Sarki *et al. Health Inf Sci Syst (2020) 8:32*
https://doi.org/10.1007/s13755-020-00125-5

Health Information Science and Systems

RESEARCH

Automated detection of mild and multi-class diabetic eye diseases using deep learning

Rubina Sarki*, Khandakar Ahmed, Hua Wang and Yanchun Zhang

Abstract

Diabetic eye disease is a collection of ocular problems that affect patients with diabetes. Thus, timely screening enhances the chances of timely treatment and prevents permanent vision impairment. Retinal fundus images are a useful resource to diagnose retinal complications for ophthalmologists. However, manual detection can be laborious and time-consuming. Therefore, developing an automated diagnose system reduces the time and workload for ophthalmologists. Recently, the image classification using Deep Learning (DL) in between healthy or diseased retinal fundus image classification already achieved a state of the art performance. While the classification of mild and multi-class diseases remains an open challenge, therefore, this research aimed to build an automated classification system considering two scenarios: (i) mild multi-class diabetic eye disease (DED), and (ii) multi-class DED. Our model tested on various datasets, annotated by an opthalmologist. The experiment conducted employing the top two pretrained convolutional neural network (CNN) models on ImageNet. Furthermore, various performance improvement techniques were employed, i.e., *fine-tune*, *optimization*, and *contrast enhancement*. Maximum accuracy of 88.3% obtained on the VGG16 model for multi-class classification and 85.95% for mild multi-class classification.

Keywords: Diabetic eye disease, Deep learning, Classification, Image processing

Introduction

The World Health Organisation (WHO) reports, 2.2 billion individuals globally have a blindness or vision loss, of which at least 1 billion have impaired vision, which could have been reversed.[1] One of the reasons for this blindness is identified as diabetes mellitus or diabetes. Approximately one-third of those with diabetes expected to diagnosed with a DED, a chronic eye disease that can cause permanent visual impairment if left unattended [10]. DED includes diabetic retinopathy (DR), glaucoma (Gl), diabetic macular edema (DME), and cataract[2] (Ca) (see Fig. 1). It is crucial to identify and diagnose these diseases for the treatment.

Motivated by the necessity of active strategies for diagnosis, and prevention to implement the broad spectrum of needs associated with retinal disorders and visual impairments throughout the lifespan. Automated DED diagnostic techniques using DL are vital to addressing these issues [1, 18]. Timely screening of DED, which is crucial to effective prognosis based on professional ophthalmologist, is time and labor intensive [16].

Although DL has generally achieved high validation accuracies for healthy and diseased (binary) classification, the results of mild and multi-class classification are less impressive, particularly for early stage impairment. Therefore, in this paper, we present a Deep Convolutional Neural Network (DCNN) based automatic DED classification model that can classify healthy images from disease pathologies. To identify the best performing framework for the mild and multi-class DED classification tasks, we initially evaluate different DCNN architectures. We aim to achieve the highest performance levels than reported in the previous works. Thus, we trained and tested mild and multi-class classification models to improve sensitivities for the different DED classes, incorporating different preprocessing and augmentation approaches to boost the accuracy of the result and enhance adequate sample volume for the dataset.

Early rectification of retinal diseases is vital, but the diagnosis of these diseases utilizing neural networks requires a substantial amount of time and memory.

*Correspondence: rubina.sarki@live.edu.vu.au
Victoria University, Ballarat Road, Melbourne, VIC 3011, USA

[1] https://www.who.int/en/news-room/fact-sheets/detail/blindness-and-visual-impairment.

[2] The National Institute of Diabetes and Digestive and Kidney Diseases.

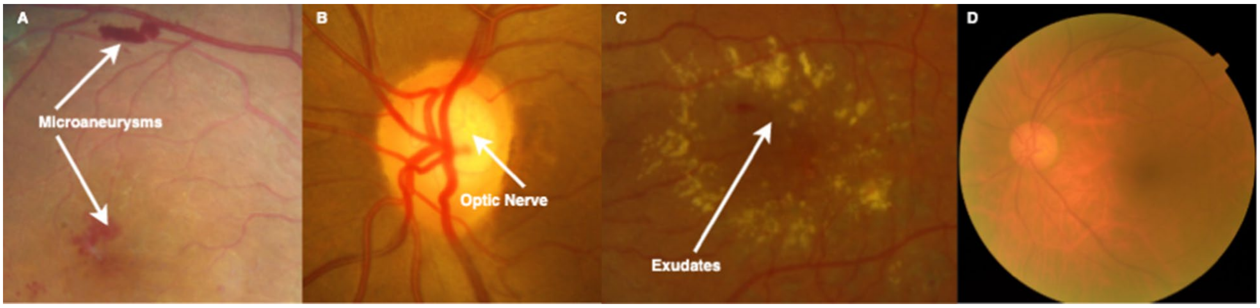

Fig. 1 Complications of DED in fundus images; **a** Microaneurysms; narrow bulges in the side of the blood vessel (Diabetic Retinopathy). **b** Progressive damage of optic nerve damage(Glaucoma). **c** Exudates formation in macular region and thickening of macula(Diabetic Macular Edema). **d** Degeneration of lens (Cataract)

Additional data must supply to enhance the precision, but this requires high computational power and a massive amount of time investment. Thus, a comparatively pretrained model can benefit the process as it adapts the design to reduce losses. Pretrained models or Transfer Learning (TL) [38] models has already been demonstrated and validated promising results in medical image classification and detection [5, 20, 26, 29]. As part of this analysis, we used the state of the art CNN models, pretrained on the broad public image repository ImageNet, following the TL principle. The top layers of the deep neural network were learned from the publicly accessible fundus image corpora for mild and multi-class classification by using initialized weights. The research performed in this paper focuses solely on different DED instances, actually challenging to classify as opposed to previous approaches. Initially, the highest-performing CNN model is selected based on a comprehensive experiment conducted. Lastly, it evaluates the set of performance improvements, including fine-tuning, and optimizer selection. Finally, an ideal intermediate scenario of accuracy achieved is selected to facilitate efficient and effective fully automated DL based system development to improve outcomes to mass screening services among the at risk population.

The arrangement of this article is as follows. "Literature review" section presents the literature review followed when surveying the articles. "Study design" section study design based on the datasets, image preprocessing and DL techniques are presented in this section. "Experiment setup" section addresses the experiments performed in two scenarios. "Results" section compared the results based on the classification evaluation. Finally, "Conclusions" section concludes the paper.

Literature review

Several previous research work concentrated on automated retinal disease detection by using machine learning algorithms [2, 11, 12, 40] to classify a substantial number of fundus images captured from retinal screening programs [7, 44]. Multiple machine learning techniques: Artificial Neural Network (ANN), K-nearest neighbour algorithm, Support Vector Machine (SVM), and Naive Bayes classifier, were implemented to the automated identification of retinal diseases [29]. Many studies implemented ANN models to identify the disparity between glaucoma and non glaucoma [8, 43]. A glaucoma research team observed visual field evaluation by discovering preperimetric glaucoma utilising DL feedforward neural networks (FNN) [3]. For the grading intensity of the nuclear cataract [15], an artificial DCNN has been applied. DL emerges as popular solution for various classification problems in the field of ML techniques [17, 19, 23, 27]. The Google research group has developed the advanced DL model capable of diagnosing Diabetes Mellitus Retinopathy (DMR) [18]. The study of the Age-related Macular degeneration (AMD) was carried out using similar DL methods, fundus photographs and optical coherence tomography [6, 24]. However, all retinal image classification studies selected binary classification through which problems of normal versus one disease were solved [21]. Furthermore, in order to classify mild and multi-class diabetic retinopathy, Lam et al. [22] employed pretrained networks (GoogleNet and AlexNet) in the Messidor dataset, authors applied selective contrast-limited adaptive histogram equalization (CLAHE) and documented improvement in the identification of subtle features. Multi-class DL algorithms for automatic detection of ten retinal disorders were studied by Choi et al. [9]. The results of this work showed that current DL algorithms were unsuccessful in classifying retinal images

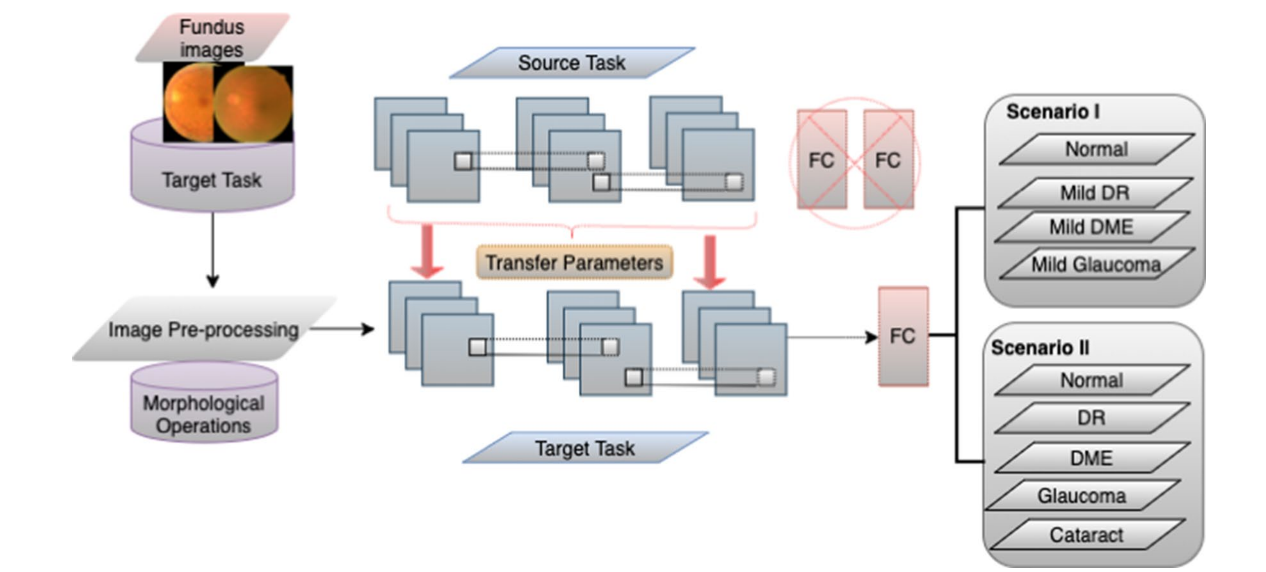

Fig. 2 Transfer learning process

from small datasets of multiple groups. While research conducted in this area have published the result that high classification performance in standardized experimental settings, it is fundamentally difficult to implement the binary classification model to the actual medical practice where patients visiting are suffering from various DED. Indeed, studies have been very limited on mild and multi-class classification aimed at recognising DED.

In this research, we adapted TL in mild and multi-class DED settings using a state of the art CNN for fundus image analysis. This paper articulates a pilot study planned with the use of small open source fundus retinal image database for TL evaluation on mild and multi-class classification.

Study design

This study's real objective is the performance enhancement of mild and multi-class DED identification via an experimental evaluation of various techniques for improving the classification. The associated goals and objectives can be identified as follows.

1. Comparative analysis of two CNN frameworks using TL concept,
2. Impact of a fine-tuning analysis on the performance of frameworks,
3. Impact of an optimizer analysis on the performance of frameworks,
4. Analysis of data improvement and contrast enhancement techniques for further classification improvements on mild and multi-class DED detection task.

The TL process is illustrated (see Fig. 2), to demonstrate the steps that followed.

Data collection

Data were obtained from the open source, including Messidor, Messidor-2, DRISHTI-GS, and retina datasets, which are publicly available. Messidor dataset includes high fidelity images with reliable labeling despite its relatively small scale. Similarly, Messidor-2 is a public database used by other individual people to evaluate DED algorithm performance. The database consists of 1748 images of 874 subjects. Messidor-2 differs from the initial 1200 image Messidor set of data, and it has two images for each item, one for each eye. The Drishti-GS dataset contains 101 retinal images, with 31 normal images and 70 lesion images. Cataract dataset acquired from retina dataset Github.[3] This dataset consists of 100 cataract images (Fig. 3).

Image processing

Data imbalance is the common problem encountered in machine learning applications and real-world data mining [25]. Image preprocessing plays a significant role: If the dataset have a small number of samples, in one or more categories that lead to the problem of misclassification. In this study, methods like under-sampling and over-sampling are performed to avoid misclassification. We implemented both the methodologies throughout

[3] https://github.com/yiweichen04/retina_dataset.

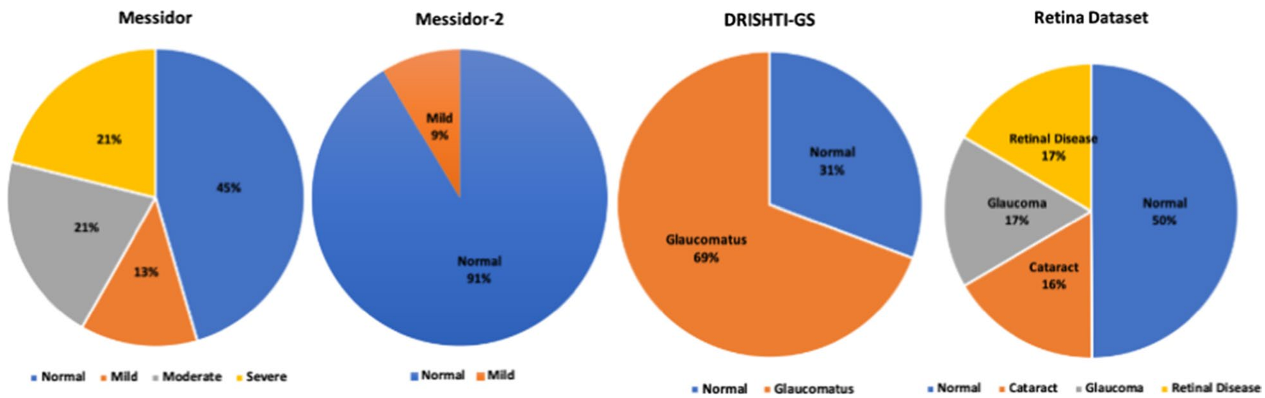

Fig. 3 Data distribution

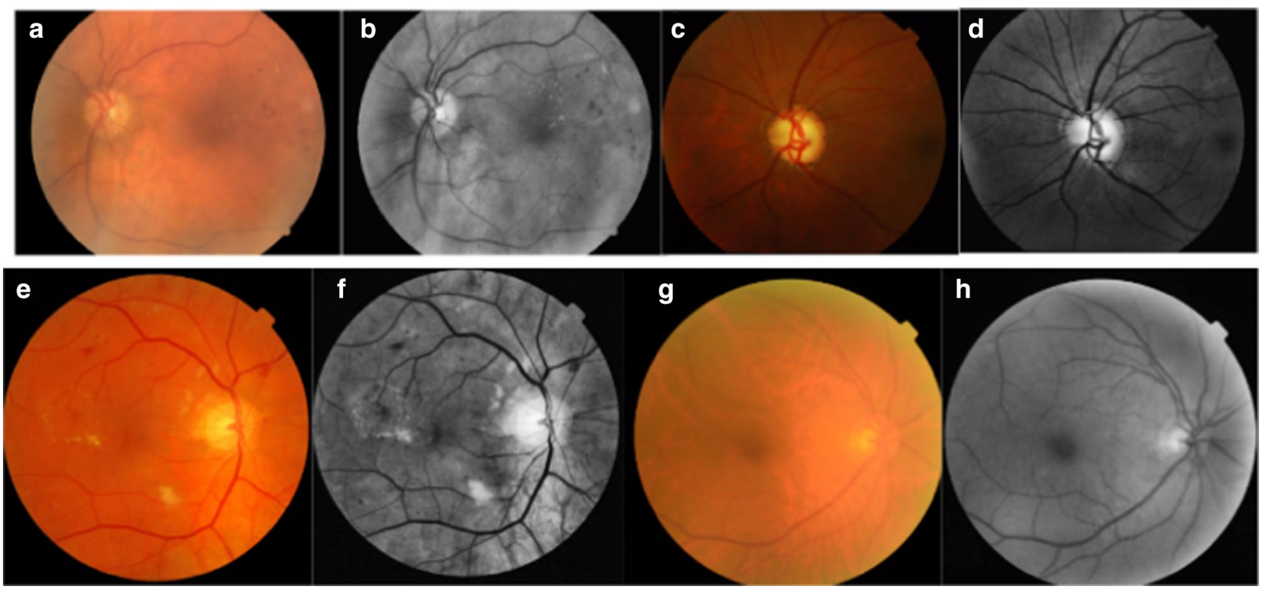

Fig. 4 Contrast enhancement **a** Original and **b** Enhanced diabetic retinopathy; **c** Original and **d** Enhanced glaucoma; **e** Original and **f** Enhanced diabetic macular edema; **g** Original and **h** Enhanced cataract

the dataset better results obtained using the under-sampling method. Followed by morphological top-hat and bottom-hat transform to enhance the contrast [37] (see Fig. 4). Two morphological transformations, top-hat and bottom-hat, are commonly used for image enhancement. These are a very effective tool for improving clarity in the presence of shading or dark areas in medical imaging. The top-hat method is defined as the difference between the input image and its opening, whereas the bottom-hat is the difference between the input image and the closing. By implementing top-hat, we can extract objects or elements smaller than the SE and brighter than their environment. On the other hand, bottom-hat produces objects more trivial than the SE and darker than their environment. So we can take advantage of these two operators by adding the top-hat and subtracting the bottom-hat result.

Pretrained convolutional neural network

In this study, we employ pretrained CNNs to incorporate the classification of the DED dataset. A deep convolutional neural network (CNN) converts a feature vector with a defined weight matrix to obtain particular feature representations without missing information about the spatial arrangement [41]. The concept uses features learned on the source task and its reuse to target jobs. TL is beneficial in areas of research that involve large quantities of data and significant computational resources [34].

Table 1 The layers and layer parameters of the VGG16 model

Layers	Layer type	Output shape	Trainable parameters
1	Cov2d	[224, 224, 64]	1792
2	Cov2d	[224, 224, 64]	36,928
4	Cov2d	[112, 112, 128]	73,856
5	Cov2d	[112, 112, 128]	147,585
6	Cov2d	[56, 56, 256]	295,168
7	Cov2d	[56, 56, 256]	590,080
8	Cov2d	[56, 56, 256]	590,080
9	Cov2d	[56, 56, 256]	590,080
10	Cov2d	[28, 28, 512]	1,180,160
11	Cov2d	[28, 28, 512]	2,359,808
12	Cov2d	[28, 28, 512]	2,359,808
13	Cov2d	[14, 14, 512]	2,359,808
14	Cov2d	[14, 14, 512]	2,359,808
15	Cov2d	[14, 14, 512]	2,359,808
16	Cov2d	[14, 14, 512]	2,359,808

Table 2 Parameters of the VGG16 model and preferred weights in this study

Model	Platform used	Image size	Optimizer	Mini-batch size	Fine-tune	Learning rate
VGG16	Anaconda	224*224	ADAM	32	15	1e−3
	Python		RMSProp			1e−3
	Keras		SGD			1e−3
	Tensorflow		AdamGrad			1e−3

Table 3 Parameters of the InceptionV3 model and preferred weights in this study

Model	Platform used	Image size	Optimizer	Mini-batch size	Fine-tune	Learning rate
InceptionV3	Anaconda	224*224	ADAM	32	100	1e−3
	Python		RMSProp			1e−3
	Keras		SGD			1e−3
	Tensorflow		AdamGrad			1e−3

Thus, we are exploring pretrained models to achieve the best possible classification outcomes. This section presents the specific information of the pretrained models.

Visual Geometry Group (VGG16) VGG was designed based on the deep convolutional neural network model in Oxford Robotics Institute by Andrew Zisserman and Karen Simonyan [35]. VGG became popular at the Large Scale Visual Recognition Challenge in 2014 (ILSVRC2014). The VGGNet operated well on the dataset of the ImageNet. To enhance the image extraction efficiency, the VGGNet used smaller 3×3 filters compared to the 11×11 AlexNet filters. There are two different versions of this deep network architecture (VGG16 and VGG19), which have different layers and depths. Moreover, the number of parameters for VGG19 is larger and more complex than VGG16 to train the model (Table 1). Table 2 Explains the parameters we used to train a system.

InceptionV3 Inception network or GoogLeNet was 22 layers network and won 2014 Image net challenge with 93.3% top-5 accuracy [36]. Later versions are referred to as Inception*VN*, where, *V* is the version and *N* is the number so inceptionV1, inceptionV2 and inceptionV3. The InceptionV3 network has several symmetrical and asymmetrical building blocks, where each block has several branches of convolutions, average pooling, max-pooling, concatenated, dropouts, and fully-connected layers. In our dataset, VGG16 obtained high accuracy than other models. Table 3 Explains the parameters we used to train a system.

Classification description

Training and testing performed by two CNN architectures, as mentioned above, in Scenario I and Scenario II classification models.

- *Scenario I* In this scenario, we will classify four classes, such as healthy, mild DR, mild DME, and mild Gl retinal fundus images (mild multi-class classification)
- *Scenario II* In this scenario, we will classify five classes, namely, healthy, DR, DME, Gl, and Ca retinal fundus images (multi-class classification).

Table 4 Confusion matrix

	Predictive positive	Predictive negative	Total
Actual positive	*TP*	*FN*	*TP* + *FN*
Actual negative	*FP*	*TN*	*FP* + *TN*
Total	*TP* + *FP*	*FN* + *TN*	

TP true positive, *FN* false negative, *FP* false positive, *TN* true negative

Classification performance analysis

The efficiency of each CNN is measured by different metrics applied to calculate the true and/or false classification for the diagnosed DED in the retinal fundus images evaluated as follows. First, the cross-validation estimator [36] is being used and resulted in a confusion matrix Table 4. When the classification model correctly classifies samples associated with a particular class, such samples placed in the TP indices. The other samples that relate to some other classes correctly identified are in the TN indices of the confusion matrix. Similarly, the FP and FN indices in the uncertainty matrix refer to the number of samples incorrectly estimated by the classifier.

Accuracy Performance parameters extracted from the confusion matrix are used for experimental analysis. Accuracy in Eq. (1) is simply the fraction of true positive and true negative by the total values of the confusion matrix. The above elements of the confusion matrix, therefore, be calculated to evaluate the effectiveness of the classifier of our presented framework.

$$Accuracy(\%) = \frac{TP + TN}{TP + FN + TN + FP}. \tag{1}$$

Similarly, *Recall*

$$Sensitivity(Recall) = \frac{TP}{TP + FN} \tag{2}$$

Precision

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

Experiment setup

All the studies conducted used Python, Keras library, TensorFlow as a back-end. The resolution of the images has been standardized to a uniform size, following each model's input requirements. The epoch number set at 15 because of the use of pretrained weights in our experiments. The distribution of training/testing dataset at 80/20. Stratified standard preference made to ensure a nearly similar dispersion of the class. Mini-batch size set to 32, and the categorical cross-entropy loss function was selected due to its suitability for multi-class classification tasks. The default ADAM was the Optimiser. The primary assessment metric for Accuracy, Specificity, and Sensitivity of test data was used for final scores validation.

Performance enhancement

Fine-tune

The neural networks used in this paper were pretrained on a large-scale ImageNet dataset covering 1000 classes, including birds, animals, and flowers. Systems obtain the highest performance in the classification tasks for objects with labeled datasets while demonstrating restricted in their assessment to specialty areas of study, such as DED detection. The diagnosis and treatment of possible pathological signs in the fundus images are based on a large number of complex characteristics and their orientation inside the fundus images [22]. There is a new representation of the feature vector on every layer of CNN by progressive extraction of the most distinguishing characteristics of [28]. The following constraints considered in the experimental work: (i) Eliminating the fully connected nodes at the end of the layer(where the true label class predictions made) and substitute the modules fully connected with those newly initialized (current pretraining approach); and (ii) eliminating the n layers and re-training the network (the suggested approach). The range of parameters used across CNN depends on the total hidden layers present on every system model. The possible classification enhancement of the DED detection task assessed as a result of the models' proposed customization options. In the analysis with Zhang et al. [45] DL based DR detection device performance accuracy improved from 95.68 to 97.15% as a result of fine-tuning (Fig. 5).

Optimizer selection

In the training phase, the neural net nodes' parameters are updated automatically to reduce the loss function. However, the direction and magnitude of the parameter adjustment are highly dependant on the optimizer utilized. The most significant weight that evaluates the performance of the Optimiser is Regularisation and Learning Rate. Too high/too lower learning rate results either in non-convergence of the loss function or in the range of the local, but not the absolute minimum. Meanwhile, Regularisation makes it possible to avoid overfitting the model by penalizing the dominant weighting factors for correct predictions. As a consequence, the generalization capability of the classifier increases when exposed to new data. The optimization techniques used for the experiments were as follows : (1) RMSprop, (2) SGD, (3) Adagrad, (4) Adam.

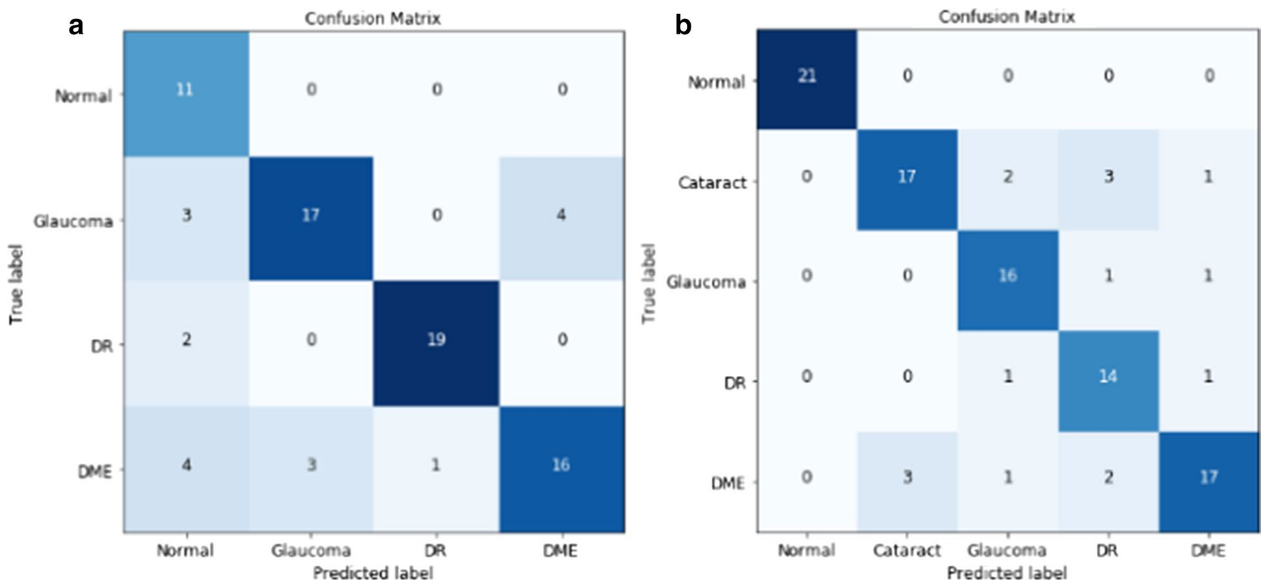

Fig. 5 Confusion Matrix in VGG16 + Adam **a** scenario I and **b** Scenario II

Table 5 Average performance of the models in mild DED classification (mild multi-classes)

Model	Optimiser	Learning rate	Accuracy* (%)	Accuracy** (%)
VGG16	*Adam*	1e−3	**82.42**	**85.94**
	RMSprop	1e−3	83.52	83.98
	SGD	1e−3	75	82.03
	Adagrad	1e−3	75	75.23
InceptionV3	*Adam*	1e−3	74	75
	RMSprop	1e−3	71	73
	SGD	1e−3	78.52	78.52
	Adagrad	1e−3	76	79.17

Bold values indicate the highest value obtained among all trained models

Accuracy* Results Before fine-tuning, Accuracy** Results after fine-tuning

Table 6 Average performance of the models in multi-class DED classification

Model	Optimiser	Learning rate	Accuracy* (%)	Accuracy** (%)
VGG16	*Adam*	1e−3	**84.88**	**88.3**
	RMSprop	1e−3	74	80
	SGD	1e−3	80	80
	Adagrad	1e−3	79.95	80
InceptionV3	*Adam*	1e−3	79	81
	RMSprop	1e−3	65	78
	SGD	1e−3	63	63
	Adagrad	1e−3	58	64

Bold values indicate the highest value obtained among all trained models

Accuracy* Results Before fine-tuning, Accuracy** Results after fine-tuning

SGD Stochastic gradient descent

Results

The two pretrained CNN models were compared with the yielded accuracy on the test dataset. Also, fine-tuning was used as a substitute for the default setting. After deletion and retraining of n layers (n was CNN-dependent), the efficiency acquired by each model was used for comparative purposes. The fine-tuning effect was evaluated in terms of Accuracy (%) increment or reduction. The highest accuracy was identified by each model (either through default or after fine-tuning) in four different optimizers. Finally, the top 1 CNN architectures + optimizers with the higher accuracy performance for the target task have been selected in Tables 5 and 6.

This study is an investigation of mild and multi-class DL algorithms for automated detection of DED. As per the British Diabetic Association (BDA) standards, a minimum amount of 80% sensitivity and 95% specificity for sight-threatening DED detection must be achieved by any method [4]. After testing our approach in DED detection tasks, the scenario I achieved maximum sensitivity of 85% and a maximum specificity of 96%. Similarly, the sensitivity of 85% and specificity of 98% for scenario II, respectively. Thus, according to the BDA standards, mild and multi-class DED detection is sufficient, in terms of its sensitivity and specificity.

Discussion

This study focuses on DL algorithms to automatically identify mild and multi-class DED. Previous researches in this topic showed the ineffectiveness of the latest DL algorithms in classifying fundus images from small datasets. It failed to demonstrate practical and effective results for a computer-aided medical applications. Therefore, this article adapted optimized DL architectures for the automated classification of normal, DR, DME, GL and Ca due to the significance of each disease in order to create an automated model for the classification.

The performance of DL models was dropped by 3%, in early stage multi-class DED classification. This finding is apparently very normal since early stage DED fundus images consist of subtle features that can be crucial for diagnosis. Interestingly, the architectures most commonly deployed were designed to identify object based features like those present in the ImageNet dataset. Then we may need a new paradigm such as lesion based (e.g. exudates) for diagnosing diseases through CNN models. Our future goals include DED lesion segmentation (region of interest) [30, 31, 42] for enhancing the identification of mild disease and moving to more complex and advantageous identification of multi-grade diseases.

Conclusions

Early identification and prompt diagnosis of DED are considered crucial to the prevention of permanent vision loss. Automated DED recognition has been the topic of several studies in the past, with the main emphasis on healthy/unhealthy binary retinal classification [32]. The results show that the identification of moderate to severe indications does not present significant difficulties due to the high visibility of the pathological features. The issue occurs with the mild identification of DED cases, in which only small lesions prove representative of the condition, sometimes undetected by the classifiers. Mild DED cases prognostication further questioned by the poor quality of fundus photography, which further complicates the recognition of subtle lesions in the eye. In the case of multi-class classification, the performance of DL models has been reduced as categories have multiplied. When categories increased, the predicted precision of the random distribution decreased. This finding corresponded to the previous studies [13]. Recent research using the GoogLeNet model to identify skin cancer has shown that increase in the number of classes has underperformed (with an accuracy of 72.1% for a three class problem and 55.4% for a nine class problem) [14]. Thus it is essential to create disease-specific strategies to differentiate between DED to enhance the efficiency of multi-class classification. Therefore, the research proposed a system that focuses entirely on the identification of mild and multi-class DED among healthy instances, as discussed in previous studies. According to the empiric nature of DL, a variety of performance optimization techniques have been applied (i) fine-tuning, (ii) optimizer choice, (iii) data increase, and (iii) contrast enhancement. Besides, the study used combined datasets from different sources to evaluate the system's robustness in its flexibility to cope with real world scenarios. As Wan et al. [39] have pointed out, the single data collection environment presents difficulties in the validation of accurate models [33].

Received: 16 August 2020 Accepted: 23 September 2020
Published online: 8 October 2020

References

1. Abràmoff MD, Reinhardt JM, Russell SR, Folk JC, Mahajan VB, Niemeijer M, Quellec G. Automated early detection of diabetic retinopathy. Ophthalmology. 2010;117(6):1147–54.
2. Arunkumar R, Karthigaikumar P. Multi-retinal disease classification by reduced deep learning features. Neural Comput Appl. 2017;28(2):329–34.
3. Asaoka R, Murata H, Iwase A, Araie M. Detecting preperimetric glaucoma with standard automated perimetry using a deep learning classifier. Ophthalmology. 2016;123(9):1974–80.
4. Association BD, et al. Retinal photography screening for diabetic eye disease. London: BDA; 1997.
5. Bargshady G, Zhou X, Deo RC, Soar J, Whittaker F, Wang H. Enhanced deep learning algorithm development to detect pain intensity from facial expression images. Expert Syst Appl. 2020;149:113305.
6. Burlina P, Freund DE, Joshi N, Wolfson Y, Bressler NM. Detection of age-related macular degeneration via deep learning. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), p. 184–88. IEEE 2016..
7. Caixinha M, Nunes S. Machine learning techniques in clinical vision sciences. Curr Eye Res. 2017;42(1):1–15.
8. Chen X, Xu Y, Wong DWK, Wong TY, Liu J. Glaucoma detection based on deep convolutional neural network. In: 2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC); 2015, p. 715–18. IEEE.
9. Choi JY, Yoo TK, Seo JG, Kwak J, Um TT, Rim TH. Multi-categorical deep learning neural network to classify retinal images: a pilot study employing small database. PLoS ONE. 2017;12(11):e0187336.
10. Congdon NG, Friedman DS, Lietman T. Important causes of visual impairment in the world today. Jama. 2003;290(15):2057–60.
11. De Fauw J, Keane P, Tomasev N, Visentin D, van den Driessche G, Johnson M, Hughes CO, Chu C, Ledsam J, Back T et al. Automated analysis of retinal imaging using machine learning techniques for computer vision. F1000Research. 2016;5.
12. De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, Askham H, Glorot X, O'Donoghue B, Visentin D, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med. 2018;24(9):1342–50.
13. Deng, J., Berg, A.C., Li, K., Fei-Fei, L.: What does classifying more than 10,000 image categories tell us? In: European conference on computer vision, p. 71–84. Springer; 2010.
14. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115–8.
15. Gao X, Lin S, Wong TY. Automatic feature learning to grade nuclear cataracts based on deep learning. IEEE Trans Biomed Eng. 2015;62(11):2693–701.
16. Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. Ophthalmology. 2017;124(7):962–9.

17. Goceri, E., Goceri, N. Deep learning in medical image analysis: recent advances and future trends; 2017.
18. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. Jama. 2016;316(22):2402–10.
19. He, J., Rong, J., Sun, L., Wang, H., Zhang, Y., Ma, J. A framework for cardiac arrhythmia detection from IoT-based ECGS. World Wide Web, p. 1–16; 2020.
20. Karri SPK, Chakraborty D, Chatterjee J. Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration. Biomed Opt Express. 2017;8(2):579–92.
21. Kwasigroch, A., Jarzembinski, B., Grochowski, M. Deep CNN based decision support system for detection and assessing the stage of diabetic retinopathy. In: 2018 International Interdisciplinary PhD Workshop (IIPhDW), p. 111–16. IEEE; 2018
22. Lam C, Yi D, Guo M, Lindsey T. Automated detection of diabetic retinopathy using deep learning. AMIA Summits Transl Sci Proc. 2018;2018:147.
23. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44.
24. Lee CS, Baughman DM, Lee AY. Deep learning is effective for classifying normal versus age-related macular degeneration oct images. Ophthalmol Retina. 2017;1(4):322–7.
25. Li H, Wang Y, Wang H, Zhou B. Multi-window based ensemble learning for classification of imbalanced streaming data. World Wide Web. 2017;20(6):1507–25.
26. Li, X., Pang, T., Xiong, B., Liu, W., Liang, P., Wang, T. Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification. In: 2017 10th international congress on image and signal processing, biomedical engineering and informatics (CISP-BMEI), p. 1–11. IEEE; 2017.
27. Ma J, Sun L, Wang H, Zhang Y, Aickelin U. Supervised anomaly detection in uncertain pseudoperiodic data streams. ACM Trans Internet Technol. 2016;16(1):1–20.
28. Mateen M, Wen J, Song S, Huang Z, et al. Fundus image classification using VGG-19 architecture with PCA and SVD. Symmetry. 2019;11(1):1.
29. Mookiah MRK, Acharya UR, Chua CK, Lim CM, Ng E, Laude A. Computer-aided diagnosis of diabetic retinopathy: a review. Comput Biol Med. 2013;43(12):2136–55.
30. Pandey D, Yin X, Wang H, Zhang Y. Accurate vessel segmentation using maximum entropy incorporating line detection and phase-preserving denoising. Comput Vis Image Unders. 2017;155:162–72.
31. Pandey D, Yin X, Wang H, Su MY, Chen JH, Wu J, Zhang Y. Automatic and fast segmentation of breast region-of-interest (ROI) and density in MRIS. Heliyon. 2018;4(12):e01042.
32. Pratt H, Coenen F, Broadbent DM, Harding SP, Zheng Y. Convolutional neural networks for diabetic retinopathy. Procedia Comput Sci. 2016;90:200–5.
33. Sarki R, Ahmed K, Zhang Y. Early detection of diabetic eye disease through deep learning using fundus images. EAI Endors Trans Pervasive Health Technol. 2020;6(22).
34. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers RM. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging. 2016;35(5):1285–98.
35. Simonyan, K., Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556; 2014.
36. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. Inf Process Manag. 2009;45(4):427–37.
37. Thapar, S., Garg, S. Study and implementation of various morphology based image contrast enhancement techniques. Int J Comput Bus Res. 2012;2229–6166.
38. Torrey, L., Shavlik, J. Transfer learning. In: Handbook of research on machine learning applications and trends: algorithms, methods, and techniques, p. 242–64. IGI Global; 2010.
39. Wan S, Liang Y, Zhang Y. Deep convolutional neural networks for diabetic retinopathy detection by image classification. Comput Electr Eng. 2018;72:274–82.
40. Wang Y, Zhang Y, Yao Z, Zhao R, Zhou F. Machine learning based detection of age-related macular degeneration (AMD) and diabetic macular edema (DME) from optical coherence tomography (OCT) images. Biomed Opt Express. 2016;7(12):4928–40.
41. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. Insights Imaging. 2018;9(4):611–29.
42. Yin X, Ng BW, He J, Zhang Y, Abbott D. Accurate image analysis of the retina using hessian matrix and binarisation of thresholded entropy with application of texture mapping. PLoS ONE. 2014;9(4):e95943.
43. Yoo TK, Hong S. Artificial neural network approach for differentiating open-angle glaucoma from glaucoma suspect without a visual field test. Investig Ophthalmol Visual Sci. 2015;56(6):3957–66.
44. Yoo TK, Park EC. Diabetic retinopathy risk prediction for fundus examination using sparse learning: a cross-sectional study. BMC Med Inform Decis Making. 2013;13(1):106.
45. Zhang W, Zhong J, Yang S, Gao Z, Hu J, Chen Y, Yi Z. Automated identification and grading system of diabetic retinopathy using deep neural networks. Knowl-Based Syst. 2019;175:12–25.