018/6874765

[13] A. Siddiqui, V. Garg. An analysis of adaptable intelligent models for pulmonary tuberculosis detection and classification. *SN Computer Science* **3**:34, 2022. https://doi.org/10.1007/s42979-021-00890-4

[14] A. Bassel, A. Abdulkareem, Z. Abdi, et al. Automatic malignant and benign skin cancer classification using a hybrid deep learning approach. *Diagnostics* **12**(10):2472, 2022. https://doi.org/10.3390/diagnostics12102472

[15] B. Samir, S. Mwanahija, B. Soumia, U. Özkaya. Deep learning for classification of chest X-ray images (Covid 19), 2023. [2023-02-20]. arXiv:2301.02468 https://doi.org/10.48550/arXiv.2301.02468

[16] S. S. M. Sheet, T.-S. Tan, M. A. As'ari, et al. Retinal disease identification using upgraded CLAHE filter and transfer convolution neural network. *ICT Express* **8**(1):142–150, 2022. https://doi.org/10.1016/j.icte.2021.05.002

[17] R. Sarki, K. Ahmed, H. Wang, Y. Zhang. Automated detection of mild and multi-class diabetic eye diseases using deep learning. *Health Information Science and Systems* **8**(1):32, 2020. https://doi.org/10.1007/s13755-020-00125-5

[18] A. M. Alqudah. AOCT-NET: a convolutional network automated classification of multiclass retinal diseases using spectral-domain optical coherence tomography images. *Medical & biological engineering & computing* **58**:41–53, 2020. https://doi.org/10.1007/s11517-019-02066-y

[19] H. Gu, Y. Guo, L. Gu, et al. Deep learning for identifying corneal diseases from ocular surface slit-lamp photographs. *Scientific Reports* **10**:17851, 2020. https://doi.org/10.1038/s41598-020-75027-3

[20] Y. Peng, S. Dharssi, Q. Chen, et al. DeepSeeNet: A deep learning model for automated classification of patient-based age-related macular degeneration severity from color fundus photographs. *Ophthalmology* **126**(4):565–575, 2019. https://doi.org/10.1016/j.ophtha.2018.11.015

[21] M. Alam, D. Le, J. I. Lim, et al. Supervised machine learning based multi-task artificial intelligence classification of retinopathies. *Journal of Clinical Medicine* **8**(6):872, 2019. https://doi.org/10.3390/jcm8060872

[22] X. Zhang, Z. Xiao, R. Higashita, et al. Adaptive feature squeeze network for nuclear cataract classification in as-oct image. *Journal of Biomedical Informatics* **128**:104037, 2022. https://doi.org/10.1016/j.jbi.2022.104037

[23] A. Imran, J. Li, Y. Pei, et al. Fundus image-based cataract classification using a hybrid convolutional and recurrent neural network. *The Visual Computer* **37**:2407–2417, 2020. https://doi.org/10.1007/s00371-020-01994-3

[24] M. A. S. Ali, K. Balasubramanian, G. D. Krishnamoorthy, et al. Classification of glaucoma based on elephant-herding optimization algorithm and deep belief network. *Electronics* **11**(11):1763, 2022. https://doi.org/10.3390/electronics11111763

[25] A. R. Prananda, E. L. Frannita, A. H. T. Hutami, et al. Retinal nerve fiber layer analysis using deep learning to improve glaucoma detection in eye disease assessment. *Applied Sciences* **13**(1):37, 2022. https://doi.org/10.3390/app13010037

[26] V. Rajinikanth, R. Sivakumar, D. J. Hemanth, et al. Automated classification of retinal images into AMD/non-AMD class—a study using multi-threshold and Gassian-filter enhanced images. *Evolutionary Intelligence* **14**:1163–1171, 2021. https://doi.org/10.1007/s12065-021-00581-2

[27] A. Miere, T. L. Meur, K. Bitton, et al. Deep learning-based classification of inherited retinal diseases using fundus autofluorescence. *Journal of Clinical Medicine* **9**(10):3303, 2020. https://doi.org/10.3390/jcm9103303

[28] D. H. Hubel, T. N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology* **195**(1):215–243, 1968. https://doi.org/10.1113/jphysiol.1968.sp008455

[29] S. Ruder. An overview of gradient descent optimization algorithms. [2023-02-26]. arXiv:1609.04747 https://doi.org/10.48550/ARXIV.1609.04747

[30] D. P. Kingma, J. Ba. Adam: A method for stochastic optimization, 2014. [2023-02-26]. arXiv:1412.6980 https://doi.org/10.48550/arXiv.1412.6980

[31] T. Tieleman, G. Hinton, et al. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning* **4**(2):26–31, 2012.

[32] J. Redmon, A. Farhadi. YOLOv3: An incremental improvement, 2018. [2023-02-29]. arXiv:1804.02767 https://doi.org/10.48550/arXiv.1804.02767

[33] S. Li, L. Wang, J. Li, Y. Yao. Image classification algorithm based on improved AlexNet. *Journal of Physics: Conference Series* **1813**:012051, 2021. https://doi.org/10.1088/1742-6596/1813/1/012051

[34] K. He, X. Zhang, S. Ren, J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778. 2016. https://doi.org/10.1109/CVPR.2016.90

[35] C. Szegedy, W. Liu, Y. Jia, et al. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9. 2015. https://doi.org/10.1109/CVPR.2015.7298594

[36] A. Krizhevsky, I. Sutskever, G. E. Hinton. ImageNet classification with deep convolutional neural networks. In F. Pereira, C. J. Burges, L. Bottou, K. Q. Weinberger (eds.), *Advances in Neural Information Processing Systems*, vol. 25, pp. 1–9. Curran Associates, Inc., 2012. https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf