

Automatic Diagnosis of Ocular Toxoplasmosis from Fundus Images with Residual Neural Networks

Rodrigo PARRA^{a,1}, Verena OJEDA^a, José Luis VÁZQUEZ NOGUERA^a, Miguel GARCÍA TORRES^b, Julio César MELLO ROMÁN^c, Cynthia VILLALBA^d, Jacques FACON^e, Federico DIVINA^b, Olivia CARDOZO^f, Verónica Elisa CASTILLO^g and Ingrid CASTRO MATTO^g

^aUniversidad Americana, Asunción, Paraguay

^bDivision of Computer Science, Universidad Pablo de Olavide, ES-41013 Seville, Spain

^cFacultad de Ciencias Exactas y Tecnológicas, Universidad Nacional de Concepción, Concepción, Paraguay

^dFacultad Politécnica, Universidad Nacional de Asunción, San Lorenzo, Paraguay

^eUniversidade Federal do Espírito Santo, Brazil

^fDepartment of Ophthalmology, Hospital General Pediátrico Niños de Acosta Ñu, Paraguay

^gDepartamento de Retina, Cátedra de Oftalmología, Hospital de Clínicas, Facultad de Ciencias Médicas, Universidad Nacional de Asunción, San Lorenzo, Paraguay

Abstract. Ocular toxoplasmosis (OT) is commonly diagnosed through the analysis of fundus images of the eye by a specialist. Despite Deep Learning being widely used to process and recognize pathologies in medical images, the diagnosis of ocular toxoplasmosis(OT) has not yet received much attention. A predictive computational model is a valuable time-saving option if used as a support tool for the diagnosis of OT. It could also help diagnose atypical cases, being particularly useful for ophthalmologists who have less experience. In this work, we propose the use of a deep learning model to perform automatic diagnosis of ocular toxoplasmosis from images of the eye fundus. A pretrained residual neural network is fine-tuned on a dataset of samples collected at the medical center of Hospital de Clínicas in Asunción, Paraguay. With sensitivity and specificity rates equal to 94% and 93%, respectively, the results show that the proposed model is highly promising. In order to replicate the results and advance further in this area of research, an open data set of images of the eye fundus labeled by ophthalmologists is made available.

Keywords. ocular toxoplasmosis, deep learning, residual neural networks, predictive model

¹ Corresponding Author: Rodrigo Parra, Universidad Americana, Brasilia 1100, Asunción, Paraguay; E-mail: rodrigo.parra@ua.edu.py.

1. Introduction

Toxoplasmosis is one of the most common parasitic diseases worldwide, with over a third of the world human population exposed to *Toxoplasma gondii*[1]. Ocular Toxoplasmosis(OT) occurs once the parasite reaches the retina and causes damage to the host cells and neighboring cells leaving primary lesions. OT requires drug treatment to eliminate the parasite and inflammation. Otherwise, it can lead to vision loss [2].

The diagnosis of OT is made by eye exams that look for lesions caused by the disease in eye fundus images. Although clinical manifestations of the disease are generally highly characteristic, atypical manifestations can lead to false-negative errors (made even by experienced ophthalmologists). Lacking a sufficiently sensitive lab test, clinical examination is still considered the diagnostic standard [3].

High seroprevalences of the disease are observed in Latin America and in tropical African countries [4], where the distribution and availability of ophthalmologists are very uneven [5]. Thus, people living in rural areas are more likely to become infected [6], and less likely to be properly diagnosed.

Deep learning models have been used to perform automatic classification of retinal images for medical diagnosis and prognosis with promising results. In particular, convo-lutional neural networks (CNNs) have been applied to identify diseases such as diabetic retinopathy [7] and age-related macular degeneration [8]. Chakravarthy et al. [9] shared promising results applying similar techniques to OT diagnosis using fundus images.

In this study, we investigated whether a particular deep learning architecture (i.e.ResNet18) can be used to assist in the automatic diagnosis of OT. We train and evaluate the model on a dataset of 160 fundus images collected by ophthalmologists at the medical center of Hospital de Clínicas in Asunción, Paraguay. With high rates of sensitivity and specificity, the results obtained show that the proposed model proved to be efficient. Another important contribution is that an open data set of eye fundus images labeled by ophthalmologists has been made available, and can be used to increase reproducibility and compare future work.

2. Methods

In this section, the data set, the algorithms are described, the experiments performed are documented and the results are contextualized.

2.1. Dataset

A data set of 160 fundus images was used to train and evaluate the predictive model. These images were collected and organized by members of the Department of Ophthalmology of the medical center at the Hospital de Clínicas in Asunción (Paraguay).The dataset is available online[10] and is free to use for research purposes. Examples of database images can be seen in Figure 1.

Images were captured by experienced ophthalmologists, using a Zeiss brand camera, model Visucam 500. Each image was manually segmented by an ophthalmologist who manually highlighted OT entities (active lesions and inactive scars). The resulting masks were used to define labels: if an image of the eye fundus contains an OT entity, the patient is considered to be infected and otherwise healthy.

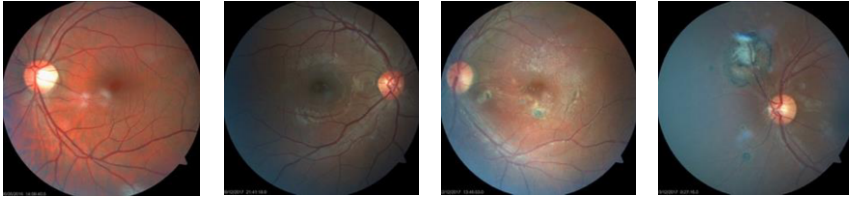


Figure 1. A sample of healthy (a, b) and unhealthy (c, d) retinal fundus images from the dataset.

2.2. Model

Deep learning models have achieved state-of-the-art results in several domains, surpassing other machine learning techniques, with computer vision being one of the most prominent use cases [11]. Transfer learning was applied given our limited availability of data [12]. This approach leverages the fact that similar tasks can often be solved with similar features. The idea is to train a model on a large dataset and then, with minor modifications to the learned weights, apply it to a problem with less data available. This is particularly useful in domains where it is difficult to build a well-annotated dataset on a large scale due to the cost of acquiring data and annotations [13].

A deep residual network (ResNet18) [14], pretrained on the Imagenet dataset [15], was fine-tuned. The final layer of the network was replaced by a softmax layer of two artificial neurons to solve the classification problem. With these adjustments, the model used was able to classify images as healthy or with the presence of OT inflammation.

3. Results

The experiments were performed on a Google Colab Pro account, which provides NvidiaT4 and P100 graphic cards and up to 25 GB of RAM. The residual network was implemented using Pytorch 1.4. The source code of our implementation is available online². The final layer of the Resnet model was replaced with a fully connected layer with appropriate output size and fine-tuned the network for 50 epochs.

The dataset was split into training (70%), validation (10%), and test (20%) sets. The training set was used for model fitting, the validation set for hyperparameter tuning, and the test set to make the final model evaluation.

3.1. Balanced Binary Accuracy

The residual network was fine-tuned with different optimization methods and a batch size of 32. For each optimizer considered, learning rates from 1E-10 to 1E-1 were considered. We ran several experiments with different hyperparameter values. In particular, we considered different batch sizes and learning rates. The results gathered in Table 1, considering accuracy, sensitivity, and specificity, show that the models trained with SGD and momentum provided a higher score in terms of accuracy in the classification.

² <https://colab.research.google.com/drive/1pBDnkVFM14e36ghCVIDfILbVmiKIAbSQ?usp=sharing>

Table 1. Metric values of the best performing models.

Optimizer	Hyperparameters	Accuracy	Sensitivity	Specificity
SGD	learning rate: 1e-2	0.9375	1.0	0.857
SGD with momentum	learning rate: 1e-2 momentum: 0.9	0.9375	0.9444	0.929
RMSProp	learning rate: 1e-2 a: 0.99	0.7825	0.7222	0.857
Adam	epsilon: 1e-8 learning rate: 1e-3 b1: 0.9 b2: 0.99 epsilon: 1e-8	0.90625	0.9444	0.857

3.2. Robustness to imbalance

To better assess the best performing model, it was trained and evaluated with data sets that had different levels of disease prevalence. Specifically, we undersampled the original dataset to generate imbalanced learning scenarios. A weighted loss function was optimized to mitigate the bias towards the majority class. Table 2 groups the results of this experiment in terms of accuracy, sensitivity, and specificity.

Table 2. Metric values of the best performing model on different levels of prevalence.

Sampling ratio	Accuracy	Sensitivity	Specificity
30% healthy / 70% sick	0.913	0.9375	0.857
50% healthy / 50% sick	0.9375	0.9444	0.929
70% healthy / 30% sick	0.826	0.7	0.923

4. Discussion

The results in terms of accuracy are lower for the adaptive optimization methods (RMSProp and Adam) than for the vanilla SGD and Momentum optimizers. This is quite interesting given the popularity of adaptive methods for neural network training. Recently, adaptive learning methods have been shown to generalize worse than SGD when training state-of-the-art deep learning models (as can be seen in our results) [16].

UK standard criteria for retinal screening tests for diabetics require a minimum of 80% to sensitivity and 95% to specificity [17]. It is worth noting that the models trained with SGD and Momentum optimizers can satisfy this criteria by modifying the classification threshold (i.e. choosing a different point along the ROC curve). Although not specific for OT, this result suggests the applicability of our model as a diagnostic support tool. Classifying samples on different levels of data imbalance was harder, as shown by the recorded metric values, but results obtained with the best model are still promising.

It is important to note that our proposed approach was simpler than Chakravarthy's technique [9]. We used a residual network with the full image as its only input, no patch-based model was trained. Another difference was the pretrained network architecture: we used ResNet18, while their work was based on VGG16. Residual networks are easier to optimize, and can gain accuracy from increased depth, which might explain the results.

5. Conclusion

We aimed to investigate whether a pretrained ResNet18 neural network could be used as an aid for automatic OT detection. For this purpose, we performed an experiment in which the model was fine-tuned on a small dataset of eye fundus images. Hyperparameter tuning was performed as part of the model selection process. The models achieved results that met the approval criteria for retinopathy tests for diabetic retinopathy. Additionally, it is important to note that we published an open dataset of fundus images labeled by ophthalmologists at the Hospital de Clínicas from Asunción, Paraguay.

Future work might include (i) adding a third label to provide more information to ophthalmologists (e.g. active/inactive/healthy), (ii) comparing our approach with others, such as the hybrid architecture proposed by Chakravarthy et al., (iii) evaluate classification with imbalanced data considering statistics of OT prevalence.

References

- [1] Tenter AM, Heckerth AR, Weiss LM. *Toxoplasma gondii*: from animals to humans. *Int J Parasitol.* 2000 Nov;30(12-13):1217–1258.
- [2] Park YH, Nam HW. Clinical features and treatment of ocular toxoplasmosis. *Korean J Parasitol.* 2013 Aug;51(4):393–399.
- [3] Garweg JG, de Groot-Mijnes JDF, Montoya JG. Diagnostic approach to ocular toxoplasmosis. *Ocul Immunol Inflamm.* 2011 Aug;19(4):255–261.
- [4] Ozgonul C, Besirli CG. Recent Developments in the Diagnosis and Treatment of Ocular Toxoplasmosis. *Ophthalmic Res.* 2017;57(1):1–12.
- [5] Hong H, Mújica OJ, Anaya J, Lansingh VC, López E, Silva JC. The Challenge of Universal Eye Health in Latin America: distributive inequality of ophthalmologists in 14 countries. *BMJ Open.* 2016 Nov;6(11):e012819.
- [6] Bigna JJ, Tochie JN, Tounouga DN, Bekolo AO, Ymele NS, Youda EL, et al. Global, regional, and country seroprevalence of *Toxoplasma gondii* in pregnant women: a systematic review, modelling and meta-analysis. *Sci Rep.* 2020 Jul;10(1):12102.
- [7] Alyoubi WL, Shalash WM, Abulkhair MF. Diabetic retinopathy detection through deep learning techniques: A review. *Informatics in Medicine Unlocked.* 2020 Jan;20:100377.
- [8] Pead E, Megaw R, Cameron J, Fleming A, Dhillon B, Trucco E, et al. Automated detection of age-related macular degeneration in color fundus photography: a systematic review. *Surv Ophthalmol.* 2019 Jul;64(4):498–511.
- [9] Chakravarthy AD, Abeyrathna D, Subramaniam M, Nguyen QD. An Approach Towards Automatic Detection of Toxoplasmosis using Fundus Images. 2019 Oct.
- [10] Parra R, Ojeda V, Noguera JLV, García-Torres M, Román JCM, Villalba C, et al. Ocular Toxoplasmosis Fundus Images Dataset. Zenodo; Available at: <https://doi.org/10.5281/zenodo.4439566>, Accessed 2021.
- [11] Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep Learning for Computer Vision: A Brief Review. *Comput Intell Neurosci.* 2018 Feb;2018:7068349.
- [12] Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, et al. A Comprehensive Survey on Transfer Learning. 2019 Nov.
- [13] Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C. A Survey on Deep Transfer Learning. 2018 Aug.
- [14] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2015 Dec.
- [15] Deng J, Dong W, Socher R, Li L, Kai Li, Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition; 2009. p. 248–255.
- [16] Wilson AC, Roelofs R, Stern M, Srebro N, Recht B. The Marginal Value of Adaptive Gradient Methods in Machine Learning. 2017 May.
- [17] Mead A, Burnett S, Davey C. Diabetic retinal screening in the UK. *J R Soc Med.* 2001 Mar;94(3):127–129.