MEDWIN PUBLISHERS
Committed to Create Value for Researchers

Open Access Journal of Ophthalmology
ISSN: 2578-465X

Development and Validation of a Code-Free Deep Learning Model for the Screening of Pathological Glaucoma and Glaucoma on Color Fundus Photographs

Carolyn YTW[1,2,3]*, Timing L[1,2], Tin LW[3] and Henry HWL[4]

[1]Institute of Ophthalmology, University College London, United Kingdom
[2]Moorfields Eye Hospital NHS Foundation Trust, United Kingdom
[3]Faculty of Medicine, The Chinese University of Hong Kong, China
[4]Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, China

Research Article
Volume 9 Issue 2
Received Date: July 10, 2024
Published Date: July 30, 2024
DOI: 10.23880/oajo-16000318

***Corresponding author:** Carolyn Yu Tung Wong, Prince of Wales Hospital, 30-32 Ngan Shing Street, Shatin, New Territories, Hong Kong, Email: carolcarol.carolyn@gmail.com

Abstract

Purpose: Myopia and glaucoma are major causes of vision impairment that are expected to rise in incidence in East Asia. Artificial intelligence (AI)-aassisted mass screenings may help reduce disease burden and improve long-term prognosis. CFDL is a deep learning (DL) subtype that facilitates non-AI-experts to derive their own AI models. With its resource-saving nature and ease of development, it may benefit resource-limited screening settings. This study evaluated the performance of CFDL in performing pathological myopia (PM) and glaucoma screening on colour fundus photographs (CFP)s.

Methods: We used labelled CFPs from the ODIR-K dataset to develop our CFDL algorithm using Google's CFDL platform i.e. Vertex AI. 3374 normal, PM and glaucoma CFPs were identified and uploaded to Vertex AI. The uploaded images were split into 8-1-1 for training, validation, and testing. Our model's performance was later compared to the state-of-the-art DL models identified through our targeted literature search. External validation of the model was performed on an independent cohort of CFPs retrieved from other datasets.

Results: At the 0.5 confidence threshold, our CFDL model achieved an area under receiver operator curve (AUROC) of 0.998, accuracy of 90.74% and recall of 90.74%. The sensitivity ranged from 94.44% (PM detection) to 40% (glaucoma detection). When externally validated, the model had a lower AUROC (0.863), accuracy (77.78%), and recall (77.78%) at the 0.5 confidence threshold.

Conclusion: The study demonstrated the feasibility of a highly accurate CFDL model for PM and glaucoma screenings on CFPs.

Keywords: Artificial Intelligence; Image Segmentation; Retinal Blood Vessel; Colour Fundus Photograph

Abbreviations

AI: Artificial Intelligence; DL: Deep Learning; CFDL: Code-Free Deep Learning; CFPs: Colour Fundus Photos; ODIR-K: Ocular Disease Intelligent Recgonition; AUROC: Area under Receiver Operator Curve; CIs: Confidence Intervals; PALM: Pathologic Myopia Challenge Dataset; SMDG-19: Standardised Multi-Channel Dataset for Glaucoma; RVO: Retinal Vein Occlusion.

Introduction

Myopia and glaucoma are major causes of visual impairment and Asian public health issues [1,2]. It is anticipated that the myopic population can reach 4.8 billion by 2050 and the number of people with glaucoma will increase to 111.8 million by 2040, disproportionately affecting people residing in East Asia, like Thailand [1,3]. Despite the expected increase in myopia and glaucoma prevalence, mass screening intiatives for the two vision-threatening conditions are still lacking. This is due to the fact that retinal specialists are continuously presented with other significant clinical obligations, such as managing various retinal disorders concurrently (e.g., surgeries and injections), whereas other healthcare professionals may lack the necessary competence to effectively identify myopia and glaucoma [1].

Furthermore, typical manually-performed mass screenings are thought to create extra load on ocular care resources in terms of clinical data processing, notably the interpretation of retinal CFPs [1]. If only human readers analyse visual data, the task might be immense [1].

Fortunately, artificial intelligence (AI) has evolved to play a significant role in automating clinical data processing, reducing the onerous burden [4]. Deep learning (DL), a subset of AI, has outperformed board-certified doctors in optical imaging classifications [5], such as diabetic retinopathy screenings [6]. Nonetheless, DL frequently necessitates significant computational and technical resources (e.g., AI professionals) for the successful construction of a model [7]. Individual physicians or institutions may not always have access to such technological and human resources [7], making it challenging to find AI solutions to fulfil unmet clinical demands for mass myopic and glaucoma screening. Code-free deep learning (CFDL) has come into the limelight as a subtype of DL in 2017, with the distinct feature of allowing physicians with no coding skills to develop their own AI algorithms [5]. CFDL platforms, which feature user-friendly interfaces and simple navigation tools, address the pain point of feature engineering, model architecture selection, and hyperparemeter optimisation in DL model building via an end-to-end automation process [5,8].

Although previous research has investigated the performance of CFDL in the classification of colour fundus photographs into different retinal disease classes [7,9], none has investigated the capacity of CFDL in discriminating PM and glaucoma on CFPs, where such a CFDL-based classifier could benefit community-based screenings with limited medical resources available. The goal of this paper is to develop and validate a CFDL tool that can accurately screen for PM and glaucoma on colour fundus photos (CFPs).

Methods

We sought to follow the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis Standards when conducting the study [10]. Standard AI reporting criteria were still being established around the time of writing, and none were readily accessible [11]. We complied with the recently released Consolidated Standards of Reporting Trials-AI extension (developed for clinical trial reporting involving AI) to properly indicate the key terms and findings [12].

Ocular Disease Intelligent Recgonition (ODIR-K) Dataset

The ODIR-K dataset [13] contained 10000 CFPs of 5000 patients acquired from clinical databases of numerous hospitals/medical centres in China utilising various cameras on the market, including Canon, Zeiss, and Kowa. Under the oversight of quality control management, trained human readers labelled the images. The CFPs were labelled into normal, DR, glaucoma, cataract, AMD, hypertension, myopia, and other diseases/abnormalities classes.

Dataset Preparation and Model Training on Google Cloud Vertex AI AutoML

We removed those CFPs that were labeled with more than one label in the ODIR-K dataset as we aimed to design a single-label multi-class classification model that assigns one label/ diagnosis to each image. Such a process ensures a confounder-free dataset. The singly-labelled normal, PM, and glaucoma CFPs were then reviewed for quality. Re-labelling or image adjustments (e.g., brightness or orientation) were not performed. To avoid selection bias, we also did not perform a manual selection of the identified CFPs [14]. Our training dataset contained 3374 labelled CFPs, with 2834 being normal, 231 being PM, and 309 being glaucoma images. Examples of the training CFPs for PM and glaucoma are displayed in (Figures 1&2), respectively.

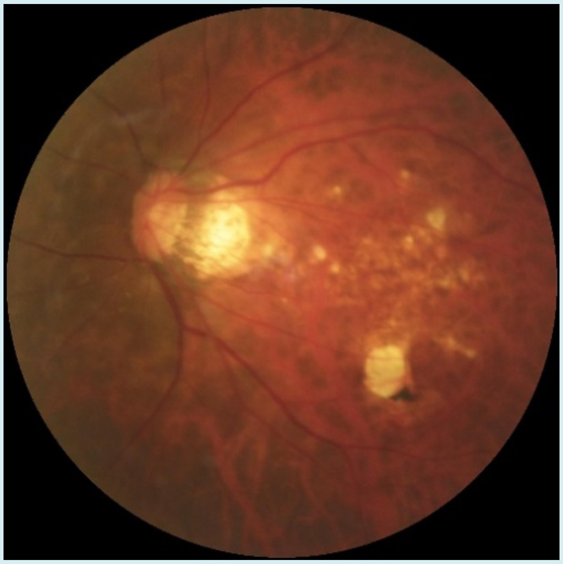

Figure 1: Example of training CFP for pathological myopia.

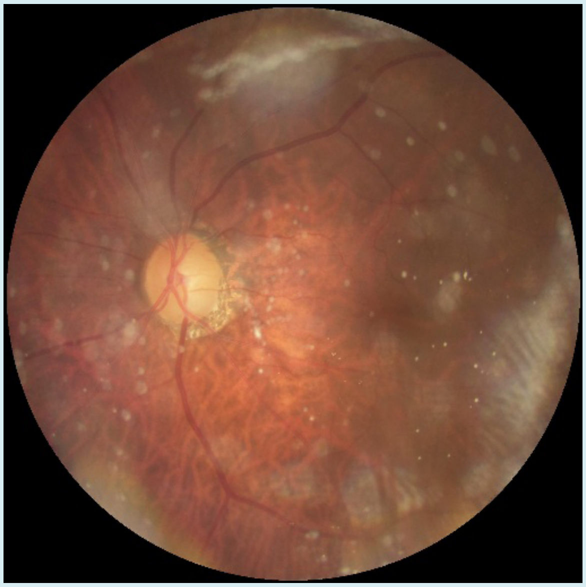

Figure 2: Example of training CFP for glaucoma.

The CFPs, as well as the CSV file with the file paths, were subsequently submitted to the Google Console Vertex AI platform for model training. For our model training, we chose a 'single-label multi-class classification job' and followed the platform's instruction to divide our CFPs into three sets: 80% for training, 10% for validation, and 10% for testing. With previous evidence suggesting reasonable repeatability for AutoML training, we performed each experiment once [15]. Early stopping was automatically enabled to minimise the computational cost. Eventually, an exportable, multi-class classification model was built automatically.

The platform also generated various performance indicators for the model automatically, such as the area under receiver operator curve (AUROC), precision, recall, and F1-score [16,17]. Confusion metrices were also created automatically, reflecting the number of times (and percentages) each label was predicted for each label in the testing set. However, variability measures such as confidence intervals (CIs) were unavailable.

External Validation

The Google Dataset Search engine and the Kaggle Dataset Search engine were used to find publicly available datasets for external validation of the model. The Pathologic Myopia Challenge dataset (PALM) and the Standardised Multi-Channel Dataset for Glaucoma (SMDG-19), both of which contained labelled CFPs for PM and glaucoma, were considered appropriate for our model context.

CFPs for PM, glaucoma, and normal were selected at random from PALM and SMDG-19 for external validation. Only CFPs with a single label for normal, PM, or glaucoma were included in the study. There were no image modifications or re-labeling. The external validation dataset includes 30 normal, 30 PM, and 30 glaucoma CFPs.

Statistical Analysis

Since the platform was unable to provide such metrics, sensitivity, specificity, and accuracy were derived manually from the overall confusion matrix using relevant formulas. True positive/(true positive+false negative) was the formula for sensitivity, whereas true negative/(true negative+false positive) was the formula for specificity [17]. The per-class accuracy has been calculated as (true positive+true negative)/(true positive+false negative+true negative+true negative+true negative+true negative+true negative+false positive) [17].

State-of-the-Art DL Models Designed for the Multi-Class Classification of Myopia and Glaucoma

A focused literature search was conducted on MEDLINE (through PubMed) on December 30th 2023, using the search terms 'deep learning' AND 'glaucoma' AND 'myopia' to locate published DL models designed for the multi-class classification of PM and glaucoma on CFPs. (Figure 3) depicts the search strategy. However, none were found to have designed a three-class classifier for the classification of PM VS glaucoma VS normal on the same training dataset as ours. Considering it was not our intention to statistically compare point estimates of performance measures, we tolerated discrepancies in datasets and model design (e.g. number of classes) for the research that built myopia and glaucoma multi-class classifiers utilising CFPs. Our objective was to make use of the expert-designed DL models to demonstrate the state-of-the-art performance attained by conventional DL in the multi-class categorization of myopia and glaucoma on CFPs.

Figure 3: Search strategy for targeted literature search.

All reviewed DL models [18-20] were convolutional neural networks (CNN) that classified PM and glaucoma, alongside other retinal conditions like hypertensive retinopathy and retinal vein occlusion (RVO) on CFPs, and had certain classification task performance metrics, such as accuracy, sensitivity, and specificity.

Results

On the CFPs, the CFDL classifier demonstrated highly accurate diagnostic performance. Internally validated AUROC, accuracy, and recall for the model were 0.998, 90.74%, and 90.74%, respectively, at the 0.5 confidence threshold.

When internally tested at the 0.5 confidence threshold, the classifier obtained per-class accuracy of 90.74%, 98.79%, and 91.76% for normal, PM, and glaucoma identification, respectively. The sensitivity ranged from 96.04% for normal and 94.44% for PM detection to 40% for glaucoma detection. When it came to per-class specificity, PM and glaucoma detections were more specific (99.13% and 97.11%, respectively) than normal class detections (62.79%). Table 1 summarises the findings of the internally validated model.

	Area-under Precision-Recall Curve	Area-under Receiver-Operator Curve	Accuracy	F-1 score	Precision	Recall	Sensitivity	Specificity
Overall	0.971	-	90.74%	-	90.74%	90.74%	-	-
Normal	-	-	90.74%	-	93.20%	96%	96.04%	62.79%
Pathological myopia	-	-	98.79%	-	89.50%	94.40%	94.44%	99.13%
Glaucoma	-	-	91.76%	-	58.80%	40%	40%	97.11%

Table 1: Overall and per-class performance of the AutoML multi-class classifier upon internal validation.

The classifier has an overall AUROC of 0.863, accuracy of 77.78%, and recall of 77.78% after external validation at the 0.5 confidence threshold. Normal, PM, and glaucoma detection accuracies are 77.78%, 80.46%, and 95.89%, respectively. Normal class detection has the highest per-class sensitivity (100%), followed by glaucoma (90%) and PM (43.33%). In terms of per-class specificity, the PM and glaucoma classes (100% and 93.48%, respectively) outperformed the normal class (60%). Table 2 summarises the model's performance outcomes after external validation.

	Area-under precision-recall curve	Area-under receiver-operator curve	Accuracy	F-1 score	Precision	Recall	Sensitivity	Specificity
Overall	0.727	0.863	77.78%	0.778	77.78%	77.78%	-	-
Normal	-	-	77.78%	0.75	60%	100%	100%	60%
Pathological myopia	-	-	80.46%	0.947	100%	90%	43.33%	100%
Glaucoma	-	-	95.89%	0.605	100%	43.30%	90%	100%

Table 2: Overall and per-class performance of the AutoML multi-class classifier upon external validation.

Discussion

A CFDL-based screening technique was created in this work to detect PM and glaucoma using just CFPs. For model training, validation, and internal testing, a publicly accessible dataset including CFPs acquired using various fundus cameras in various geographical regions in China was used. Our results reveal that our CFDL performed well in PM and glaucoma screening, with an AUROC of 0.998, precision of 90.74%, and recall of 90.74% in the internal test set. The CFDL algorithm performed reasonably in diagnosing myopia and glaucoma in the external test set, with an AUROC of 0.863, accuracy of 77.8%, and recall of 77.8%. Our study went on to compare the model's potential to detect myopia and glaucoma with the state-of-the-art performance attained by traditional DL models in detecting numerous retinal conditions including myopia and glaucoma, on CFPs. This is the first study to our knowledge that examines CFDL's ability to accurately recognise PM and glaucoma on CFPs. We utilised CFPs because they are simpler to operate and more cost-effective as compared to imaging modalities like optical coherence tomography, which made them suited for mass screening throughout the community [21]. The CFDL model we developed may aid in resource-constrained mass screenings for PM and glaucoma at the community level or in isolated geographical areas with restricted access to healthcare.

Previous research used conventional DL to screen for myopia and glaucoma, as well as other prevalent retinal diseases on CFPs. Table 3 summarises the findings of the traditional DL models. Guo C, et al. [18] developed a DL-based five-class classifier to identify PM and glaucoma on CFPs, as well as retinitis pigmentosa and maculopathy. The model had an average accuracy of 0.962 in detecting glaucoma and myopia. Zhu S, et al. [19], on the other hand, built a six-class classifier to predict high myopia and glaucoma in a cohort of CFPs bearing diverse retinal conditions (e.g., macular degeneration and retinal vein occlusion). The model achieved an overall accuracy of 95.59%. Han Y, et al. [20] constructed a generative adversarial network-based-DL model to distinguish glaucoma and myopia from other retinal conditions such as diabetic retinopathy and age-related macular degeneration on CFPs. The model correctly classified glaucoma and myopia with 83.89% and 88.78% accuracy, respectively. We attempted to locate bespoke DL models trained on identical datasets as ours (ODIR-K) for a three-class classification of glaucoma, myopia, and normal fundi on CFPs, but were unsuccessful. As a result, we used existing DL models developed for the multi-class classification of myopia and glaucoma. Despite differences in datasets, model size, and architecture, our CFDL model obtained per-class accuracy of 98.79% and 91.76% for PM and glaucoma diagnosis, respectively, which is comparable to the state-of-the-art performance attained by DL classifiers. Hence, CFDL is a potentially useful AI tool for PM and glaucoma diagnosis. Future research should directly compare the discriminative performance of our technique to custom DL models constructed on the same dataset.

	Dataset source	Dataset Size	Method	Task	Area under Receiver Operator Curve	Accuracy	F-1 score	Sensitivity	Specificity
Guo C, et al. [18]	Colour fundus photographs retrieved from the Kaggle public dataset	250 CFPs used for training	Deep learning convolutional neural network (CNN) (MobileNetV2)	5-class classification (healthy, glaucoma, pathological myopia, maculopathy and retinitis pigmemntosa)	-	Average: 0.962 Glaucoma: 0.96 Pathological myopia (PM): 0.944	-	Average: 90.4% Glaucoma: 90% Pathological myopia (PM): 86%	Average: 97.6% Glaucoma: 97.5% Pathological myopia (PM): 96.5%
Zhu S, et al. [19]	CFPs retrieved retrospectively from the Intelligent Ophthalmology Database of the Ophthalmology Hospital of Nanjing Medical University	2400 CFPs used for training	Deep learning CNN (EfficientNet-B4)	6-class classification (healthy, glaucoma, high myopia, diabetic retinopathy, macular degeneration and retinal vein occlusion)	Glaucoma: 0.983 High myopia (HM): 0.979	Overall: 95.59%	Glaucoma: 94.62% HM: 97.37%	Glaucoma: 97.62% HM: 96.10%	Glaucoma: 99.07% HM: 99.60%
Han Y, et al. [20]	CFPs retrieved from 3 public-available and 1 proprietary dataset	64,351 normal CFPsand 26,148 CFPswith lesions	Generative adversarial network-based DL model	6-class classification (glaucoma, myopia, diabetic retinopathy, age-related macular degeneration, catarct and hypertensive retinopathy)	Glaucoma: 0.916 Myopia: 0.961	Glaucoma: 83.89% Myopia: 88.78%	-	Glaucoma: 83.70% Myopia: 88.83%	Glaucoma: 84.0% Myopia: 88.75%

Table 3: Summary of the traditional deep learning multi-class classification models.

In East Asia, glaucoma and PM have now become the leading causes for blindness [22]. A routine screening program to identify asymptomatic patients and refer them for treatment in a timely fashion could reduce disease burden and improve prognosis [23]. Routine screening for glaucoma and PM was shown to be very cost-effective, with AI-assisted screening being the most cost-effective method for regular screenings [23]. While AI promises to underpin broad-scale myopic and glaucoma screenings, the affordability, accessibility, and flexibility of the AI tool should play crucial roles in establishing an economically viable, sustainable, and adaptable screening program in the long term [23].'

Our model demonstrated characteristics that are in line with the goals of broad PM and glaucoma screenings. First, CFDL is resource-saving in nature. Expert resources, like AI engineers and analysts, can be saved because CFDL platforms automate the necessary steps in DL model construction (model architecture selection and hyperparameter optimisation) [24] and allow developers to gain immediate access to the model results via a graphical user interface without the need for analysts to interpret results [25]. Time can also be saved since AutoML eliminates the requirement for developers to go through a trial-and-error process to adjust a DL model that is appropriate for the application setting [26]. The optimal settings for the DL model [26] are determined automatically by AutoML. Furthermore, more complicated and sophisticated algorithms for processing a greater quantity of data and variables may be generated swiftly via Cloud, without the need for heavy hardware or GPU resources [7,24]. This is especially useful for doctors dealing with population-level data in a screening scenario. CFDL's resource-friendliness supports screening programmes with minimal budget and expertise accessibility [26]. Second, physicians may better tailor screening algorithms to their specific application circumstances. It is worth noting that there may be a mismatch in expectations between model builders and medical researchers when physicians rely only on ML specialists with no public health experience to design models, resulting in models that do not meet the needs of the clinicians [24]. Adjusting the confidence threshold to fine-tune the algorithm's sensitivity and specificity for detecting pathologies is a form of customisation [8]. Following model deployment, fine-tuning is also easily accomplished using CFDL platforms. CFDL solutions offer post-deployment monitoring of algorithm performance and allow developers to easily improve their models with new data acquired in deployment scenarios without having to retrain the model [8]. CFDL is more flexible than manually constructed models and can be readily improved in response to changes in the deployment scenario [24].

Although our CFDL model gives high diagnostic accuracy in general and may be useful in screening contexts, performance is not always constant. Sensitivity varies with class, ranging from 96.04% in normal class detection to 40.0% in glaucoma class detection. This is due to the unbalanced dataset, in which the normal class comprised much more CFPs (2834 CFPs) than the glaucoma group (309 CFPs). This problem may be remedied by implementing a thresholding technique in which the algorithm's sensitivity and specificity can be modified by adjusting the confidence threshold [8]. Ophthalmologists can tailor the classification algorithm to their unique needs by selecting the best combination of sensitivity and specificity [27]. An in silico screening system for PM and glaucoma, for example, should be very sensitive [27]. This may be accomplished by lowering the confidence level [8]. However, given the already constrained hospital resources in the community where false positives are not desired [8], fine-tuning the confidence threshold to achieve a fine balance of sensitivity and specificity that suits the deployment environment context can be easily and efficiently executed in a postprocessing step [28].

Our CFDL model, like other DL models, has been shown to have strong internal validation performance but worsened in external testing settings [29-31]. This is because there is a data shift [32]. Given the relatively small size of our training dataset and the homogeneous patient demographics (ethnicity), model training might have been biased and non-generalized [33]. When confronted with a new CFP dataset with out-of-distribution characteristics such as different disease prevalences (myopia and glaucoma) in the population, distinct diagnostic criteria (e.g., glaucoma definition ground truth), and varied image properties (e.g., field of view or colour distribution), the CFDL model may have been insufficiently robust to handle the distribution changes [34,35].

There are several strengths in our study. First, we did not make adjustments (e.g., brightness or pixel change) to the eligible CFPs, as it was previously reported that tampering with images in minor ways could potentially undermine the algorithm's classification [36]. A previous study discovered that human-undetectable changes in the pixels of CFPs caused the model to misclassify half of the DR CFPs as normal [37]. Moreover, we utilised datasets made available to the public and detailed our methodology in study design that allows research reproducibility. Additionally, we only included retinal diseases that are most prevalent and distressing to our community, such as glaucoma and pathological myopia [38]. Unlike other multi-class classifiers that performed glaucoma and PM classification on CFPs, less prevalent disease classes, such as hypertensive retinopathy and RVO, that are irrelevant to our pressing screening needs were excluded in our algorithm's development [38]. This increases the screening utility of the model [38].

Our algorithm, however, has certain drawbacks that necessitate cautious interpretation of our findings. First, because our model only used CFPs from a single publicly available dataset, inherent biases may have existed. Limited data availability, the single-image capture technique, and limited camera model availability are all sources of bias [38]. Compiling a training dataset using CFPs obtained from different publicly available databases would better reflect the varied and dynamic real-world environment, while reducing data bias associated with single-source acquisitions [38]. Furthermore, the model's performance has not been compared to that of physicians. Comparing the model's performance to that of physicians demonstrates the model's clinical relevance and efficacy [38]. Additionally, without being validated on real-world data, it is uncertain if the model's diagnostic performance can be transferred into clinical efficacy. Finally, our study solely looked at the model's diagnostic accuracy and did not look at other application issues. Considerations should include the model's cost-effectiveness, the model's comprehensibility, the algorithm's acceptance, and any safety or liability problems associated with model use. To ensure the model's effective implementation in real-world situations, a multi-dimensional examination of both its performance and application factors should be done.

Conclusion

Using CFDL, we proved the capability of physicians with little coding knowledge to generate accurate DL algorithms for PM and glaucoma screening. Our model achieved state-of-the-art overall performance, but it still requires refinements with improved training data and setting customisation to the deployment circumstance.

Acknowledgement

None

Funding/ Support

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of Interest

All authors have disclosed no conflicts of interest

Data Availability Statement

All data is publicly available and can be retrieved from open-source platforms like Google Dataset Search and Kaggle. The links to the datasets used were also cited in the reference.

Ethics Statements and Patient Consent for Publication

Not required.

Author Contributions

All authors contributed to the (1) concept or design, (2) acquisition of data, (3) analysis or interpretation of data, (4) drafting of the manuscript, and (5) critical revision for important intellectual content. All authors had full access to the data, contributed to the study, approved the final version for publication, and take responsibility for its accuracy and integrity.

References

1. Lu L, Zhou E, Yu W, Chen B, Ren P, et al. (2021) Development of deep learning-based detecting systems for pathologic myopia using retinal fundus images. Commun Biol 4(1): 1225.

2. Lim WS, Ho HY, Ho HC, Chen YW, Lee CK, et al. (2022) Use of multimodal dataset in AI for detecting glaucoma based on fundus photographs assessed with OCT: focus group study on high prevalence of myopia. BMC Med Imaging 22(1): 206.

3. Tham YC, Li X, Wong TY, Quigley HA, Aung T, et al. (2014) Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. Ophthalmology 121(11): 2081-2090.

4. Hamet P, Tremblay J (2017) Artificial intelligence in medicine. Metabolism 69S: S36-S40.

5. O'Byrne C, Abbas A, Korot E, Keane PA (2021) Automated deep learning in ophthalmology: AI that can build AI. Curr Opin Ophthalmol 32(5): 406-412.

6. Nakayama LF, Ribeiro LZ, Novaes F, Miyawaki IA, Miyawaki AE, et al. (2020) Artificial intelligence for diabetic retinopathy screening: a review. Eye 34(3): 451-460.

7. Korot E, Guan Z, Ferraz D, Wagner SK, Zhang G, et al. (2021) Code-free deep learning for multi-modality medical image classification. Nature Machine Intelligence 3(4): 288-298.

8. Jacoba CMP, Doan D, Salongcay RP, Aquino LAC, Silva JPY, et al. (2023) Performance of Automated Machine Learning for Diabetic Retinopathy Image Classification from Multi-field Handheld Retinal Images. Ophthalmol Retina 7(8): 703-712.

9. Antaki F, Coussa RG, Kahwati G, Hammamji K, Sebag M, et al. (2023) Accuracy of automated machine learning in classifying retinal pathologies from ultra-widefield pseudocolour fundus images. Br J Ophthalmol 107(1): 90-95.

10. Collins GS, Reitsma JB, Altman DG, Moons KG (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 350: g7594.

11. Collins GS, Moons KGM (2019) Reporting of artificial intelligence prediction models. Lancet 393(10181): 1577-1579.

12. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, et al. (2020) Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med 26(9): 1364-1374.

13. Larxel (2020) Ocular Disease Recognition.

14. Yu AC, Eng J (2020) One Algorithm May Not Fit All: How Selection Bias Affects Machine Learning Performance. Radiographics 40(7): 1932-1937.

15. D'Amour A, Heller K, Moldovan D, Adlam B, Alipanahi B, et al. (2020) Underspecification Presents Challenges for Credibility in Modern Machine Learning. arXiv.

16. Davis J, Goadrich M (2006) The relationship between Precision-Recall and ROC curves. In: Proceedings of the 23rd international conference on Machine learning (ICML '06). Association for Computing Machinery. New

York, USA, pp: 233-240.

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

18. Guo C, Yu M, Li J (2021) Prediction of Different Eye Diseases Based on Fundus Photography via Deep Transfer Learning. J Clin Med Res 10(23): 5481.

19. Zhu S, Lu B, Wang C, Wu M, Zheng B, et al. (2022) Screening of Common Retinal Diseases Using Six-Category Models Based on EfficientNet. Front Med 9: 808402.

20. Han Y, Li W, Liu M, Wu Z, Zhang F, et al. (2021) Application of an Anomaly Detection Model to Screen for Ocular Diseases Using Color Retinal Fundus Images: Design and Evaluation Study. J Med Internet Res 23(7): e27822.

21. Hagiwara Y, Koh JEW, Tan JH, Bhandary SV, Laude A, et al. (2018) Computer-aided diagnosis of glaucoma using fundus images: A review. Comput Methods Programs Biomed 165: 1-12.

22. Ahn J, Kim G, Choi M (2023) A Bibliometric Analysis of Myopia Research in East Asia in the 21st Century: The Socio-Economic Status and Quantitative Analysis. Inquiry 60: 469580231174333.

23. Liu H, Li R, Zhang Y, Zhang K, Yusufu M, et al. (2023) Economic evaluation of combined population-based screening for multiple blindness-causing eye diseases in China: a cost-effectiveness analysis. Lancet Glob Health 11(3): e456-e465.

24. Zhang T, Rabhi F, Behnaz A, Chen X, Paik HY, et al. (2022) Use of automated machine learning for an outbreak risk prediction tool. Informatics in Medicine Unlocked 34: 101121.

25. Papoutsoglou G, Karaglani M, Lagani V, Thomson N, Røe OD, et al. (2021) Automated machine learning optimizes and accelerates predictive modeling from COVID-19 high throughput datasets. Sci Rep 11(1): 15107.

26. Callender T, Schaar MDV (2023) Automated machine learning as a partner in predictive modelling. Lancet Digit Health 5(5): e254-e256.

27. Banerjee P, Dehnbostel FO, Preissner R (2018) Prediction Is a Balancing Act: Importance of Sampling Methods to Balance Sensitivity and Specificity of Predictive Models Based on Imbalanced Chemical Data Sets. Front Chem 6: 362.

28. Esposito C, Landrum GA, Schneider N, Stiefl N, Riniker S (2021) GHOST: Adjusting the Decision Threshold to Handle Imbalanced Data in Machine Learning. J Chem Inf Model 61(6): 2623-2640.

29. Li Z, He Y, Keel S, Meng W, Chang RT, et al. (2018) Efficacy of a Deep Learning System for Detecting Glaucomatous Optic Neuropathy Based on Color Fundus Photographs. Ophthalmology 125(8): 1199-1206.

30. Li L, Xu M, Wang X, Jiang L, Liu H (2019) Attention Based Glaucoma Detection: A Large-scale Database and CNN Model. arXiv 1: 1-10.

31. Orlando JI, Fu H, Breda JB, Keer KV, Bathula DR, et al. (2020) REFUGE Challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. Med Image Anal 59: 101570.

32. Gao Y (2024) Radiology: Artificial Intelligence.

33. Ho SY, Phua K, Wong L, Goh WWB (2020) Extensions of the External Validation for Checking Learned Model Interpretability and Generalizability. Patterns (N Y) 1(8): 100129.

34. Beede E, Baylor E, Hersch F, Iurchenko A, Wilcox L, et al. (2020) A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy (CHI '20). In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery. New York, USA, pp: 1-12.

35. Hemelings R, Elen B, Schuster AK, Blaschko MB, Barbosa-Breda J, et al. (2023) A generalizable deep learning regression model for automated glaucoma screening from fundus images. NPJ Digit Med 16(1): 112.

36. Mayro EL, Wang M, Elze T, Pasquale LR (2020) The impact of artificial intelligence in the diagnosis and management of glaucoma. Eye 34(1): 1-11.

37. Lynch SK, Shah A, Folk JC, Wu X, Abramoff MD (2017) Catastrophic Failure in Image-Based Convolutional Neural Network Algorithms for Detecting Diabetic Retinopathy. Invest Ophthalmol Vis Sci 58(8): 3776-3776.

38. Chłopowiec AR, Karanowski K, Skrzypczak T, Grzesiuk M, Chłopowiec AB, et al. (2023) Counteracting Data Bias and Class Imbalance-Towards a Useful and Reliable Retinal Disease Recognition System. Diagnostics (Basel) 13(11): 1904.