Transfer Learning from the Domain of Diabetic Retinopathy to Aid in the Detection of Age-Related Macular Degeneration

Roberto Romero-Oraá[1,2](✉), María Herrero-Tudela[1], Roberto Hornero[1,2], María I. López[1], and María García[1,2]

[1] Biomedical Engineering Group, University of Valladolid, Valladolid, Spain
roberto.romero@uva.es
[2] Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Instituto de Salud Carlos III, Valladolid, Spain

Abstract. Age-related macular degeneration (AMD) is the leading cause of vision loss in the elderly population. Transfer learning has proven useful in fundus image analysis for early diagnosis. Previous models were pre-trained on the ImageNet database. However, a source domain related to retinal diagnosis would facilitate model learning. Our objective was to apply transfer learning from the domain of diabetic retinopathy (DR) to aid in the detection of AMD (binary classification). The proposed model was based on the ResNet-RS architecture. Pre-training aimed at DR diagnosis was conducted using the Kaggle database. Then, fine-tuning was performed using the Automatic Detection challenge on Age-related Macular degeneration (ADAM) dataset. We carried out 3 experiments with different number of images used for fine-tuning. As the main result, our method showed a much faster convergence than the corresponding models pre-trained on ImageNet. Additionally, the proposed source domain was proven especially useful when scarce data in the destination domain was available.

Keywords: Aged-related macular degeneration · Diabetic retinopathy · transfer learning · convolutional neural network · computer-aided diagnosis system

1 Introduction

Age-related macular degeneration (AMD) is the leading cause of vision loss among people aged 50 years and older [1]. This vision loss is irreversible, eventually causing complete blindness [2]. Like most retinal conditions, AMD does not show any external symptoms in its early stages and, thus, continuous ophthalmological examinations are required [3]. In this sense, fundus images are the most cost-effective imaging modality for screening this disease [4], allowing changes in the retina to be detected, associated anomalies to be inspected, and progress to be monitored [3]. However, the high prevalence of AMD and the lack of healthcare resources is overloading the patient care systems, causing treatments to be delayed and error-prone [2]. For this reason, artificial intelligence algorithms have proven to be useful in classifying different types of data, including fundus images [2].

T. Jarm et al. (Eds.): EMBEC 2024, IFMBE Proceedings 113, pp. 345–352, 2024.
https://doi.org/10.1007/978-3-031-61628-0_38

In recent years, deep-learning algorithms are outperforming traditional classifiers [2]. Furthermore, they allow obtaining relevant representations of the data without the need for manual feature extraction [2]. In this context, convolutional neural networks (CNN) are the most popular models for computer vision tasks, including medical image analysis [2]. Although transformers have revolutionized the state of the art of image classification models [5], the latest CNN architectures have managed to surpass their performance in terms of accuracy and scalability [6].

However, the training of deep-learning models require large amounts of annotated data [4]. In medical image analysis, dedicated capturing devices and expert annotations are required [4]. For these reasons, there are no large fundus image databases designed for AMD research that are easily accessible. In this situation, transfer learning aims to help improve the learning from one domain by re-using the knowledge from a related domain [7]. In practice, transfer learning is based on pre-trained models and fine-tuning [8]. Numerous previous studies have followed this approach for the automatic detection of AMD [4]. However, all of them were based on models pre-trained on the ImageNet database [9]. This database consists of tens of millions of annotated images organized by the semantic hierarchy of WordNet [9]. While this dataset has become the de facto standard in transfer learning applications, we hypothesize that using a source domain related to fundus image analysis (closer to our problem) would facilitate model learning for AMD detection. Our objective in this study was to apply transfer learning from the domain of diabetic retinopathy (DR) to aid in the diagnosis of AMD (binary classification problem). We used the Kaggle (EyePACs) fundus image dataset, released for DR research, consisting of more than 80,000 fundus images [10]. This way, the knowledge learned for DR diagnosis using a large fundus image database is transferred to automatically detect AMD. The convergence would be faster and, presumably, more accurate. For the development of the model, we used the ResNet-RS architecture [11], which has never been explored in this field.

2 Fundus Image Databases

The pre-training of the model was conducted using the Kaggle database, aimed at DR diagnosis [10]. Fine-tuning was performed with the Automatic Detection challenge on Age-related Macular degeneration (ADAM) dataset, which was designed for AMD research [4].

2.1 Kaggle (EyePACs)

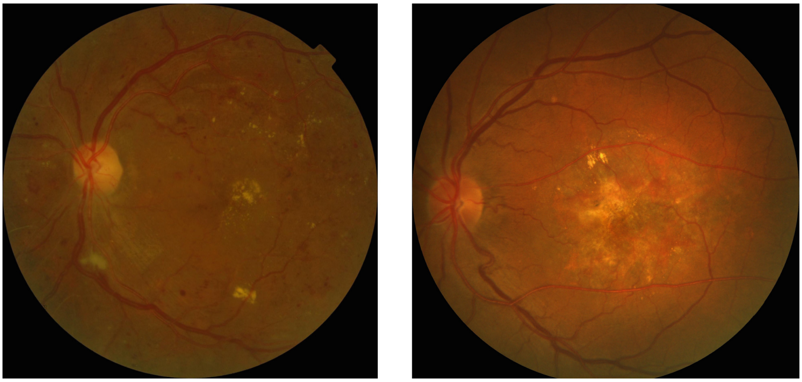
This database was provided by EyePACS for the DR Detection competition published on Kaggle [10]. It is the largest, public fundus image dataset for DR with 88,702 images. However, many of them had poor quality and were discarded using an automatic algorithm from our previous work [12]. Thus, 52,973 images were finally used and divided into a training set of 52,073 images (13,099 with DR and 38,974 with no DR) and a validation set of 900 images (200 with DR and 700 with no DR). All images come from different models and types of cameras. Therefore, a great variety of visual features and resolutions are covered. Figure 1(a) shows an example of a fundus image with DR lesions.

2.2 ADAM

This database, aimed at AMD research, consists of 1,200 fundus images provided by the Zhongshan Ophthalmic Center of the Sun Yat-sen University (China). A Zeiss Visucam 500 fundus camera was used to capture 824 of the images at 2124 × 2056 pixels and a Canon CR-2 device was used to capture the remaining 376 images at 1444 × 1444 pixels. All images had an adequate quality for automatic processing. The dataset was divided into 3 groups: training (400 images), validation (400 images) and test (400 images). Each group had the same number of images with AMD (89 images). Figure 1(b) shows an example of a fundus image with AMD signs.

3 Methods

The proposed method had a preprocessing stage for input normalization. Then, a deep CNN backbone was linked to a set of fully-connected layers. The model was first pre-trained for DR detection and, finally, fine-tuned for AMD classification.



(a) (b)

Fig. 1. (a) Fundus image example showing DR signs from the Kaggle database. (b) Fundus image example with AMD signs from the ADAM database.

3.1 Preprocessing

To normalize the input data and reduce computation time, all images were resized to 512 × 512 pixels [4]. Next, to adapt the pixel values to the input expected by the CNN backbone, the images were normalized in the range [−1,1] [11].

3.2 Data Augmentation

Deep neural networks require a large number of training data. To increase the number of images with which to train the model of this work, the online data augmentation technique was used [2]. This technique allows new synthetic images to be randomly generated in every epoch from the training set. For this task, simple transformations were applied: rotations (±50°), shifts (±7%), flips and scaling (±10%), as in other studies [2].

3.3 Convolutional Neural Network Backbone

We selected the architecture ResNet-RS [11] for the backbone of our CNN model as a feature extractor. This architecture can be seen as an improved version of the canonical ResNet. By introducing a new training methodology and two scaling strategies, ResNet-RS achieves state-of the-art accuracies on ImageNet while showing a much faster performance than previous architectures [11].

3.4 Fully-Connected Layers

At the end of the architecture, 3 fully-connected layers of 1024, 512 and 1 neurons, respectively, were added to act as a binary classifier (AMD / no AMD). We used the ReLu activation function in the first two layers and the sigmoid activation function in the last one. Between each fully-connected layer, the dropout technique was applied with a factor of 0.25 to minimize overtraining [2].

3.5 Pre-training

The proposed model was firstly trained to automatically detect DR. For this task, we used a large set of 52,973 images from the Kaggle database. It should be noted that the model was previously pre-trained on ImageNet to accelerate convergence and facilitate the learning of this stage [9].

As for the training procedure, the model was trained up to a maximum of 200 epochs while applying early stopping when the error on the validation set did not improve during 7 epochs. This parameter, known as patience, was empirically set. Also, the learning rate was reduced by half every time this error reached a minimum and remained constant for 3 epochs [1]. In order to deal with data imbalance, we used the weighted binary cross entropy as the loss function. Finally, the Adam optimization algorithm was applied with an initial learning rate of 0.00005 and a batch size of 4 images [1].

3.6 Fine-Tuning

Having a model capable of accurately detecting DR (source domain), the main idea was to transfer such knowledge to aid in the detection of AMD (destination domain). Since both problems (DR and AMD detection) share the same type of input data (fundus images) and the retinal signs characterizing both conditions might be similar, only a small number of images for fine-tuning is required. Thus, the small training set of 400 images from ADAM was considered in this work.

To evaluate the relevance of our approach based on the size of the training set in the destination domain (AMD detection), we performed two additional experiments by using reduced subsets of the training data (randomly selected). This way, AMD classification was assessed, in 3 different experiments, using N = 50, 100 and 400 images, respectively, as the training set of the fine-tuning stage.

The training procedure for the fine-tuning phase was identical to the one applied for the pre-training step, described in the previous section.

4 Results

The proposed method for AMD detection was evaluated on the 400 images of the test set of the ADAM database in terms of accuracy (*ACC*), sensitivity (*SE*), specificity (*SP*), *F1*-score and area under the receiver operating characteristic curve (*AUC*). Additionally, we measured the time spent in fine-tuning (t). Table 1 shows the results for each experiment in which a different size of training set (N) was considered for fine-tuning. The general performance of the model increased with the value of N. In particular, we achieved an $AUC = 0.89$ when $N = 50$, an $AUC = 0.91$ when $N = 100$, and an $AUC = 0.94$ when $N = 400$. Likewise, the time t became longer as we used a larger N. Thus, $t = 5'24''$ was required with $N = 50$, $t = 9'21''$ was necessary when $N = 100$, and $18'40''$ was spent when $N = 400$. Additionally, balanced values of SE and SP were obtained for every experiment: between 81.35% and 87.64%.

The experiments were computed using a GPU NVIDIA GeForce RTX 4080 with 16 GB GDDR6X memory.

5 Discussion

In this work, we developed an automatic method for AMD detection that exploits the knowledge learned from the domain of DR diagnosis. Since both retinal conditions represent two similar domains with a common type of input data (fundus images), it was expected that the learning was straightforwardly transferrable. To the best of our knowledge, no previous studies have followed this approach.

Table 1. Results on the test set of the ADAM database using the model pre-trained on the Kaggle dataset. N is the number of images used for fine-tunning and t is the computation time spent in such task.

N	*ACC*	*SE*	*SP*	*F1*	*AUC*	*t*
50	82.25	82.02	82.32	0.67	0.89	5′24”
100	82.75	87.64	81.35	0.69	0.91	9′21”
400	86.75	87.14	85.39	0.91	0.94	18′40”

In view of the results exposed in Table 1, we can observe that an $AUC = 0.89$ was surpassed in every experiment. As expected, the greater the number of images used for fine-tuning (N), the greater the performance the model achieved. Nevertheless, even for the scenario where $N = 50$, a remarkable $ACC = 82.25\%$ was obtained. It is also worth noting that the time spent for fine-tuning was very short in all three experiments. A small number of epochs was enough for a fast convergence.

In order to evaluate the relevance of our approach, we conducted the same experiments while using a model pre-trained on the ImageNet database and directly fine-tunned for AMD classification. In this case, the DR domain and the Kaggle database were not involved. The results can be seen in Table 2 and a direct comparison can be established for all experiments. The most notable difference between our approach and the alternative has to do with the computation times. For each value of $N = [50,100,400]$, our method performed about 5 times faster than the corresponding models pre-trained on ImageNet. Since the pre-trained weights of the proposed model were already adapted for fundus image analysis, a smaller number of epochs were required to converge in the fine-running stage. This rapid convergence together with the early stopping technique allowed the computing time to be greatly reduced. As an additional advantage, our approach achieved much more balanced SE and SP values.

When analyzing the experiments with $N = 50$, we can observe how our proposal surpasses the traditional approach in every metric except for SP. In particular, the value of $AUC = 0.89$ is notably higher than the $AUC = 0.78$. A similar conclusion was observed with $N = 100$ although the differences between the computed metrics were smaller. When $N = 400$, however, results using both approaches were very similar. That is, our approach does not imply better performance for this case (even though it is much faster). These comparisons show that the proposed approach is especially useful when scarce data in the destination domain is available.

Our results are in line with those of other related methods in the literature [4]. Only three of the contributions to the ADAM challenge showed a higher AUC (between 0.95 and 0.97) than ours (0.94) using the same test set ($N = 400$) [4]. It is important to note that the mentioned contributions were based on ensemble strategies. Conversely, in this work we designed a simple architecture to focus the experiments on the different transfer learning strategies that were used.

Results on this work are promising. However, some limitations were found. First, a great number of images from the Kaggle dataset were discarded due to poor quality. Even though an automatic algorithm was applied, we would like to measure the impact of the image quality on the results. Second, fine-tuning was conducted using one public dataset only. In the future, it would be desirable to include additional databases in our experiments. Third, there is still room for improvement in the prediction capacity. In this sense, it would be desirable to explore emerging deep learning architectures as well as attention mechanisms and ensemble strategies.

Table 2. Results on the test set of the ADAM database using the model pre-trained on the ImageNet dataset. N is the number of images used for fine-tunning and t is the computation time spent in such task.

N	ACC	SE	SP	$F1$	AUC	t
50	80.75	22.47	97.42	0.34	0.78	26′16”
100	88.25	64.04	95.17	0.71	0.86	50′03”
400	90.50	68.50	96.78	0.76	0.94	90′34”

6 Conclusions

The present work revealed that an appropriate source domain when applying transfer learning helps improve the model effectiveness and efficiency. In contrast to ImageNet pre-training, the use of a pre-trained network aimed at DR diagnosis allowed us to achieve a much faster convergence for AMD detection as well as a more balanced *SE* and *SP* values. Additionally, the proposed approach showed a higher performance when a reduced subset of images was available for fine-tuning purposes. The strategy presented in this study could be useful for a fast development of novel, accurate methods to aid in the diagnosis of AMD. As a result, clinical costs and workload would be reduced while eye care for AMD patients would be improved.

Acknowledgements. This research has been developed under the grants TED2021-131913B-I00, PID2020-115468RB-I00 and PGC2018-098214-A-I00 funded by ’Ministerio de Ciencia e Innovación/Agencia Estatal de Investigación/https://doi.org/10.13039/501100011033/’, European Regional Development Fund (ERDF) A way of making Europe and European Union NextGenerationEU/PRTR; and by ‘CIBER en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN)’ through ‘Instituto de Salud Carlos III’ cofunded with ERDF funds. M. Herrero-Tudela was in receipt of a PIF-UVa grant of the University of Valladolid.

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

References

1. Li, T., et al.: Applications of deep learning in fundus images: a review. Med. Image Anal. **69**, 101971 (2021)
2. Sengupta, S., Singh, A., Leopold, H.A., Gulati, T., Lakshminarayanan, V.: Ophthalmic diagnosis using deep learning with fundus images-a critical review. (2019).https://doi.org/10.1016/j.artmed.2019.101758
3. Iqbal, S., Khan, T.M., Naveed, K., Naqvi, S.S., Nawaz, S.J.: Recent trends and advances in fundus image analysis: a review. Comput. Biol. Med. **151**, 106277 (2022). https://doi.org/10.1016/J.COMPBIOMED.2022.106277
4. Fang, H., et al.: ADAM challenge: detecting age-related macular degeneration from fundus images. IEEE Trans. Med. Imaging **41**, 2828–2847 (2022). https://doi.org/10.1109/TMI.2022.3172773

5. Dosovitskiy, A., et al.: An image is worth 16 × 16 words: transformers for image recognition at scale. In: ICLR 2021 - 9th International Conference on Learning Representations (2020)
6. Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S.: A ConvNet for the 2020s. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2022-June, pp. 11966–11976 (2020). https://doi.org/10.1109/CVPR52688.2022.01167
7. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**, 1345–1359 (2010). https://doi.org/10.1109/TKDE.2009.191
8. 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: Proceedings - 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, CISP-BMEI 2017, 2018-January, pp. 1–11 (2018). https://doi.org/10.1109/CISP-BMEI.2017.8301998
9. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database
10. Kaggle Diabetic Retinopathy Detection Competition. https://www.kaggle.com/c/diabetic-retinopathy-detection/
11. Bello, I., et al.: Revisiting ResNets: improved training and scaling strategies
12. Romero Oraá, R., García, M., Oraá-Pérez, J., López, M.I., Hornero, R.: Automatic fundus image quality assessment: diagnostic accuracy in clinical practice. Invest Ophthalmol Vis Sci **61**, 2033 (2020)