Automated Machine Learning Versus Expert-Designed Models in Ocular Toxoplasmosis: Detection and Lesion Localization Using Fundus Images

Abbreviated Title: Automated Machine Learning in Ocular Toxoplasmosis

Daniel Milad, MD[1,2,3], **Fares Antaki,** MD, CM[1,2,3,4], **Allison Bernstein**, MD[1,2,3] , **Samir Touma,** MD[1,2,3] **Renaud Duval**, MD[1,2]

1. Department of Ophthalmology, Université de Montréal, Montreal, Québec, Canada
2. Centre Universitaire d'Ophtalmologie (CUO), Hôpital Maisonneuve-Rosemont, CIUSSS de l'Est-de-l'Île-de-Montréal, Montreal, Quebec, Canada
3. Department of Ophthalmology, Centre Hospitalier de l'Université de Montréal (CHUM), Montreal, Quebec, Canada
4. The CHUM School of Artificial Intelligence in Healthcare (SAIH), Centre Hospitalier de l'Université de Montréal (CHUM), Montreal, Quebec, Canada

ORCID of the authors: Daniel Milad (0000-0002-0693-3421), Fares Antaki (0000-0001-6679-7276), Samir Touma (0000-0002-6365-0946) Renaud Duval (0000-0002-3845-3318)

Corresponding author: Renaud Duval, MD, CM; Centre Universitaire d'Ophtalmologie (CUO), Hôpital Maisonneuve-Rosemont, CIUSSS de l'Est-de-l'Île-de-Montréal, 5415 Boulevard de l'Assomption, Montréal, Québec, Canada, H1T 2M4. Email: renaud.duval@gmail.com

Meeting Presentations: Podium presentation, 2023 Canadian Ophthalmological Society Annual Meeting, June 2023, Quebec City, Quebec, Canada

Financial Disclosures: D. Milad, None; F. Antaki, Bayer (Grant); A. Bernstein, None; Samir Touma, Bayer (Grant); R. Duval, Roche (Consultant), Novartis (Research grant), Bayer (Consultant, Education grant), Apellis (Consultant), Biogen (Consultant).

Funding Support: None
Conflict of interest: No conflicting relationship exists for any author.
Word Count: 2752
Number of Figures: 4
Number of Tables: 3
Number of Supplemental Figures: 0
Number of Supplemental Tables: 2

Summary Statement

Automated machine learning excels at detecting ocular toxoplasmosis lesions in fundus images, performing on par with expert-designed bespoke models. Automated object detection models can also visually identify these lesions with strong performance metrics

Keywords

Ocular toxoplasmosis, Public Health, Artificial Intelligence, Deep Learning

Abstract

Purpose: Automated machine learning (AutoML) allows clinicians without coding experience to build their own deep learning (DL) models. This study assesses the performance of AutoML in detecting and localizing ocular toxoplasmosis (OT) lesions in fundus images and compares it to expert-designed models.

Methods: Ophthalmology trainees without coding experience designed AutoML models using 304 labelled fundus images. We designed a binary model to differentiate OT from normal and an object detection model to visually identify OT lesions.

Results: The AutoML model had an area under the precision-recall curve (AuPRC) of 0.945, sensitivity of 100%, specificity of 83% and accuracy of 93.5% (vs. 94%, 86% and 91% for the bespoke models). The AutoML object detection model had an AuPRC of 0.600 with a precision of 93.3% and recall of 56%. Using a diversified external validation dataset, our model correctly labeled 15 normal fundus images (100%) and 15 OT fundus images (100%), with a mean confidence score of 0.965 and 0.963, respectively.

Conclusion: AutoML models created by ophthalmologists without coding experience were comparable or better than expert-designed bespoke models trained on the same dataset. By creatively using AutoML to identify OT lesions on fundus images, our approach brings the whole spectrum of DL model design into the hands of clinicians.

Ocular toxoplasmosis (OT), caused by the *Toxoplasma gondii* infection, is the leading cause of posterior infectious uveitis. The prevalence of OT varies significantly, ranging from 1% to 85% according to various studies.[1–8] Higher prevalence rates are observed in regions such as South America, Latin America, Africa, and parts of Asia, where it is associated with poor environmental and socioeconomic conditions.[9] The *Toxoplasma gondii* parasite can be transmitted congenitally but is most often acquired through the ingestion of undercooked meat, contaminated vegetables, soil, or water containing oocysts.[9] Toxoplasmic retinochoroiditis is the most common presentation, consisting of a focal area of intense retinitis with moderate to severe overlying vitreous inflammation, with recurring active disease indicated by an adjacent chorioretinal scar.[9,10]

The ability to identify OT lesions is essential since accurate diagnosis guides treatment, especially for immunosuppressed patients, pregnant individuals, those with extensive lesions, and cases of retinitis that threaten the macula.[11] Amongst uveitis specialists, agreement on clinical diagnosis of OT is moderate to good, but sensitivity and specificity estimates vary greatly. [12,13] Although there is no widely agreed upon treatment regimen for OT, the traditional approach includes a combination of pyrimethamine, sulfadiazine, and prednisone taken systemically, or clindamycin with dexamethasone administered intravitreally.[14] The management of OT in regions with constrained healthcare,

especially prevalent in developing countries, faces hurdles regarding limited professional expertise and medication access. Nonetheless, the potential for automated OT diagnosis using artificial intelligence (AI), specifically machine learning (ML), may help augment diagnostic accuracy and streamline patient management.

Historically, the development of ML models was limited to AI experts possessing coding proficiency and access to significant computing resources, which are frequently unavailable to most clinicians. Automated machine learning platforms (AutoML) are among the latest technologies that have enabled democratization of AI development, providing clinicians with simple graphical interfaces to achieve comparable outcomes. AutoML models have demonstrated similar diagnostic performances to expert-designed (bespoke) models in various data types including electronic medical records, fundus photos and surgical videos.[15–21] These models can be incorporated into smartphones, enabling offline diagnosis using the device's camera and connectivity features. We have previously demonstrated that this approach is particularly beneficial for the diagnosis of trachoma, for example, as it facilitates greater accessibility to diagnostic tools in regions with limited healthcare access.[22]

In this paper, we showcase how it is possible to construct an inexpensive AutoML deep learning (DL) model with the ability to detect and localize OT lesions in fundus photos. To the best of our knowledge, this study is the first to demonstrate the feasibility of such models in the diagnosis of infectious uveitis.

Methods

Ophthalmology trainees without coding experience designed DL models using Google Cloud AutoML Vision to diagnose and localize OT lesions from a public fundus image dataset.[23] The study's high-level flow chart is detailed in **Figure 1**.

1. Dataset

We utilized an open dataset of fundus images from Paraguay published by Parra and colleagues in May 2021.[23] The images were obtained by the Department of Ophthalmology of the Hospital de Clínicas in Asunción, captured by experienced ophthalmologists using a Zeiss Visucam 500 camera. An ophthalmologist, with access to patients' clinical history and ancillary tests, labeled the images, representing the ground truth in the dataset. After reviewing the dataset for accuracy, no data was removed and we included 304 labeled fundus images. Although Parra et al.'s study only included 160 of the 304 published images, attempts to contact the authors to determine which images were used in their study were unsuccessful.

Data set preparation before model development

We uploaded the images to the Google Cloud platform without performing any image augmentation. Examples of fundus photos used for training are shown in **Figure 2**. For the object detection model, we included fundus images centered on the posterior pole with visible OT lesions. We did not selectively include pathology and rather included images covering a range of OT lesions from least to

most evident. Despite the presence of various degrees of vitreous haze, some limiting the visualization of the lesions, AutoML was able to identify the lesions. Two authors (DM, AB) placed bounding boxes on the OT lesions, and any differences were reviewed to reach a consensus, establishing the ground truth for our object detection model.

2. Model development

We developed models using Google Cloud AutoML, employing single-label classification for the first model to assign a single label (OT or normal) to each classified image. To ensure the meaningfulness and benchmarking of our results, our model measured the performance of OT against normal, thus enabling a comparison with Parra and colleagues' bespoke model. [23] Since there was a 62% prevalence of OT in our data set, we compared our model's performance to Parra et al.'s best performing model with a sampling ratio of 30% healthy / 70% sick. [23] For the second model, image object detection was used to localize OT lesions on the image. Google Cloud AutoML automatically splits the dataset, using approximately 80% for training, 10% for validation, and 10% for testing. Currently, the AutoML platform automatically optimizes hyperparameters for model development. Attempts to contact the authors for their train/validation/test splits were unsuccessful. Based on previous evidence of reasonable repeatability for AutoML training, we performed each experiment once. [15,24] Early stopping was automatically enabled on training to avoid overfitting.

3. External Validation

Open datasets, AutoML, and cloud-based platforms for AI deployment democratize access to AI in ophthalmology.[25] We searched the Google Dataset Search engine for public datasets for external validation but found none suitable. Since OT is not prevalent in Canada, producing a dataset was not feasible. This is in keeping with a recent systematic review of all potential data sets for ophthalmological imaging, none of which were suited for our study.[26] We compiled images from Google Images search engine, obtaining 15 normal fundus images and 15 OT fundus images, which were evaluated using our model. To validate the images displaying OT, authors agreed upon the diagnosis and verified the sources of each image. The web addresses, ground truth, and predictions of the external validation set are provided in Supplemental Table 1. The OT fundus images were also used to externally validate the object detection model.

4. Statistical analysis and performance metrics

Google Cloud AutoML provides the area under the precision-recall curve (AuPRC), a performance metric ranging from 0 to 1.0, with higher values indicating more accurate models. Confusion matrices, sensitivity (recall), and positive predictive value (PPV) for the overall classifier and each classification subtask are also reported.[27]

To compare binary classification tasks, sensitivity and specificity were obtained at the same threshold as in Parra and colleagues' model. Due to the absence of confidence intervals, we qualitatively compared the point estimate of sensitivity and specificity of the AutoML models to those in the

literature. Specificity was manually calculated from the overall confusion matrix using the formula: true negative / (true negative + false positive). [27]

For the object detection AutoML model, AuPRC was reported but could not be compared to the literature. Other performance metrics included precision (the percentage of correct predictions) and recall (the percentage of all ground truth items successfully predicted by the model). The intersection over union (IoU) threshold value, set to 0.2, determines which predictions to return; the higher the threshold, the closer the predicted bounding box values must be to the ground truth bounding box values. During the model's development, we systematically explored various IoU threshold values through an iterative process. This experimentation revealed that increasing the IoU threshold resulted in marginal changes in bounding box dimensions, yet there was a noticeable decrease in accuracy. At an IoU threshold of 0.2, the bounding boxes demonstrated the highest level of accuracy whilst maintaining reasonable bounding box dimensions.

5. Saliency maps

We developed saliency maps using four representative, correctly predicted images from the external validation dataset with Google Cloud's AutoML. These maps enabled visualization of areas in the images that contributed most to disease classification. We employed a region-based saliency method called eXplanation with Ranked Area Integrals, which works well with natural images like fundus images. [28] The viridis color map indicated areas of lesser influence in blue and greater influence in yellow.

Results

We included 304 labeled fundus images, with 187 (62%) having active or inactive OT and 117 (38%) being normal. For the object detection model, we included all labeled OT fundus images (n=187).

At the 0.5 confidence threshold cut-off, the binary AutoML models performed as well as or better than the bespoke models (Tables 1 and 2). Our AutoML single-label classification model had an AuPRC of 0.945, sensitivity of 100%, specificity of 83%, and accuracy of 93.5% (versus 94%, 86%, and 91% for the bespoke models). Saliency maps highlighting areas of interest in correctly predicted images can be found in **Figure 3**. Our AutoML object detection model had an AuPRC of 0.600, with a precision of 93.3% and recall of 56% (**Figure 4**).

Using the external validation dataset, our binary model correctly labeled 15 (100%) normal fundus images and 15 (100%) OT fundus images, with mean confidence scores of 0.965 and 0.963, respectively (Table 3). Our AutoML object detection model also highlighted all (100%) OT fundus lesions at the 0.5 confidence threshold cut-off (Supplemental Table 2). At a 0.5 confidence threshold cut-off, there were 15 (100%) true positives, 0 (0%) false positives, and 0 (0%) false negatives.

Discussion

In this pilot study, we demonstrate the value of AutoML in democratizing AI, making it accessible to clinicians. We showed the feasibility of using code-free platforms by ophthalmologists to create ML models that detect and identify OT lesions from fundus images. The model's effectiveness was evaluated by comparing its performance to bespoke models designed by AI experts, trained and validated on the same dataset. Our AutoML models performed well, with metrics similar to or better than those designed by AI experts. Additionally, we created an object detection model to visually identify OT fundus lesions for clinicians. Although these findings showcase the potential of AI applications in improving patient care, it is crucial to thoroughly evaluate these applications before implementing them in clinical settings. Our study serves as a proof of concept for the design strategy combining binary and object detection, introducing a new dimension of AI model design into the hands of clinicians.

Clinical findings currently guide OT diagnosis, with features such as inflammation and retinochoroiditis being characteristic. [9] However, diagnosis can be subjective due to varying clinical presentations, leading to differences in prevalence across studies. Among uveitis specialists, agreement on OT fundus image interpretation is moderate to good, but sensitivity and specificity estimates vary greatly. [12,13] In unusual cases, serologies or aqueous humor tap analyses may confirm the diagnosis. [29] Standardizing the OT diagnostic process is crucial, as blindness rates can reach up to 24%.[13] Using AI models could standardize and aid clinicians' diagnostic process. While the current model is limited to detecting OT lesions without distinguishing active from inactive lesions, it serves as a promising proof-of-concept for deep learning in uveitis diagnosis. This was done to compare our results to Parra and colleagues' bespoke model. [23] By developing models that differentiate active from

inactive lesions, as well as other uveitis etiologies using larger datasets, we could significantly improve diagnostic accuracy and speed. These models hold the potential to enhance patient outcomes by offering a consistent standard for OT diagnosis, thereby mitigating diagnostic variability. This uniformity could help guide clinicians' management strategies.

For OT, retinal imaging is a cornerstone in the diagnosis and consequent treatment of the condition. As such, fundus image data sets can be used to program ML models, such as the ones in this study. Our study used fundus images from a medical center in Paraguay that had only previously been used in a similar context. The comparable study used a ResNet18 neural network training program on the data set and found results of 93% specificity and 94% sensibility.[23] There is a scarcity of research on the application of AI in diagnosing OT, with only a small number of studies utilizing DL and more complex Convolutional Neural Networks (CNN) dual-input models. To date, there has been no investigation into the effectiveness of these bespoke models compared to the emerging technique of AutoML.[23,30] Developing and optimizing such models demands a high level of expertise and extensive computing resources. While these models offer the flexibility to build custom-made solutions to specific clinical settings and incorporate domain knowledge to improve performance, they require a significant investment of time, effort, and finances. This renders them inaccessible to many clinicians and researchers with limited resources and expertise. In contrast, AutoML provides a user-friendly and accessible interface that allows for the rapid generation of ML models, without necessitating programming skills or significant resources. AutoML algorithms automatically optimize hyperparameters, choose algorithms, and tune models to achieve optimal performance. This allows for the production of high-performing models in a cost-effective, accessible and time-efficient

manner. Therefore, while bespoke models remain a valuable approach, AutoML offers a promising alternative for clinicians and researchers seeking to incorporate AI in their practice.

Our study presents a pioneering approach in developing an object detection DL model for OT, which to our knowledge, has not been done before. This object detection model holds potential for clinicians, providing visual feedback that enables localization of anomalies with increased accuracy and enables timely intervention when necessary. This approach represents a significant step forward in image-based diagnosis and illustrates the versatility and adaptability of DL models in clinical settings. The results obtained from applying the object detection model to the external validation dataset are of particular interest. As presented in Supplemental Table 2, the model successfully identified all visible OT fundus lesions by outlining them with bounding boxes. Notably, the model exhibited an initially high false positive rate, indicating that some bounding boxes were incorrectly attributed to lesions. However, a key feature of the model is that it provides clinicians with an indication of its level of confidence for each bounding box. When a bounding box was a false positive, the model emitted a very low confidence score. Thus, when applying a 0.5 confidence threshold cut-off, there were no false positives. This approach empowers clinicians to exercise their clinical judgment and carefully examine areas flagged as potentially containing lesions, ensuring that no lesions are overlooked.

Our results emphasize the value of open-access big data in addressing global public health issues. With a decentralized dataset, collaborators worldwide can tackle multiple issues simultaneously. A strength of our study is the use of an open-access dataset and a detailed explanation of our

methodology, allowing researchers to reproduce our results. We benchmarked our results by comparing them to expert-designed models trained on the same dataset. In a move towards transparency and model explainability, we generated saliency maps highlighting areas of interest from correctly predicted cases. However, these maps do not fully explain the model's decision-making process, and their usefulness remains to be proven. [31] Our findings are limited by the scale of external validation since OT is not prevalent in Canada. To address this, we used normal and OT fundus images found on Google Images. External validation in a region affected by OT would be crucial to evaluate our application's performance in a real-world setting. Additionally, we did not train or challenge our model with pathologies from the differential diagnosis of OT, such as tuberculosis, viral retinitis, toxocariasis, or syphilis. Investigating this aspect is crucial before considering any clinical applications. Moreover, we chose a low IoU threshold to detect and localize lesions within clinically meaningful margins. This approach, however, limited our capacity to evaluate discriminative performance with precise lesion margin delineation, as typically done in segmentation studies. Future work could explore segmenting OT lesions to distinguish active from inactive lesions, which may prove beneficial in clinic. Lastly, AutoML empowers clinicians to develop ML models, despite lacking coding proficiency. Nevertheless, it is imperative to exercise caution in their clinical implementation. These models require thorough validation by experienced AI engineers prior to their application in clinical settings.

In summary, we demonstrated the feasibility of DL model design by ophthalmologists without coding experience for the diagnosis of OT from fundus images. The AutoML models were comparable or better than the bespoke models trained on the same dataset. We also developed an object detection

model, visually identifying the lesion for the clinician. Thus, we showed how creative use of AutoML democratises access to the world of ML and brings useful tools into the hands of clinicians. This approach has the potential to bring world-class diagnostic technology to the fingertips of underserved populations worldwide.

Bibliography

1. Kamani J, Mani AU, Egwu GO, Kumshe HA. Seroprevalence of human infection with Toxoplasma gondii and the associated risk factors, in Maiduguri, Borno state, Nigeria. Ann Trop Med Parasitol. 2009 Jun;103(4):317–21.

2. Swai ES, Schoonman L. Seroprevalence of Toxoplasma gondii infection amongst residents of Tanga district in north-east Tanzania. Tanzan J Health Res. 2009 Oct;11(4):205–9.

3. Jones JL, Kruszon-Moran D, Sanders-Lewis K, Wilson M. Toxoplasma gondii infection in the United States, 1999 2004, decline from the prior decade. Am J Trop Med Hyg. 2007 Sep;77(3):405–10.

4. Silveira C, Belfort R, Muccioli C, Abreu MT, Martins MC, Victora C, et al. A follow-up study of Toxoplasma gondii infection in southern Brazil. Am J Ophthalmol. 2001 Mar;131(3):351–4.

5. Terazawa A, Muljono R, Susanto L, Margono SS, Konishi E. High Toxoplasma antibody prevalence among inhabitants in Jakarta, Indonesia. Jpn J Infect Dis. 2003 Jun;56(3):107–9.

6. Shin DW, Cha DY, Hua QJ, Cha GH, Lee YH. Seroprevalence of Toxoplasma gondii infection and characteristics of seropositive patients in general hospitals in Daejeon, Korea. Korean J Parasitol. 2009 Jun;47(2):125–30.

7. Udonsom R, Lekkla A, Chung PTT, Cam PD, Sukthana Y. Seroprevalence of Toxoplasma gondii antibody in Vietnamese villagers. Southeast Asian J Trop Med Public Health. 2008 Jan;39(1):14–8.

8. Bahia-Oliveira LMG, Jones JL, Azevedo-Silva J, Alves CCF, Oréfice F, Addiss DG. Highly endemic, waterborne toxoplasmosis in north Rio de Janeiro state, Brazil. Emerg Infect Dis. 2003 Jan;9(1):55–62.

9. Maenz M, Schlüter D, Liesenfeld O, Schares G, Gross U, Pleyer U. Ocular toxoplasmosis past, present and new aspects of an old disease. Prog Retin Eye Res. 2014 Mar;39:77–106.

10. Smith JR, Ashander LM, Arruda SL, Cordeiro CA, Lie S, Rochet E, et al. Pathogenesis of ocular toxoplasmosis. Prog Retin Eye Res. 2021 Mar;81:100882.

11. Butler NJ, Furtado JM, Winthrop KL, Smith JR. Ocular toxoplasmosis II: clinical features, pathology and management. Clin Experiment Ophthalmol. 2013;41(1):95–108.

12. Stanford MR, Gras L, Wade A, Gilbert RE. Reliability of expert interpretation of retinal photographs for the diagnosis of toxoplasma retinochoroiditis. Br J Ophthalmol. 2002 Jun;86(6):636–9.

13. Bosch-Driessen LEH, Berendschot TTJM, Ongkosuwito JV, Rothova A. Ocular toxoplasmosis: clinical features and prognosis of 154 patients. Ophthalmology. 2002 May;109(5):869–78.

14. Feliciano-Alfonso JE, Muñoz-Ortiz J, Marín-Noriega MA, Vargas-Villanueva A, Triviño-Blanco L, Carvajal-Saiz N, et al. Safety and efficacy of different antibiotic regimens in patients with ocular toxoplasmosis: systematic review and meta-analysis. Syst Rev. 2021 Jul 19;10(1):206.

15. Korot E, Guan Z, Ferraz D, Wagner SK, Zhang G, Liu X, et al. Code-free deep learning for multi-modality medical image classification. Nat Mach Intell. 2021 Mar 1;3(4):288–98.

16. O'Byrne C, Abbas A, Korot E, Keane PA. Automated deep learning in ophthalmology: AI that can build AI. Curr Opin Ophthalmol. 2021 Sep 1;32(5):406–12.

17. Faes L, Wagner SK, Fu DJ, Liu X, Korot E, Ledsam JR, et al. Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study. Lancet Digit Health. 2019 Sep;1(5):e232–42.

18. Antaki F, Coussa RG, Hammamji K, Duval R. Revisiting the Problem of Optic Nerve Detection in a Retinal Image Using Automated Machine Learning. Asia-Pac J Ophthalmol Phila Pa. 2021 Jun 1;10(3):335–6.

19. Antaki F, Coussa RG, Kahwati G, Hammamji K, Sebag M, Duval R. Accuracy of automated machine learning in classifying retinal pathologies from ultra-widefield pseudocolour fundus images. Br J Ophthalmol. 2023 Jan;107(1):90–5.

20. Antaki F, Kahwati G, Sebag J, Coussa RG, Fanous A, Duval R, et al. Predictive modeling of proliferative vitreoretinopathy using automated machine learning by ophthalmologists without coding experience. Sci Rep. 2020 Nov 11;10(1):19528.

21. Touma S, Antaki F, Duval R. Development of a code-free machine learning model for the classification of cataract surgery phases. Sci Rep. 2022 Feb 14;12(1):2398.

22. Milad D, Antaki F, Robert MC, Duval R. Development and deployment of a smartphone application for diagnosing trachoma: Leveraging code-free deep learning and edge artificial intelligence. Saudi J Ophthalmol [Internet]. [cited 2023 Mar 16];Ahead of Print. Available from: https://www.saudijophthalmol.org/preprintarticle.asp?id=369836;type=0

23. Parra R, Ojeda V, Vázquez Noguera JL, García Torres M, Mello Román JC, Villalba C, et al. Automatic Diagnosis of Ocular Toxoplasmosis from Fundus Images with Residual Neural Networks. Stud Health Technol Inform. 2021 May 27;281:173–7.

24. D'Amour A, Heller K, Moldovan D, Adlam B, Alipanahi B, Beutel A, et al. Underspecification Presents Challenges for Credibility in Modern Machine Learning. 2020 [cited 2023 Mar 16]; Available from: https://arxiv.org/abs/2011.03395

25. Khan SM, Liu X, Nath S, Korot E, Faes L, Wagner SK, et al. A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability. Lancet Digit Health. 2021 Jan;3(1):e51–66.

26. Trevethan R. Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. Front Public Health. 2017;5:307.

27. Davis J, Goadrich M. The relationship between Precision-Recall and ROC curves. In: Proceedings of the 23rd international conference on Machine learning - ICML '06 [Internet]. Pittsburgh, Pennsylvania: ACM Press; 2006 [cited 2023 Mar 16]. p. 233–40. Available from: http://portal.acm.org/citation.cfm?doid=1143844.1143874

28. Kapishnikov A, Bolukbasi T, Viégas F, Terry M. XRAI: Better Attributions Through Regions. 2019 [cited 2023 Mar 18]; Available from: https://arxiv.org/abs/1906.02825

29. Fardeau C, Romand S, Rao NA, Cassoux N, Bettembourg O, Thulliez P, et al. Diagnosis of toxoplasmic retinochoroiditis with atypical clinical features. Am J Ophthalmol. 2002 Aug;134(2):196–203.

30. Chakravarthy AD, Abeyrathna D, Subramaniam M, Chundi P, Halim MS, Hasanreisoglu M, et al. An Approach Towards Automatic Detection of Toxoplasmosis using Fundus Images. In: 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE) [Internet]. Athens, Greece: IEEE; 2019 [cited 2023 Mar 16]. p. 710–7. Available from: https://ieeexplore.ieee.org/document/8941984/

31. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, et al. Artificial intelligence and deep learning in ophthalmology. Br J Ophthalmol. 2019 Feb;103(2):167–75.

Figure Legends

Figure 1: AutoML OT Models Development Flow Chart

OT ocular toxoplasmosis, *N* normal

Figure 2: Examples of ocular toxoplasmosis classification of selected fundus images

Representative fundus images of OT and normal fundi. (A) Normal fundus image. (B, C) Cases of recurring retinochoroiditis with an adjacent chorioretinal scar. (D) Case of active OT retinochoroiditis along the superior vascular arcade.

OT ocular toxoplasmosis

Figure 3. Saliency maps of correctly predicted cases of normal fundus images and OT fundus images using the AutoML Single-label classification model.

(A) Normal fundus image. (B) Saliency map of the normal image highlighting the vessels as the most important region for the predicted class; 100% confidence score. (C) Fundus image with active OT lesions. (D) Saliency map of the OT image highlighting the area of retinochoroiditis and macular exudates as the most important regions for the predicted class; 100% confidence score.

XRAI eXplanation with Ranked Area Integrals*, OT* ocular toxoplasmosis

Figure 4. Examples of ocular toxoplasmosis classification of selected fundus images using the AutoML Object Detection Model

(A) Correctly predicted active OT lesion with a confidence score of 0.935. The model creates a visual box surrounding the lesion, seen in blue. (B) Fundus image of an OT lesion that was not detected with strong enough confidence (confidence score of 0.245) to be considered a lesion. The bounding box seen in blue is the ground truth outline by the clinicians. (C) Correctly predicted OT lesion with a confidence score of 0.992. (C) Ground truth label of fundus image (C), showing the similarity between our ground truth and the object detection model's prediction.

OT ocular toxoplasmosis