2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Ocular Diseases Detection using Recent Deep Learning Techniques*

Takfarines Guergueb, Moulay A. Akhloufi, *Senior Member IEEE*

***Abstract*— Early fundus screening is a cost-effective and efficient approach to reduce ophthalmic disease-related blindness in ophthalmology. Manual evaluation is time-consuming. Ophthalmic disease detection studies have shown interesting results thanks to the advancement in deep learning techniques, but the majority of them are limited to a single disease. In this paper we propose the study of various deep learning models for eyes disease detection where several optimizations were performed. The results show that the best model achieves high scores with an AUC of 98.31% for six diseases and an AUC of 96.04% for eight diseases.**

I. Introduction

The leading cause of human blindness around the globe is due to retinal fundus diseases [1]. Among the largest ophthalmic diseases are cataract, age related macular degeneration (AMD), glaucoma, and diabetic retinopathy (DR). By 2030, glaucoma patients will reach around 95.4 million patients and the number of people with myopia will increase from 1.95 billion in 2010 to 3.36 billion [2]. Additionally, various studies have identified that individuals with cases of myopia especially in early age or before they turn 20 y/o are more likely to suffer cataracts later [3]. Despite the fact that the actual cause remains unidentified, various research show that the growing axial eyeball length may prohibit nutrient delivery to the hind site of the lenses [3].

Research work on eye diseases detection has gained a lot of attention from the research community. The detection is mainly based on three type of images slit lamp, retro illumination and fundus images [4].

Jing *et al.* [5], proposed a convolutional neural network (CNN) based ensemble model to detect one or more diseases in the fundus images. The method first converts the multi-label classification problem of each individual image into a two-classification problem for each label. Second, transfer learning and ensemble learning techniques are used for addressing the issue of limited dataset. The model is based on two parts, the first part is used for feature extraction which is based on an EfficientNet model. The second part is based on neural networks for multi-label classification problem. Finally, the output probabilities of both models are fused together for the final result. The model was tested on ODIR dataset and the results showed that the proposed model can achieve a good performance even when training with fewer data. The results show that EffecientNetB3 outperforms other models and achieves an accuracy of 90% and an Area Under Curve (AUC) of 67%.

Junjun *et al.* [6], proposed a CNN-based multi-label ocular disease classification model for addressing the problem of correlation between left and right eyes. The proposed framework can process severe ocular diseases. The authors designed a dense correlation network (DCNet) which is based on three modules. For feature extraction, different backbone CNNs are employed, including ResNet-18, 34, 50, and 101. The Spatial Correlation Module (SCM) is used for feature correlation. Finally, the classifier is used for classification and score generation. The model was trained and tested on ODIR dataset using the color fundus photography (CFP) images. The results show that the model using ResNet-101 for feature extraction and SCM achieves the best performance with an AUC of 92.7%.

In this work, we propose the study of recent deep learning architectures, with various data augmentation techniques and fine-tuning to improve the performance of the classification of multiple ocular diseases.

II. Proposed Approach

A. Image dataset

The data used in this study is collected from the Ocular Disease Intelligent Recognition (ODIR) [7]. The ODIR challenge is a structured ophthalmic dataset of 5,000 patients with age, color fundus photographs from left and right eyes and doctors' diagnostic keywords collected by Shanggong Medical Technology Co., Ltd. from different hospitals/medical centers in China.

This competition consist of eight types of ocular diseases and a total of 6,392 images with 2,873 Normal, 1,608 Diabetes, 284 Glaucoma, 293 Cataract, 266 Age Related Macula Degeneration, 128 Hypertension, 232 Pathological Myopia and 708 Other Diseases/Abnormalities.

In our experiments we worked on two scenarios 6 and 8 ocular diseases, where the six ocular diseases are Normal, Glaucoma, Cataract, Age Related Macula Degeneration, Hypertension and Pathological Myopia. Figure 1 shows example images from ODIR 2019 [7].

B. Data augmentation

1) Mixup: Mixup is a data augmentation technique consisting of two parts: random convex combination of raw inputs, and correspondingly, convex combination of one-hot label encodings [8].

$$\begin{aligned} \tilde{x} &= \lambda x_i + (1-\lambda)x_j \\ \tilde{y} &= \lambda y_i + (1-\lambda)y_j \end{aligned} \tag{1}$$

*This research was enabled in part by support provided by the New Brunswick Health Research Foundation (NBHRF).

T. Guergueb and M.A. Akhloufi are with the Perception, Robotics, and Intelligent Machines Research Group (PRIME), Dept of Computer Science, Université de Moncton, Moncton, NB, Canada {etg7520, moulay.akhloufi}@umoncton.ca

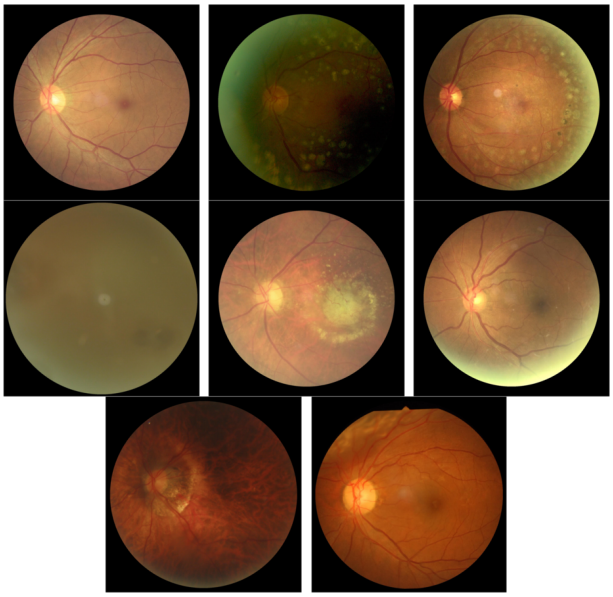

Fig. 1: Example images from ODIR. From up to down, left to right, we have: Normal; Diabetic Retinopathy; Glaucoma; Cataract; Age Related Macula Degeneration; Hypertension; Pathological Myopia; Other Diseases/Abnormalities

where $(\tilde{x}, \tilde{y})$ are the virtual feature-target pair, x_i, x_j are raw input vectors, y_i, y_j are one-hot label encodings and $\lambda \in [0, 1]$.

The main reason for the use of Mixup in our study is that Mixup helps reduce the memorization of corrupt labels, increases the robustness of the model to adversarial examples [8].

2) CutMix: CutMix is a data augmentation that shares some similarity with Mixup which mixes two samples by interpolating both the image and labels [9]. In CutMix the removed regions are replaced with a patch from another image and the ground truth labels are also mixed proportionally to the number of pixels of combined images [9].

The goal of CutMix is to generate a new training sample $(\tilde{x}, \tilde{y})$ by combining two training samples (x_A, y_A) and (x_B, y_B) [9]. The combining operation are defined as:

$$\begin{aligned} \tilde{x} &= \mathbf{M} \odot x_A + (\mathbf{1} - \mathbf{M}) \odot x_B \\ \tilde{y} &= \lambda y_A + (1 - \lambda) y_B, \end{aligned} \tag{2}$$

where $\mathbf{M} \in \{0, 1\}^{W \times H}$ is a binary mask indicating where to drop out and fill in from two images (W, H are the width and height respectively), λ is the he combination ratio, 1 is a binary mask filled with ones and $\odot$ is an element-wise multiplication [9].

CutMix can improve the model robustness, alleviate the model over-confidence and increase training efficiency [9].

3) More data augmentations: In addition to Mixup and CutMix, we used a translation of +- 10, rotations between -5 and +5 degrees with a 1 degree increment.

C. Learning rate scheduler

1) Stochastic gradient descent with warm restarts: Stochastic Gradient Descent with Warm Restarts (SGDR) is an SGD that uses warm restarts instead of learning rate annealing. Where at every new restart the learning rate is initialized to some value and is scheduled to decrease [10]. Essentially, the warm restarts are not performed from scratch but from the parameters of the last step that the model generated during its convergence. Furthermore, we use an aggressive cosine annealing schedule to decrease the learning rate fastly. Mathematically the proposed formula for SGDR is as follows:

$$\eta_t = \eta^i_{min} + \frac{1}{2}(\eta^i_{max} - \eta^i_{min})(1 + \cos(\frac{T_{cur}}{T_i}\pi)) \tag{3}$$

where η^i_{min}, η^i_{max} are ranges for the learning rate, T_{cur} indicates the number of epochs since the last restart and T_i determines the epoch of the next restart [10]. In our work we set the η^i_{max} to 0.0008, the η^i_{min} to 0 and the T_i to 50, meaning that at every 50 epochs a new warm restart is applied.

2) 1cycle: The 1cyclic policy is a cyclical learning rate policy (CLR) for the super-convergence with a slight modification [11]. Where the authors recommend to do one cycle of learning rate. During the training process LR starts increasing from an initial learning rate to the maximum LR for a fixed number of epochs then decreasing to a minimum LR less or equal to the initial learning rate for the remaining epochs [11]. The formula for CLR is as follows:

$$\eta_t = (\eta_{min} + (\eta_{max} - \eta_{min})(max(0, 1 - x)) \tag{4}$$

where x is defined as:

$$x = \left| (\frac{iterations}{stepsize} - 2(cycle) + 1) \right| \tag{5}$$

and $cycle$ can be calculated as:

$$cycle = floor\left(1 + \frac{iterations}{2(stepsize)}\right) \tag{6}$$

where η_{min}, η_{max} are the boundaries of the LR, $iterations$ represents the number of completed mini-batches and $stepsize$ defines one half of a cycle length. The term $(1-x)$ should always be positive [11].

In our study the LR starts increasing from an initial LR of 0.00001 to a maximum LR of 0.00040 for a fixed period of 25 epochs then decreasing for the remaining epochs to a minimum LR of 0.00001. The decay rate is fixed to 0.8, the LR changes during all training process because in our experiments LR sustain is set to 0.

3) Fixed learning rate: Beside of SGDR and 1cycle we tried also different fixed learning rates such as 0.0008, 0.0004, 0.0001, where 0.0008 is the η^i_{max} of the SGDR, 0.0004 is $(\frac{1}{2}(\eta^i_{max}))$ and 0.0001 is $(\frac{1}{8}(\eta^i_{max}))$.

D. Focal Loss

The Focal Loss was proposed for dense object detection to address the extreme imbalance between foreground and background classes during training [12]. The Focal Loss is an extension of the cross-entropy (CE) loss for binary classification:

$$\mathrm{CE}(p, y) = \begin{cases} -\log(p) & \text{if } y = 1 \\ -\log(1-p) & \text{otherwise} \end{cases} \quad (7)$$

Where y specifies the ground-truth class and $y \in \{\pm 1\}$, p is the model's estimated probability for the class with label $y = 1$ and $p \in [0, 1]$ [12]. Mathematically the proposed formula for Focal Loss is defined as:

$$\mathrm{FL}\left(p_{\mathrm{t}}\right) = -\alpha_{\mathrm{t}}\left(1 - p_{\mathrm{t}}\right)^{\gamma} \log\left(p_{\mathrm{t}}\right) \quad (8)$$

where p_t is defined as:

$$p_{\mathrm{t}} = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{otherwise} \end{cases} \quad (9)$$

α is the weight factor to address the class imbalance problem with $\alpha \in \{\pm 1\}$, and γ is tunable focusing parameter. In our experience we set α to 0.25 and γ to 2 according to the work of Tsung-Yi Lin *et al.* where they show that the use of 0.25 and 2 for α and γ respectively achieve better results [12].

E. Deep learning models

A comparative analysis was carried out to compare and evaluate the implementation of various pre-trained models. We chose to work with two architectures EfficientNet [13] and DenseNet [14], [15]. In our experiments, six models were trained and tested on ODIR dataset including EfficientNetB5, EfficientNetB6, EfficientNetB7, DenseNet121, DenseNet169 and DenseNet201.

To improve our models in terms of performance, accuracy and to reduce training time, the hyperparameters of the proposed models, for the multi classification of six diseases, are fine-tuned except the last six layers. Meaning that the model will train the six frozen hidden layers in addition to the final layers. The same process was applied for the eight diseases, where we replace the six frozen layers with eight frozen layers.

Generally CNNs need a large number of data to generalize properly. To handle this issue and to avoid overfitting, a Global Average Pooling layer (GAP) and a dropout of 0.5 are added. A Softmax layer is used for the final prediction to normalize the outputs.

III. Experimental Results

A. Training and test datasets

Some past work [5], [6] used two images left and right of the eye as input. In our work, our models take a single image as input (left or right) because the disease can be present in one eye alone.

The proposed models are trained and evaluated on ODIR dataset [7]. All input images are resized to 512 × 512 pixels. The dataset is randomly divided into 80% for training and 20% for testing. Data augmentation to the training set were used (Mixup, CutMix and other data augmentations as described above). Adam is used as optimizer.

Different Learning Rate (LR) techniques are used such as SGDR and 1cycle to see the impact of each LR on models training. In addition a fixed LR was used based on SGDR boundaries. The models are trained for 100 epochs using a batchsize of 64.

B. Performance evaluation

To evaluate the ocular diseases detection performance of our models, we used the following metrics: Area Under Curve (AUC), Sensitivity (SN), Specificity (SP), and Accuracy (ACC).

Table I shows that a fixed learning rate can help model achieve good results. EfficientNetB7 reached an ACC of 88.85%, a SN of 94.44%, a SP of 94.98% and an AUC of 98.25% with a learning rate of 0.0001.

Table II shows the results on six ocular diseases for different deep learning models using SGDR learning rate, where EfficientNetB7 achieves an interesting result with an ACC of 88.24%, a SN of 88.88%, a SP of 94.11% and an AUC of 98.10%. 1cycle learning rate was very interesting as can be seen in table II, where EfficientNetB7 outperforms other models with an ACC of 87.13%, a SN of 90.74%, a SP of 94.63% and an AUC of 98.31%.

Table III shows the results on eight ocular diseases, EfficientNetB7 was used based on its highest performance on the six diseases (Tables I and II). Overall, 1cycle LR was the best performing and achieved the highest AUC for both six and eight ocular diseases detection. EfficientNetB7 achieves an interesting result with an ACC of 74.98%, a SN of 73.57%, a SP of 85.23% and an AUC of 96.04%.

When comparing with past work dealing with the detection of eight diseases, we can see that our best model achieves state-of-the-art performance with an AUC of 96.04% in comparison to Jing *et al.* [5] with an AUC of 67.01% and Junjun *et al.* [6] with an AUC of 92.7%.

Figure 2 shows images with low score prediction, where our model was unable to predict some images properly due to extreme low light. Also the class Other Diseases/Abnormalities (OD) was problematic as many diseases are present in this class, thus increasing the risk of incorrect predictions.

IV. Conclusion

In this paper, we studied various models for ocular diseases detection, where six different eye diseases were considered including Normal, Glaucoma, Cataract, Age Related Macula Degeneration, Hypertension and Pathological Myopia. We have also tested our best model on eight ocular diseases. A comparative analysis was carried out to compare the impact of some learning rate strategies such as SGDR, 1cycle and different fixed learing rates on our models. Mixup, CutMix and other proposed data augmentations were used to generate more data, to avoid overfitting and help the models generalize properly. Additionally, Focal Loss was used to overcome the uneven distribution of the classes in the dataset.

TABLE I: Results of deep learning models for ocular diseases detection using different fixed LR (0.0008, 0.0004, 0.0001)

Methodes	LR = 0.0008				LR = 0.0004				LR = 0.0001			
	ACC	SN	SP	AUC	ACC	SN	SP	AUC	ACC	SN	SP	AUC
EfficientNetB5	87.78	**94.44**	93.25	97.41	**86.64**	**94.44**	91.86	**97.98**	86.40	92.59	91.69	97.86
EfficientNetB6	86.64	88.88	92.90	**97.77**	85.54	**94.44**	92.90	97.06	87.87	87.03	94.11	98.23
EfficientNetB7	**87.86**	88.88	93.25	97.48	**86.64**	85.18	94.46	97.75	**88.85**	**94.44**	94.98	**98.25**
DenseNet121	84.31	83.33	**96.71**	96.83	86.15	88.88	**96.88**	97.63	87.50	**94.44**	96.19	97.32
DenseNet169	71.69	88.88	76.81	93.44	85.91	87.03	96.36	97.41	85.91	88.88	95.32	97.38
DenseNet201	84.44	87.03	93.94	97.18	**86.64**	88.88	95.84	97.38	87.50	85.18	**97.05**	97.97

TABLE II: Results of deep learning models for six ocular diseases detection for SGDR and 1cycle LR

Methodes	LR = SGDR				LR = 1cycle			
	ACC	SN	SP	AUC	ACC	SN	SP	AUC
EfficientNetB5	87.25	**88.88**	93.07	97.78	**87.99**	**92.59**	94.80	98.27
EfficientNetB6	86.52	**88.88**	92.38	97.75	87.01	**92.59**	93.42	98.18
EfficientNetB7	**88.24**	**88.88**	94.11	**98.10**	87.13	90.74	94.63	**98.31**
DenseNet121	88.11	**88.88**	96.02	98.09	86.76	87.03	96.02	97.62
DenseNet169	86.76	87.03	**96.53**	97.77	87.13	**92.59**	96.02	98.03
DenseNet201	87.13	**88.88**	**96.53**	97.80	86.27	88.88	**97.23**	97.69

TABLE III: Results of deep learning models for eight ocular diseases detection using different LR

Methodes	ACC	SN	SP	AUC	LR
EfficientNetB7	74.82	77.47	80.49	96.01	0.0001
EfficientNetB7	74.98	73.57	85.23	**96.04**	1cycle
EfficientNetB7	**76.70**	**78.67**	**85.94**	95.64	SGDR

The results show that EfficientNetB7 using 1cyle LR is the best performing model with an AUC of 98.31% for six ocular diseases and 96.04% for eight ocular diseases.

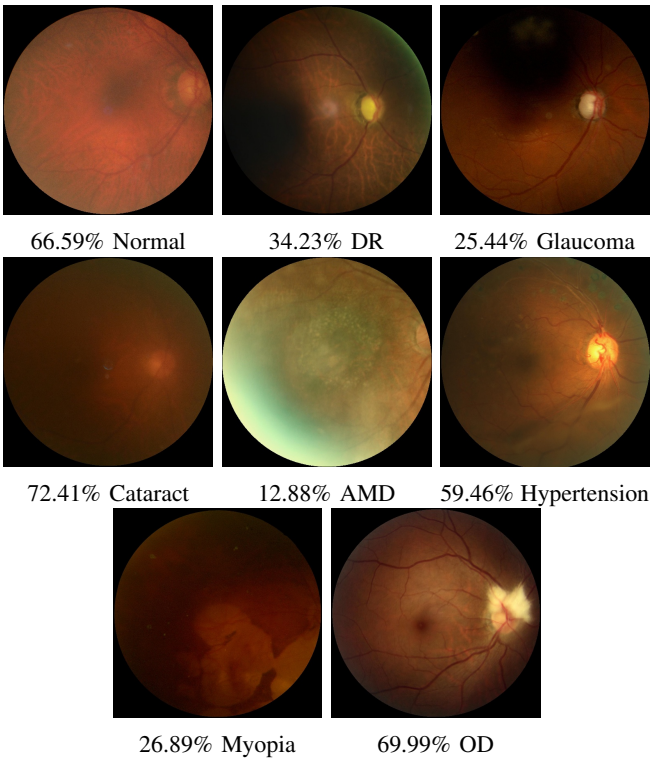

Fig. 2: Bad prediction examples with the obtained score

REFERENCES

[1] H. A. Quigley and A. T. Broman, "The number of people with glaucoma worldwide in 2010 and 2020," *British Journal of Ophthalmology*, p. 262–267, 2006.

[2] World Report On Vision, "The international agency for the prevention of blindness." https://www.iapb.org/advocate/eye-health-and-universal-health-coverage/world-report-on-vision/, online; accessed Apr. 25, 2021.

[3] Myopia Institute, "The link between myopia and cataracts," https://www.myopiainstitute.com/eye-care/the-link-between-myopia-and-cataracts/, online; accessed March 2021.

[4] H. L. Li Xiong and L. Xu., "An approach to evaluate blurriness in retinal images with vitreous opacity for cataract diagnosis," *J Healthc Eng*, vol. 2017, 2017.

[5] L. Y. Jing Wang, W. H. Zhanqiang Huo, and J. Luo., "Multi-label classification of fundus images with efficientnet," *IEEE Access*, vol. 8, p. 212499–212508, 2020.

[6] C. L. Junjun He, Y. Q. Jin Ye, and L. Gu., "Multi-label ocular disease classification with a dense correlation deep neural network," in *Biomedical Signal Processing and Control*, 2021, vol. 63, p. 102167.

[7] ODIR, "Ocular Disease Intelligent Recognition ODIR-5K," https://odir2019.grand-challenge.org/, online; accessed March 2021.

[8] M. C. Hongyi Zhang and D. L.-P. Yann N. Dauphin, "Mixup: Beyond empirical risk minimization," *arxiv.org/abs/1710.09412*, 2017.

[9] D. H. Sangdoo Yun, S. C. Seong Joon Oh, and Y. Y. Junsuk Choe, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, p. 6022–6031.

[10] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," *arxiv.org/abs/1608.03983*, 2016.

[11] L. N. Smith and N. Topin, "Super-convergence: very fast training of neural networks using large learning rates," in *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, T. Pham, Ed., vol. 11006, International Society for Optics and Photonics. SPIE, 2019, pp. 369 – 386. [Online]. Available: https://doi.org/10.1117/12.2520589

[12] P. G. Tsung-Yi Lin, K. H. Ross Girshick, and J. P. Dollár, "Focal loss for dense object detection," in *2017 IEEE International Conference on Computer Vision (ICCV)*. Los Alamitos, CA, USA: IEEE Computer Society, oct 2017, pp. 2999–3007. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/ICCV.2017.324

[13] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *ICML*, 2019.

[14] G. Huang, Z. Liu, G. Pleiss, L. Van Der Maaten, and K. Weinberger, "Convolutional networks with dense connectivity," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[15] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.