Privacy-Preserving AI-based Glaucoma Referral using Multi-Centric Real-World Data: A Feasibility Study with Federated Learning

Filipe Soares[1 [0000-0002-2881-313X]] (✉), Telmo Barbosa[1 [0000-0001-9211-8900]], João Barbosa-Breda[2 [0000-0001-7816-816X]], Luís Abegão Pinto[3,4 [0000-0002-9960-7579]], Carlos Marques-Neves[4,5 [0000-0002-3842-2466]]

[1] Fraunhofer Portugal AICOS, Rua Alfredo Allen 455/461, 4200-135 Porto, Portugal
filipe.soares@aicos.fraunhofer.pt
[2] Department of Ophthalmology, Unidade Local de Saúde São João, Portugal
[3] Department of Ophthalmology, Unidade Local de Saúde Santa Maria, Portugal
[4] Center for the Study of Vision Sciences, University Ophthalmology Clinic, Faculty of Medicine, University of Lisbon, Portugal
[5] alm-PRIMUM, Lisbon, Portugal

Abstract. Glaucoma is a leading cause of irreversible blindness worldwide, often progressing undiagnosed due to asymptomatic early stages and limited access to specialist care. To address these barriers, we present a feasibility study of Glaucoma-PAIR (Privacy-preserving AI-based Referral), a computer-aided system developed using federated learning (FL). The study was conducted across a multi-centric network of three heterogeneous clinical sites in Portugal, two tertiary large public hospitals and one private clinic, each with distinct patient demographics, imaging equipment, and data distributions. The system leverages color fundus photography and expert-labeled cases to train a glaucoma classification model, without transferring sensitive patient data across institutions by employing FL, ensuring compliance with institutional governance and data protection regulations. Our work addresses major challenges in clinical AI, including privacy, generalizability, and integration into real-world workflows. Through close collaboration with ophthalmologists, we identified key constraints in existing referral pathways and incorporated those insights into the study design. Notably, the federated global model achieved performance comparable to a centralized model trained on pooled data, improved the average sensitivity which is a critical metric for a screening tool, and showed significant performance gains at the most clinically diverse site. This study provides a practical demonstration of responsible machine learning, combining privacy-preserving operations with clinical feasibility. Our findings highlight the potential of federated learning to enable the development of scalable and equitable AI tools, to support patient triage for a glaucoma specialist, particularly in settings with limited ophthalmology accessibility, promoting access to

To appear in the Proceedings of the First Workshop on Responsible Healthcare using Machine Learning (RHCML 2025)

earlier diagnosis and care. We discuss the implications for future deployment and integration into national screening workflows.

Keywords: Federated Learning · Computer Vision · Ophthalmology.

1 Introduction

Glaucoma is a leading cause of irreversible vision loss, with millions of cases remaining undiagnosed globally. Studies estimate that around 50% of glaucoma cases are undiagnosed in developed countries, with even higher rates in developing regions [16]. This lack of diagnosis results in patients with advanced stages of the disease, leading to higher treatment costs and a greater burden on healthcare systems [2]. Current screening methods for glaucoma are not cost-effective, resulting in inefficient use of healthcare resources. The cost of examination and specialized equipment limits widespread screening, particularly in resource-limited settings [35,38]. One of the clinical sites integrating this work, has been pioneering a centre in a prospective clinical trial with artificial intelligence on glaucoma screening (Clinicaltrial.gov:NCT05875090), which suggested a sustained circuit between primary care settings and hospital-based clinics [20,21]. AI has shown promise in improving the efficiency of ophthalmology services [34], mirroring its impact in radiology [7]. However, the scientific community, public entities, and corporate innovators face significant challenges in transferring computer-aided diagnosis (CAD) systems to the real world due to: (a) lack of diversity in training data – limiting their applicability across different populations [18]; (b) struggles with generalizability – may perform well in controlled environments but underperforms in real-world settings [38]; (c) data privacy concerns and security concerns [14]; (d) interoperability issues – integrating AI tools with existing healthcare systems is challenging; [15]; (e) sound integration into workflows – effective integration requires a mixed methods approach, combining ethnographic research, service experience, process engineering, and participatory design for actionable insights [10]. These barriers hinder the feasible implementation of AI solutions in clinical settings, delaying the benefits of advanced technologies. To overcome these, recent approaches comprise:

Federated Learning (FL) addresses data diversity and generalizability by enabling AI training across multiple institutions without sharing raw data [33,38], and other privacy-preserving layers can be added as the homomorphic encryption during weight communication [11]. In addition, collaborative approaches involving multidisciplinary teams and implementing feedback loops between AI systems and medical experts enhance the system's reliability and continuous improvement [7].

1.1 Main Contributions

To address these challenges, we have carried out an observational study with retrospective data in three clinical sites in Portugal, to test FL for training a computer-aided system called Glaucoma-PAIR (Privacy-preserving AI-based Referral) starting from a pre-trained model by [17]. The multi-centric design of our study, based on real-world data, contributes to generate evidence on how to develop ML reliability close to the clinical environment. The main contributions of this study are summarised below.

Privacy-Preserving Federated Learning for Glaucoma Diagnosis: We developed a computer-aided referral system trained using FL across three independent clinical sites. This approach enables the use of sensitive real-world data without requiring cross-institutional data sharing, preserving patient privacy and institutional autonomy.

Multi-Centric Real-World Validation: Unlike many AI models trained on limited or homogeneous datasets, our study includes private data from diverse clinical environments enhancing model generalizability across. The system maintains a strong performance even in lower-resource sites, showcasing its robustness.

Clinician-Guided Labeling and Development: The AI model is trained on cases curated and labeled by glaucoma specialists, ensuring meaningful grading as the ground truth for machine learning development. This enhances both the clinical relevance and the heterogeneity in data required for the AI model robustness aiming to support the detection of glaucoma suspects, in the scope of patient triage and referral for a glaucoma specialist.

Responsible AI in Practice: This work demonstrates how responsible AI principles—privacy, robustness, and stakeholder collaboration—can be operationalised in real healthcare facilities. It offers a model for ethically grounded ML development in sensitive domains such as ophthalmology.

2 Related Work

The need for early glaucoma detection, coupled with the limited number of ophthalmologists, has led to the proposal of various automatic methods for glaucoma diagnosis or referral, using color fundus photographs, also called retinal fundus images [40]. In recent years, several Deep Learning (DL) architectures have been employed for the Glaucoma classification purpose, including Inception-V3 [19], MobileNetV2 [22], ResNet-50 [13,9], EfficientNetB0 [17], VGG-16 [6], DenseNet-201 [29], and vision transformers [8,37]. To improve the outcomes in terms of the glaucoma diagnosis, techniques like attention-based CNNs [31], transfer learning [26], active learning [13], hierarchical DL models [39], and even generative approaches [17] have been explored in the last years.

The work of Martins et al. [22] was a key contribution in segmentation around the optic nerve, often included in the detection typically based on U-shaped architectures (e.g. U-Net) [24,26]. Although, it should be noted that relevant information for diagnosing glaucoma extends beyond the optic disc [12].

CAD systems still largely depend on the use of publicly available datasets such as ORIGA [42], REFUGE [28] or AIROGS [36]. However, most of them are skewed towards specific demographic groups, failing to capture variations in the population. Accessing datasets with a wide range of variables such as age, ethnicity, and genetics, along with diverse clinical presentations of glaucoma, is a significant challenge [41] as there are changes in retinal structures among different racial and ethnic groups [5,30].

3 Methods

3.1 Clinical Context and Data Sources

The study was conducted in collaboration with three ophthalmology centers in Portugal: two large tertiary public hospitals, ULS São João (ULSSJ) located in Porto, ULS Santa Maria (ULSSM) and one private clinic, alm-PRIMUM (ALM), both in Lisbon. These sites differ in patient demographics, clinical workflows, glaucoma prevalence, and imaging equipment. At ULS São João, fundus images were acquired during ophthalmology consultations, upon request and not as a routine, resulting in a dataset with a high proportion of advanced glaucoma cases and patients with high myopia—reflecting a population with more severe disease manifestations and discs that are hard to judge on fundoscopy. At ULS Santa Maria, the cases stemmed from a virtual clinic for which patients are referred to when they are deemed glaucomatous or at least with high risk factors (such as high IOP). This results in greater clinical heterogeneity and a more balanced distribution of glaucoma stages in the dataset, due to a relatively high frequency of wrong referrals (ie, normal cases) often based on single factors. At ALM, images are primarily sourced from general ophthalmology and neuro-ophthalmology consultations (approximately 34% from the latter), leading to a case mix characterized by higher complexity and a greater incidence of diagnostically challenging or atypical presentations. Table 1 summarizes the key characteristics of the datasets from each site, underscoring the heterogeneity of clinical contexts used to assess the robustness and generalizability of the proposed AI-based glaucoma referral system.

Institutional ethical approvals were obtained from the three local ethical committees in health, and a data governance protocol was secured in accordance with national health data protection regulations. The study followed a retrospective design, utilizing color fundus photographs already acquired during clinical practice. All data remained within the IT infrastructure of each institution; no raw image data was exchanged, in line with privacy-preserving principles.

Each clinical coordinator oversaw the preparation of the dataset at their respective institution, following a standardized procedure: selection of eligible images, anonymization, and compilation of demographic and clinical data. To better characterise the samples, the recorded variables included: anonymized patient ID, anonymized retinal image ID, eye laterality (right/left), optic disc-centered field of view (yes/no), age, sex, and intraocular pressure (IOP).

Each institution's dataset was assigned to a panel of three glaucoma specialists from that institution for image annotation and case labeling. Using a user interface customised for this study, the nine experts independently reviewed each case and classified it into one of three categories: non-classifiable (NC), glaucoma suspect, or normal, based solely on digital visual inspection of the fundus images. Each expert labeled an average of 712 images, including the NC. The final ground-truth label for each case was established through majority voting among the three local experts.

Table 1. Dataset Characteristics by Clinical Site

Characteristic	**ULSSJ**	**ULSSM**	**ALM**
Total number of images	404	1252	480
Images after excluding NC	322	1110	467
Age average (years)	65	69	65
Female (%)	51%	57%	42%
IOP average (mmHg)	15.94	15.81	15.75
Fundus camera brand	Zeiss	Canon	Canon
Image quality	Usable	Good	Good
Optic disc centered (%)	62	100	0
Expert experience (years)	7; 9; 9	12; 4; 3	25; 6; 30
Glaucoma suspect label (%)	69	66	31

3.2 Federated Learning based on a pre-trained model

A FL architecture was developed and implemented using the Flower [3] software framework, enabling decentralized model training across participating hospitals without data centralization. Each site locally trained the model on its own dataset, and only model weight updates were transmitted to a central coordinating server. The server aggregated these updates using a federated averaging (FedAvg) algorithm and redistributed the global model to all participants.

The baseline model for image classification was based on a previous work described in [17], which is the pre-trained model used in this study to be finedtuned in federated learning. The backbone model selected was EfficientNetB0, chosen for its balance between performance and computational efficiency. Training was performed on a harmonized dataset, compiled by combining multiple publicly available glaucoma-labeled datasets (ORIGA, Drishti-GS, REFUGE, RIM-ONE, ACRIMA). Since these datasets varied in image acquisition, some containing a standard field-of-view (FOV) in retinal imaging and others focused on the optic disc area, the images were standardized by cropping around the optic disc. These crops were resized to 224 $\times$ 224 pixels and normalized to ensure consistency across samples.

For the federated learning experiments in this study, the optic disc was detected automatically and then cropped following the work in [22], without applying any image processing or histogram equalization. The same squared input sized was used for training with a batch size of 4. The model optimization employed a multi-label Dice loss function with the ADAM optimizer, initialized with a learning rate of 0.00001. The training process continued for up to 20 local epochs in each federated round, for a total of 10 rounds (e.g. 200 epochs in total). Early stopping was triggered after 15 epochs without validation loss improvement to mitigate local overfitting.

We benchmarked the federated model against the baseline model at each clinical site. Performance was evaluated using site-wise test sets with metrics including accuracy, F1-score, precision, sensitivity, and specificity published in a MLFLow service [4]. It was only possible to apply a dataset split of 45% Train, 15% Validation and 40% Test sets, following the study protocol, in the site with less samples. Due to memory limitations in the local computer, delays in local training had caused some synchronicity issues in federated learning orchestration process using Flower. For this reason, to approximate the training conditions of all local sites, we used a training set equivalent in absolute size, keeping the 15% in validation and using the remaining data as a local test set.

3.3 System Deployment

To enable a privacy-preserving, decentralized workflow for AI development and validation, a custom FL infrastructure was deployed at each clinical site. The system architecture was designed to support on-premise data processing and model training while ensuring compliance with institutional governance protocols and European data protection regulations (GDPR).

Each participating site was equipped with a microcomputer edge box with GPU (NVIDIA Jetson with 32GB RAM) configured to run containerized services via Docker. These services included the local machine learning training pipeline, data management tool, image annotation interface, inference engine and the respective Application Programming Interfaces. Deployment was performed collaboratively with the hospitals' IT teams, who were responsible for configuring secure network access. This included enabling SSL encryption, setting up HTTPS port forwarding, and defining firewall rules to allow secure communication with the central FL coordination server (a cloud virtual machine running in West Europe on Microsoft Azure) while preventing any outbound transfer of raw patient data.

A custom web-based image annotation tool (see Figure 1) was installed onsite and made accessible through user-specific login credentials. This interface allowed glaucoma specialists to securely review, label, and annotate fundus images within the local hospital network. Although not fully integrated with local clinical workflows to speed up the study execution, the image annotator usability enabled efficient interaction of the users with the dataset while maintaining data confidentiality.

The entire system operated under a FL paradigm, where each site trained local models on their respective datasets and shared only model updates (weights) with the central server. No images or identifiable patient data were transmitted externally. This decentralized setup ensured that sensitive health data never left the premises, aligning with institutional ethics approvals and national data handling standards.

This deployment strategy not only preserved data privacy but also created a realistic operational environment for evaluating the performance and feasibility of AI-assisted glaucoma referral tools near clinical practice.

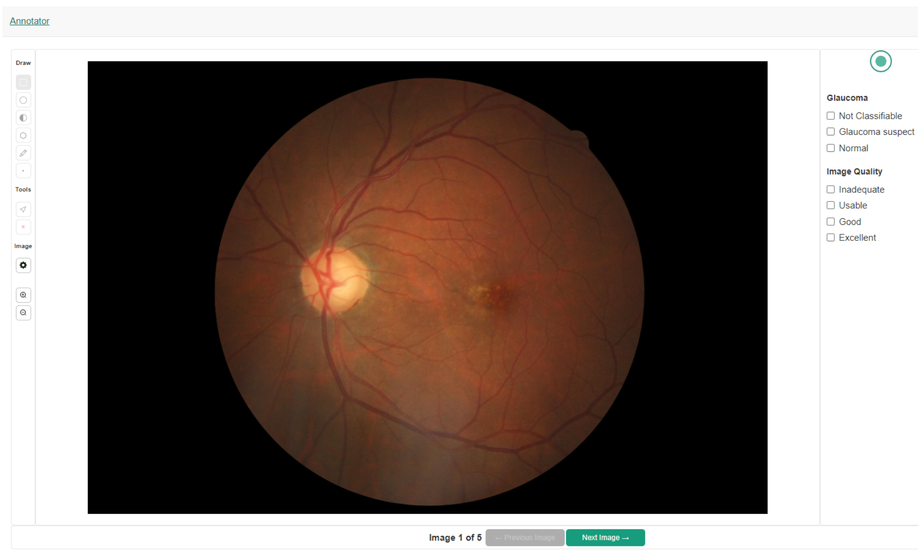

Fig. 1. Image annotation tool.

4 Results

We compared the performance of the centralized baseline model with the FL model across the three participating clinical sites, ULSSJ, ULSSM, and ALM. Table 2 presents the results per site and averaged across all sites. Table 3 depicts the confusion matrices of the tests sets for baseline and FL models.

The average performance across the three centers indicates that the FL model achieved comparable accuracy (0.72) to the baseline (0.73), with a slight improvement in F1-score from 0.71 to 0.72, reflecting more balanced performance across classes. Notably, the federated model outperformed the baseline in sensitivity, increasing from 0.65 to 0.72 on average, which is a desirable outcome in a screening context where minimizing false negatives is critical. However, this gain in sensitivity came at the cost of reduced specificity, which dropped from 0.79 in the baseline model to 0.68 in the FL model.

Table 2. Performance Comparison of Baseline and Federated Models Across Clinical Sites

Metrics	**ULSSJ**		**ULSSM**		**ALM**		**Average**	
	Baseline	FL	Baseline	FL	Baseline	FL	Baseline	FL
Accuracy	0.74	0.71	0.66	0.70	0.79	0.75	0.73	0.72
F1-score	0.80	0.79	0.68	0.75	0.64	0.62	0.71	0.72
Precision	0.83	0.79	0.90	0.84	0.66	0.57	0.80	0.73
Sensitivity	0.78	0.80	0.55	0.68	0.63	0.67	0.65	0.72
Specificity	0.63	0.51	0.88	0.74	0.86	0.78	0.79	0.68

Table 3. Confusion Matrices of Baseline and Federated Models Across Clinical Sites

Metrics	**ULSSJ**		**ULSSM**		**ALM**		**Average**	
	Baseline	FL	Baseline	FL	Baseline	FL	Baseline	FL
False Negative	20	18	233	164	27	24	93	69
False Positive	15	20	31	68	23	36	23	41
True Negative	26	21	234	197	141	128	134	115
True Positive	71	73	285	354	45	48	134	158

At ULSSJ, the FL model slightly decreased in accuracy (0.74 to 0.71) and precision (0.83 to 0.79), while showing a modest improvement in sensitivity (0.78 to 0.80). Performance at ULSSM exhibited the most pronounced gains from FL, with accuracy improving from 0.66 to 0.70 and F1-score rising from 0.68 to 0.75. Here, sensitivity increased substantially from 0.55 to 0.68, with a trade-off in specificity (0.88 to 0.74). At ALM, both models performed strongly in terms of accuracy and specificity, although the FL model showed a slight performance decline across most metrics compared to the baseline.

Overall, these results suggest that federated learning can help improve generalization in settings with heterogeneous data distributions by boosting sensitivity and F1-score—key metrics in medical screening, while slightly compromising specificity. The federated approach may be particularly advantageous in underrepresented or more diverse clinical environments, as observed in the ULSSM results.

5 Generalizable Insights about Responsible Application of Machine Learning in Healthcare

This study explored the feasibility and performance trade-offs of applying FL for glaucoma detection across multiple clinical sites with heterogeneous data sources and workflows. While the baseline model, trained centrally on pooled data, achieved a good overall predictive performance, the federated model demonstrated competitive results for the context of a AI-powered glaucoma screening, particularly in terms of F1-score and sensitivity. Moreover, these results were obtained having a considerably small and private training set, and despite overcoming some of the challenges of decentralized training in real-word clinical settings. These findings reflect common trade-offs in some FL applications, where performance may be modestly compromised due to local sample size limitations, and domain shifts across institutions.

Importantly, the federated approach preserved institutional data privacy by enabling local model training without transferring raw images, thus aligning with GDPR requirements and the data governance policies of each hospital. This privacy-preserving setup is especially relevant for scaling AI adoption in healthcare, where regulatory and ethical constraints often impede centralized data sharing. This setup can be relevant to obtain more, real and diverse data, needed to develop and validate high-risk AI systems that include generalizable machine learning models.

The observed performance gaps between the baseline and FL models suggest that further advances are needed, namely with the expansion of the available data in the federated network. Future work should explore personalized FL strategies, domain adaptation techniques, and more robust model aggregation methods as the ones experimented in [1], to better address inter-site variability and improve convergence.

This study also highlights the value of interdisciplinary collaboration in AI for healthcare. Clinical input guided image selection, annotation, and contextual interpretation, while machine learning expertise shaped model design and evaluation. The diverse clinical settings, ranging from public hospitals to private clinics, with varying patient profiles and imaging acquisition setups, offered a realistic testbed for evaluating the generalizability of AI models across different healthcare contexts.

Looking ahead, future research should prioritize the integration of fairness, transparency, and interpretability into privacy-preserving FL networks. Clinician trust and patient safety hinge not only on model performance but also on the ability to identify failure cases, and adapt to local clinical practices. To this end, incorporating continuous stakeholder feedback, and aligning development processes with real-world clinical needs will be critical for the responsible and effective deployment of AI-assisted glaucoma screening tools.

The practical application of this work is emphasized by new workflows, such as the virtual hospital-based clinics [32,23,25,27]. For instance, a "Virtual Clinic for Glaucoma" is being initiated at ULSSJ, representing a new asynchronous and remote digital consultation model designed to optimize healthcare resources for both glaucoma specialists and patients. Given that glaucoma is a chronic disease requiring constant monitoring, and with patient numbers expected to rise, current clinical schemes are becoming unsustainable. In this virtual clinic model, a technician performs a set of exams, and the physician evaluates the results asynchronously, increasing the number of patients that can be monitored. This process aids in the triage of patients referred for initial glaucoma consultations and in evaluating low-risk patients, freeing up specialists' time for more complex and urgent cases. Such a workflow aims to improve the patient experience, accelerate diagnoses, and increase clinical capacity by leveraging technology for data collection and analysis. The responsible integration of an AI-based support solution for the robust classification of fundus images, as showed in this study, is intended to directly assist physicians in their decision-making and optimize the quality of patient referrals within this new paradigm.

6 Conclusion

This study demonstrates the feasibility of deploying a federated learning framework for glaucoma detection using color fundus photography across multiple ophthalmology centers, preserving patient data privacy while enabling collaborative model development. Our results show that although the centrally trained baseline model achieved slightly higher predictive performance considering the set of metrics evaluated, the FL model offered a viable alternative with competitive sensitivity and F1-score, particularly valuable in the scope of screening and AI-based patient referrals (e.g. from general ophthalmology to glaucoma expert units), and in contexts where training data cannot be shared due to regulatory constraints.

By leveraging a diverse, multi-centric dataset and integrating clinical expertise throughout the process—from labeling to deployment—this work illustrates a realistic path toward scalable, privacy-conscious AI in ophthalmology. The study also highlights current limitations in federated learning performance due to data heterogeneity and inter-institutional variability, pointing to future research directions in personalization, domain adaptation, and model explainability.

Ultimately, our findings support the growing potential of federated learning to democratize access to high-performing AI models in healthcare, while respecting ethical and legal boundaries critical for clinical adoption.

Acknowledgments. This work was supported by European funds through the Recovery and Resilience Plan, under the project "Center for Responsible AI" (C645008882-00000055), and by the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101189689.

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Baptista, T., Soares, C., Oliveira, T., Soares, F.: Federated Learning for Computer-Aided Diagnosis of Glaucoma Using Retinal Fundus Images. Applied Sciences **13**(21), 11620 (Jan 2023). `https://doi.org/10.3390/app132111620`, `https://www.mdpi.com/2076-3417/13/21/11620`, number: 21 Publisher: Multidisciplinary Digital Publishing Institute
2. Barbosa-Breda, J., Gonçalves-Pinho, M., Santos, J.V., Rocha-Sousa, A., Abegão-Pinto, L., Stalmans, I., Freitas, A.: Trends in Glaucoma Surgical Procedures in Portugal: A 16-Year Nationwide Study (2000-2015). Journal of Glaucoma **27**(8), 682–686 (Aug 2018). `https://doi.org/10.1097/IJG.0000000000001011`

3. Beutel, D.J., Topal, T., Mathur, A., Qiu, X., Fernandez-Marques, J., Gao, Y., Sani, L., Li, K.H., Parcollet, T., de Gusmão, P.P.B., Lane, N.D.: Flower: A Friendly Federated Learning Research Framework. Tech. Rep. arXiv:2007.14390, arXiv (Mar 2022). https://doi.org/10.48550/arXiv.2007.14390, http://arxiv.org/abs/2007.14390, arXiv:2007.14390 [cs, stat] type: article
4. Chen, A., Chow, A., Davidson, A., DCunha, A., Ghodsi, A., Hong, S.A., Konwinski, A., Mewald, C., Murching, S., Nykodym, T., Ogilvie, P., Parkhe, M., Singh, A., Xie, F., Zaharia, M., Zang, R., Zheng, J., Zumar, C.: Developments in MLflow: A System to Accelerate the Machine Learning Lifecycle. In: Proceedings of the Fourth International Workshop on Data Management for End-to-End Machine Learning. pp. 1–4. DEEM'20, Association for Computing Machinery, New York, NY, USA (Jun 2020). https://doi.org/10.1145/3399579.3399867, https://doi.org/10.1145/3399579.3399867
5. Christopher, M., Nakahara, K., Bowd, C., Proudfoot, J.A., Belghith, A., Goldbaum, M.H., Rezapour, J., Weinreb, R.N., Fazio, M.A., Girkin, C.A., Liebmann, J.M., De Moraes, G., Murata, H., Tokumo, K., Shibata, N., Fujino, Y., Matsuura, M., Kiuchi, Y., Tanito, M., Asaoka, R., Zangwill, L.M.: Effects of Study Population, Labeling and Training on Glaucoma Detection Using Deep Learning Algorithms. Translational Vision Science & Technology **9**(2), 27 (Apr 2020). https://doi.org/10.1167/tvst.9.2.27, https://tvst.arvojournals.org/article.aspx?articleid=2765468
6. De Sales Carvalho, N.R., Da Conceição Leal Carvalho Rodrigues, M., De Carvalho Filho, A.O., Mathew, M.J.: Automatic method for glaucoma diagnosis using a three-dimensional convoluted neural network. Neurocomputing **438**, 72–83 (May 2021). https://doi.org/10.1016/j.neucom.2020.07.146, https://linkinghub.elsevier.com/retrieve/pii/S0925231221001090
7. Dikici, E., Bigelow, M., Prevedello, L.M., White, R.D., Erdal, B.S.: Integrating AI into radiology workflow: levels of research, production, and feedback maturity. Journal of Medical Imaging (Bellingham, Wash.) **7**(1), 016502 (Jan 2020). https://doi.org/10.1117/1.JMI.7.1.016502
8. Fan, R., Alipour, K., Bowd, C., Christopher, M., Brye, N., Proudfoot, J.A., Goldbaum, M.H., Belghith, A., Girkin, C.A., Fazio, M.A., Liebmann, J.M., Weinreb, R.N., Pazzani, M., Kriegman, D., Zangwill, L.M.: Detecting Glaucoma from Fundus Photographs Using Deep Learning without Convolutions. Ophthalmology Science **3**(1), 100233 (Mar 2023). https://doi.org/10.1016/j.xops.2022.100233, https://linkinghub.elsevier.com/retrieve/pii/S2666914522001221
9. Gheisari, S., Shariflou, S., Phu, J., Kennedy, P.J., Agar, A., Kalloniatis, M., Golzan, S.M.: A combined convolutional and recurrent neural network for enhanced glaucoma detection. Scientific Reports **11**(1), 1945 (Jan 2021). https://doi.org/10.1038/s41598-021-81554-4, https://www.nature.com/articles/s41598-021-81554-4
10. Gu, H., Liang, Y., Xu, Y., Williams, C.K., Magaki, S., Khanlou, N., Vinters, H., Chen, Z., Ni, S., Yang, C., Yan, W., Zhang, X.R., Li, Y., Haeri, M., Chen, X. Improving Workflow Integration with xPath: Design and Evaluation of a Human-AI Diagnosis System in Pathology. ACM Trans. Comput.-Hum. Interact. **30**(2), 28:1–28:37 (Mar 2023). https://doi.org/10.1145/3577011, https://doi.org/10.1145/3577011
11. Gu, X., Sabrina, F., Fan, Z., Sohail, S.: A Review of Privacy Enhancement Methods for Federated Learning in Healthcare Systems. International Journal of Environmental Research and Public Health **20**(15), 6539 (Aug 2023). https://doi.

org/10.3390/ijerph20156539, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10418741/

12. Hemelings, R., Elen, B., Barbosa-Breda, J., Blaschko, M.B., De Boever, P., Stalmans, I.: Deep learning on fundus images detects glaucoma beyond the optic disc. Scientific Reports **11**(1), 20313 (Oct 2021). https://doi.org/10.1038/s41598-021-99605-1, https://www.nature.com/articles/s41598-021-99605-1, publisher: Nature Publishing Group
13. Hemelings, R., Elen, B., Barbosa-Breda, J., Lemmens, S., Meire, M., Pourjavan, S., Vandewalle, E., Veire, S.V.d., Blaschko, M.B., Boever, P.D., Stalmans, I.: Accurate prediction of glaucoma from colour fundus images with a convolutional neural network that relies on active and transfer learning. Acta Ophthalmologica **98**(1), e94–e100 (2020). https://doi.org/10.1111/aos.14193, https://onlinelibrary.wiley.com/doi/abs/10.1111/aos.14193, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/aos.14193
14. Holzinger, A., Haibe-Kains, B., Jurisica, I.: Why imaging data alone is not enough: AI-based integration of imaging, omics, and clinical data. European Journal of Nuclear Medicine and Molecular Imaging **46**(13), 2722–2730 (Dec 2019). https://doi.org/10.1007/s00259-019-04382-9
15. Kanakaraj, P., Ramadass, K., Bao, S., Basford, M., Jones, L.M., Lee, H.H., Xu, K., Schilling, K.G., Carr, J.J., Terry, J.G., Huo, Y., Sandler, K.L., Netwon, A.T., Landman, B.A.: Workflow Integration of Research AI Tools into a Hospital Radiology Rapid Prototyping Environment. Journal of Digital Imaging **35**(4), 1023–1033 (Aug 2022). https://doi.org/10.1007/s10278-022-00601-2
16. Lee, S.S.Y., Mackey, D.A.: Glaucoma – risk factors and current challenges in the diagnosis of a leading cause of visual impairment. Maturitas **163**, 15–22 (Sep 2022). https://doi.org/10.1016/j.maturitas.2022.05.002, https://www.sciencedirect.com/science/article/pii/S0378512222000950
17. Leonardo, R., Gonçalves, J., Carreiro, A., Simões, B., Oliveira, T., Soares, F.: Impact of Generative Modeling for Fundus Image Augmentation With Improved and Degraded Quality in the Classification of Glaucoma. IEEE Access **10**, 111636–111649 (2022). https://doi.org/10.1109/ACCESS.2022.3215126, conference Name: IEEE Access
18. Li, R.C., Asch, S.M., Shah, N.H.: Developing a delivery science for artificial intelligence in healthcare. npj Digital Medicine **3**(1), 1–3 (Aug 2020). https://doi.org/10.1038/s41746-020-00318-y, https://www.nature.com/articles/s41746-020-00318-y, publisher: Nature Publishing Group
19. Li, Z., He, Y., Keel, S., Meng, W., Chang, R.T., He, M.: Efficacy of a Deep Learning System for Detecting Glaucomatous Optic Neuropathy Based on Color Fundus Photographs. Ophthalmology **125**(8), 1199–1206 (Aug 2018). https://doi.org/10.1016/j.ophtha.2018.01.023, https://linkinghub.elsevier.com/retrieve/pii/S0161642017335650
20. Lima-Cabrita, A.V., Duarte, S.R., Fernandes, E., Marques-Neves, C., Rodrigues, W., Marques, R., santos, v., carrapico, E., Pazos, M., Stalmans, I., Ferreira, J.T., Pinto, L.A.: AI-(EM)POWERED SCREENING: GLAUCOMA POST STUDY. Investigative Ophthalmology & Visual Science **65**(7), 605 (Jun 2024)
21. Marques-Neves, C., Lima-Cabrita, A.V., Marques, R., santos, v., carrapico, E., Rodrigues, W., Stalmans, I., Pinto, L.A., Ferreira, J.T.: Associating glaucoma and diabetic retinopathy: opportunistic AI-powered glaucoma screening. Investigative Ophthalmology & Visual Science **65**(7), 604 (Jun 2024)

22. Martins, J., Cardoso, J.S., Soares, F.: Offline computer-aided diagnosis for Glaucoma detection using fundus images targeted at mobile devices. Computer Methods and Programs in Biomedicine **192**, 105341 (Aug 2020). https://doi.org/10.1016/j.cmpb.2020.105341, https://linkinghub.elsevier.com/retrieve/pii/S0169260719312015
23. Matos, D.B., Barão, R.C., José, P., Cabrita, A., Barata, A.D., Pinto, L.A.: Glaucoma triage system: results of implementing a virtual clinic. Graefe's Archive for Clinical and Experimental Ophthalmology = Albrecht Von Graefes Archiv Fur Klinische Und Experimentelle Ophthalmologie **261**(8), 2367–2374 (Aug 2023). https://doi.org/10.1007/s00417-023-06039-8
24. M.B, S., .M, S., Raja, S., Joyson, A., Latha, C., Rachel, S., Thavasimuthu, A., Thavasimuthu, R., Waji, Y.: Segmentation and Classification of Glaucoma Using U-Net with Deep Learning Model. Journal of Healthcare Engineering **2022**, 1–10 (Feb 2022). https://doi.org/10.1155/2022/1601354
25. Mercer, R., Alaghband, P.: The value of virtual glaucoma clinics: a review. Eye **38**(10), 1840–1844 (Jul 2024). https://doi.org/10.1038/s41433-024-03056-7, https://www.nature.com/articles/s41433-024-03056-7, publisher: Nature Publishing Group
26. Natarajan, D., Sankaralingam, E., Balraj, K., Karuppusamy, S.: A deep learning framework for glaucoma detection based on robust optic disc segmentation and transfer learning. International Journal of Imaging Systems and Technology **32**(1), 230–250 (2022). https://doi.org/10.1002/ima.22609, https://onlinelibrary.wiley.com/doi/abs/10.1002/ima.22609, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ima.22609
27. Nikolaidou, A., Tsaousis, K.T.: Teleophthalmology and Artificial Intelligence As Game Changers in Ophthalmic Care After the COVID-19 Pandemic. Cureus **13**(7), e16392 (Jul 2021). https://doi.org/10.7759/cureus.16392
28. Orlando, J.I., Fu, H., Breda, J.B., Keer, K.v., Bathula, D.R., Diaz-Pinto, A., Fang, R., Heng, P.A., Kim, J., Lee, J., Lee, J., Li, X., Liu, P., Lu, S., Murugesan, B., Naranjo, V., Phaye, S.S.R., Shankaranarayana, S.M., Sikka, A., Son, J., Hengel, A.v.d., Wang, S., Wu, J., Wu, Z., Xu, G., Xu, Y., Yin, P., Li, F., Zhang, X., Xu, Y., Bogunović, H.: REFUGE Challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. Medical Image Analysis **59**, 101570 (2020). https://doi.org/https://doi.org/10.1016/j.media.2019.101570, https://www.sciencedirect.com/science/article/pii/S1361841519301100
29. Ovreiu, S., Paraschiv, E.A., Ovreiu, E.: Deep Learning & Digital Fundus Images: Glaucoma Detection using DenseNet. In: 2021 13th International Conference on Electronics, Computers and Artificial Intelligence (ECAI). pp. 1–4. IEEE, Pitesti, Romania (Jul 2021). https://doi.org/10.1109/ECAI52376.2021.9515188, https://ieeexplore.ieee.org/document/9515188/
30. Schuster, A.K., Wagner, F.M., Pfeiffer, N., Hoffmann, E.M.: Risk factors for open-angle glaucoma and recommendations for glaucoma screening. Der Ophthalmologe **118**(2), 145–152 (Jul 2021). https://doi.org/10.1007/s00347-021-01378-5, https://doi.org/10.1007/s00347-021-01378-5
31. Shyamalee, T., Meedeniya, D., Lim, G., Karunarathne, M.: Automated Tool Support for Glaucoma Identification With Explainability Using Fundus Images. IEEE Access **12**, 17290–17307 (2024). https://doi.org/10.1109/ACCESS.2024.3359698, https://ieeexplore.ieee.org/document/10416867/

32. Simons, A.S., Vercauteren, J., Barbosa-Breda, J., Stalmans, I.: Shared Care and Virtual Clinics for Glaucoma in a Hospital Setting. Journal of Clinical Medicine **10**(20), 4785 (Jan 2021). https://doi.org/10.3390/jcm10204785, https://www.mdpi.com/2077-0383/10/20/4785, number: 20 Publisher: Multidisciplinary Digital Publishing Institute
33. Soltan, A.A.S., Thakur, A., Yang, J., Chauhan, A., D'Cruz, L.G., Dickson, P., Soltan, M.A., Thickett, D.R., Eyre, D.W., Zhu, T., Clifton, D.A.: A scalable federated learning solution for secondary care using low-cost microcomputing: privacy-preserving development and evaluation of a COVID-19 screening test in UK hospitals. The Lancet Digital Health **6**(2), e93–e104 (Feb 2024). https://doi.org/10.1016/S2589-7500(23)00226-1, https://www.thelancet.com/journals/landig/article/PIIS2589-7500(23)00226-1/fulltext, publisher: Elsevier
34. Taribagil, P., Hogg, H.J., Balaskas, K., Keane, P.A.: Integrating artificial intelligence into an ophthalmologist's workflow: obstacles and opportunities. Expert Review of Ophthalmology **18**(1), 45–56 (Jan 2023). https://doi.org/10.1080/17469899.2023.2175672, https://doi.org/10.1080/17469899.2023.2175672, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/17469899.2023.2175672
35. US Preventive Services Task Force: Screening for Primary Open-Angle Glaucoma: US Preventive Services Task Force Recommendation Statement. JAMA **327**(20), 1992–1997 (May 2022). https://doi.org/10.1001/jama.2022.7013, https://doi.org/10.1001/jama.2022.7013
36. de Vente, C., Vermeer, K.A., Jaccard, N., Wang, H., Sun, H., Khader, F., Truhn, D., Aimyshev, T., Zhanibekuly, Y., Le, T.D., Galdran, A., Ballester, M.G., Carneiro, G., Devika, R.G., Sethumadhavan, H.P., Puthussery, D., Liu, H., Yang, Z., Kondo, S., Kasai, S., Wang, E., Durvasula, A., Heras, J., Zapata, M., Araújo, T., Aresta, G., Bogunović, H., Arikan, M., Lee, Y.C., Cho, H.B., Choi, Y.H., Qayyum, A., Razzak, I., van Ginneken, B., Lemij, H.G., Sánchez, C.I.: AIROGS: Artificial Intelligence for Robust Glaucoma Screening Challenge. IEEE Transactions on Medical Imaging **43**(1), 542–557 (2024). https://doi.org/10.1109/TMI.2023.3313786
37. Wassel, M., Hamdi, A.M., Adly, N., Torki, M.: Vision Transformers Based Classification for Glaucomatous Eye Condition. In: 2022 26th International Conference on Pattern Recognition (ICPR). pp. 5082–5088. IEEE, Montreal, QC, Canada (Aug 2022). https://doi.org/10.1109/ICPR56361.2022.9956086, https://ieeexplore.ieee.org/document/9956086/
38. Wenderott, K., Gambashidze, N., Weigl, M.: Integration of Artificial Intelligence Into Sociotechnical Work Systems—Effects of Artificial Intelligence Solutions in Medical Imaging on Clinical Efficiency: Protocol for a Systematic Literature Review. JMIR Research Protocols **11**(12), e40485 (Dec 2022). https://doi.org/10.2196/40485, https://www.researchprotocols.org/2022/12/e40485, company: JMIR Research Protocols Distributor: JMIR Research Protocols Institution: JMIR Research Protocols Label: JMIR Research Protocols Publisher: JMIR Publications Inc., Toronto, Canada
39. Xu, Y., Hu, M., Liu, H., Yang, H., Wang, H., Lu, S., Liang, T., Li, X., Xu, M., Li, L., Li, H., Ji, X., Wang, Z., Li, L., Weinreb, R.N., Wang, N.: A hierarchical deep learning approach with transparency and interpretability based on small samples for glaucoma diagnosis. npj Digital Medicine **4**(1), 48 (Mar 2021). https://doi.org/10.1038/s41746-021-00417-4, https://www.nature.com/articles/s41746-021-00417-4

40. Zedan, M., Zulkifley, M., Ibrahim, A., Moubark, A., Kamari, N., Abdani, S.: Automated Glaucoma Screening and Diagnosis Based on Retinal Fundus Images Using Deep Learning Approaches: A Comprehensive Review. Diagnostics **13**(13), 2180 (Jun 2023). https://doi.org/10.3390/diagnostics13132180, https://www.mdpi.com/2075-4418/13/13/2180
41. Zhu, Y., Salowe, R., Chow, C., Li, S., Bastani, O., O'Brien, J.M.: Advancing Glaucoma Care: Integrating Artificial Intelligence in Diagnosis, Management, and Progression Detection. Bioengineering **11**(2), 122 (Feb 2024). https://doi.org/10.3390/bioengineering11020122, https://www.mdpi.com/2306-5354/11/2/122, number: 2 Publisher: Multidisciplinary Digital Publishing Institute
42. Zhuo Zhang, Feng Shou Yin, Jiang Liu, Wing Kee Wong, Ngan Meng Tan, Beng Hai Lee, Jun Cheng, Tien Yin Wong: ORIGA$^{\text{-light}}$: An online retinal fundus image database for glaucoma analysis and research. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology. pp. 3065–3068. IEEE, Buenos Aires (Aug 2010). https://doi.org/10.1109/IEMBS.2010.5626137, http://ieeexplore.ieee.org/document/5626137/