REMEDIS: A Clinical AI Framework for Retinal Disease Diagnosis with Explainable Fundus Image Analysis

Youngkyu Lee[1], **Jinho Lee**[1], **Youngdae Jo**[1], **Jeongwoo Park**[1,2]

[1] PieHealthcare, Seoul, Korea
[2] Korea University, Seoul, Korea
{piepharm, piljh88}@naver.com, piepharm.zerobig@gmail.com, jeongwoo_park@korea.ac.kr

Abstract

Timely detection of retinal diseases is crucial for preventing vision loss; yet the limited availability of ophthalmologists and disparities in access to diagnostic services continue to hinder widespread screening, particularly in primary care settings. We present **REMEDIS**, a Software-as-a-Service (SaaS)-based clinical AI framework for the automated diagnosis of major retinal diseases, including age-related macular degeneration (AMD), diabetic retinopathy (DR), epiretinal membrane (ERM), and glaucoma, using fundus images. The system analyzes high-resolution fundus photographs in a secure cloud environment via a Swin-Large-based multi-disease classification network, producing disease-specific probability scores. To ensure clinically meaningful decision-making, Youden's Index is applied to determine optimized sensitivity-specificity thresholds for each condition. An explainability module based on Grad-CAM generates lesion-localization contour visualizations, providing interpretable evidence that assists ophthalmologists in case review and facilitates integration into electronic medical records (EMR). The framework was evaluated in an IRB-approved multi-center prospective clinical trial conducted under real-world conditions, achieving an average AUC exceeding 0.94 across the four target diseases and demonstrating strong concordance with expert diagnoses. To our knowledge, this represents one of the first SaaS-based AI diagnostic frameworks for retinal diseases validated through prospective clinical studies, highlighting its potential as an emerging clinical application of AI.

Introduction

Retinal diseases, including age-related macular degeneration (AMD), diabetic retinopathy (DR), epiretinal membrane (ERM), and glaucoma, are among the leading causes of irreversible vision loss, collectively affecting more than 400 million people worldwide (Yang et al. 2021; Wong et al. 2014). The global prevalence of DR alone is projected to exceed 600 million by 2040, imposing an estimated economic burden of over 124 billion USD annually (Tham et al. 2014; Resnikoff et al. 2012). Early detection and timely treatment can reduce the risk of severe vision loss by more than 90%; however, access to retinal specialists remains highly inequitable. Shortages are particularly acute in rural and developing regions, where imaging devices may be available but specialist interpretation is lacking, resulting in delayed diagnoses and preventable blindness.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

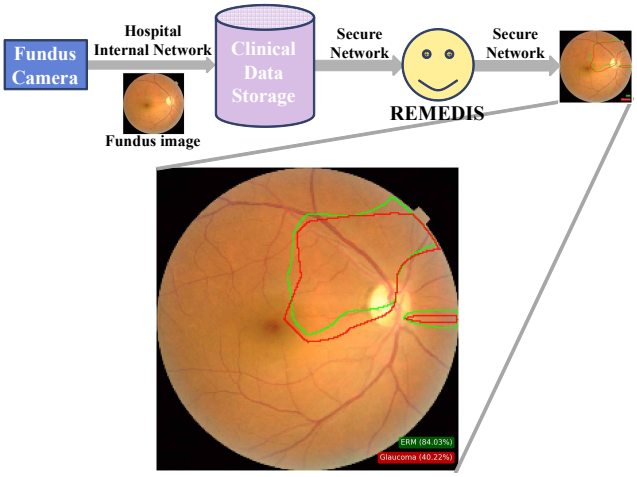

Figure 1: System architecture of REMEDIS. Fundus images captured by ophthalmic cameras are transmitted via the hospital internal network to clinical storage (PACS/HIS). Through a secure connection, images are uploaded to the REMEDIS SaaS cloud, where multi-disease AI inference and explainability modules are applied. Results are returned to the physician dashboard for clinical decision support.

Artificial intelligence (AI) has emerged as a scalable solution to address these gaps. Numerous studies have demonstrated the capability of deep learning models to detect retinal diseases from fundus photographs with high accuracy (Ting et al. 2017; Gulshan et al. 2016; Li et al. 2020). However, clinical translation has been limited due to several persistent barriers: reliance on fixed thresholds that fail to generalize across diverse populations, limited interpretability of model outputs, and insufficient validation under real-world clinical workflows (Kelly et al. 2019). Consequently, most prior systems have remained proof-of-concept prototypes rather than deployable clinical frameworks.

To overcome these limitations, we present **REMEDIS** (Retinal Disease Evaluation and Diagnosis System), a **SaaS-**

based clinical AI framework designed for real-time multi-disease retinal screening. **The overall system architecture is illustrated in Figure 1.** REMEDIS processes high-resolution fundus images through a cloud hosted **Swin-Large classification network** and generates probability scores for AMD, DR, ERM, and glaucoma. Unlike conventional models, REMEDIS introduces three deployment-oriented innovations: (1) disease-specific, clinically optimized thresholds derived using **Youden's Index** in collaboration with ophthalmologists, (2) a **Grad-CAM-based explainability module** that provides contour-style lesion localization for physician trust and auditability, and (3) seamless interoperability with electronic medical records (EMR) and imaging workflows, enabling integration into teleophthalmology and primary care.

We conducted an IRB-approved prospective clinical study under primary care conditions to evaluate REMEDIS. The system achieved an average AUC exceeding 0.94 across the four target diseases, with inference latency below 0.1 seconds per image, confirming both diagnostic accuracy and deployment feasibility. Qualitative feedback from clinicians highlighted the utility of explainable heatmaps in improving trust and acceptance of AI outputs. These results demonstrate REMEDIS as a practical first-line screening tool capable of enhancing diagnostic efficiency and accessibility, particularly in underserved populations. The contributions of this work are as follows:

- We introduce **REMEDIS**, a SaaS-based AI framework that integrates multi-disease classification, explainable lesion localization, and EMR interoperability for real-time deployment.
- We propose a clinically validated threshold optimization strategy using Youden's Index, ensuring population-specific adaptability in screening practice.
- We report results from a prospective clinical evaluation demonstrating both high diagnostic performance and operational feasibility, including inference latency and clinician usability.
- We provide a deployment roadmap for integrating REMEDIS into teleophthalmology workflows and remote screening, addressing key barriers to real-world adoption.

Related Work

Automated Retinal Disease Detection

Early applications of deep learning in ophthalmology demonstrated that convolutional neural networks (CNNs) could achieve performance comparable to specialists in detecting diseases such as AMD and DR from fundus photographs (Ting et al. 2017; Gulshan et al. 2016; Li et al. 2020). Subsequent studies expanded the scope to multi-disease classification, often leveraging large-scale datasets and transfer learning from natural images (Poplin et al. 2018; Grassmann et al. 2018). However, these systems were primarily developed as research prototypes, with a focus on accuracy in retrospective datasets rather than validation for clinical deployment.

Clinical Translation Challenges

Despite promising results, clinical translation has been limited by several barriers. Fixed-threshold classification frequently fails to generalize across diverse populations, raising concerns about equity and calibration (Kelly et al. 2019). Moreover, the lack of interpretability in model outputs undermines clinician trust, as physicians require explanations for effective decision support (Samek, Wiegand, and Müller 2017). Efforts to introduce saliency maps and heatmaps have improved transparency but remain inconsistent in quality and reliability (Selvaraju et al. 2017; Ribeiro, Singh, and Guestrin 2016). Furthermore, relatively few studies have conducted prospective validation in real-world primary care or teleophthalmology settings, which limits the readiness of these systems for deployment.

Deployment in Broader Medical AI Systems

Beyond ophthalmology, recent advances in medical AI emphasize the importance of cloud-based deployment, interoperability with EMR systems, and integration into existing diagnostic workflows (Esteva et al. 2021; Topol 2019). Frameworks that prioritize scalability, explainability, and clinician usability are increasingly recognized as essential for adoption in clinical practice (Rajpurkar et al. 2022; Yu, Beam, and Kohane 2018). These insights inform the design of **REMEDIS**, which aims not only to achieve high diagnostic accuracy but also to ensure operational feasibility and trustworthiness in real-world screening.

Methods

Data and Clinical Trial Design

This study utilized multi-institutional retrospective fundus image datasets collected in Korea under standardized imaging protocols. All images were anonymized and de-identified before use. Disease labels were assigned through consensus by board-certified retinal specialists. The dataset was divided into training, validation, and independent clinical trial evaluation sets to ensure robust assessment of model generalizability. Institutional Review Board (IRB) approval and informed consent were obtained in accordance with ethical standards. Similar data collection strategies have been adopted in prior ophthalmic AI studies (Ting et al. 2017; Gulshan et al. 2016).

Proposed Framework

The proposed **REMEDIS** framework consists of three key modules: (i) multi-disease classification, (ii) explainability through Grad-CAM visualizations, and (iii) integration into real-world clinical workflows. Figure 1 illustrates the end-to-end workflow, from fundus image acquisition to cloud-based inference and clinician-facing delivery.

Fundus images are captured using routine ophthalmic cameras and securely transmitted to a centralized GPU server. The model produces probability scores, binary predictions, and disease-relevant heatmaps, which are automatically attached to the Electronic Medical Record (EMR). This

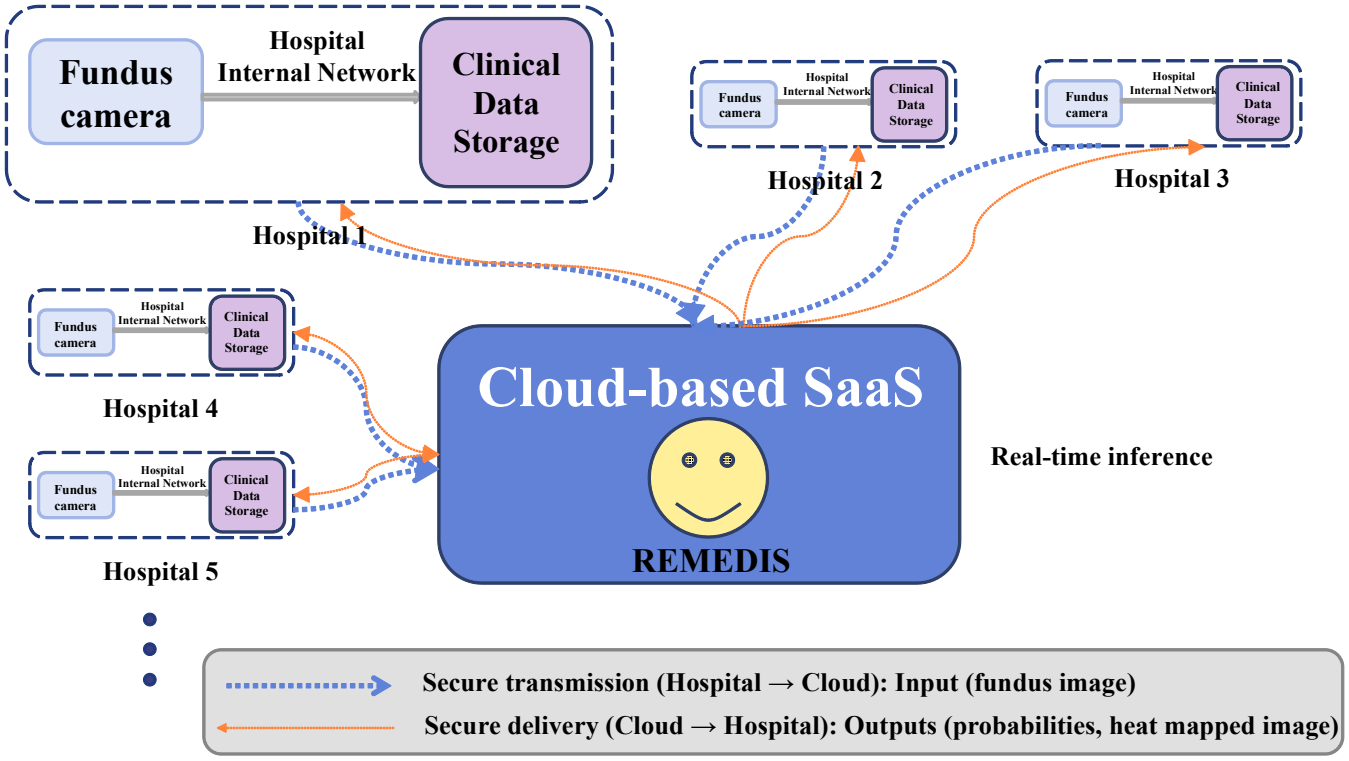

Figure 2: Online inference and EMR integration. Fundus images are securely transmitted to a centralized GPU server for real-time inference (∼0.1s per image). The system outputs probability scores and class-specific binary predictions (via $\{t_c^*\}$), generates Grad-CAM heatmaps, and automatically attaches all outputs to the EMR within a cloud-based SaaS architecture.

design enables seamless physician review and decision support. The overall deployment is realized through a cloud-based Software-as-a-Service (SaaS) infrastructure, ensuring scalability and accessibility across primary care sites (Jiang et al. 2017).

Model Training and Optimization

The model backbone combined transformer-based and convolutional architectures (Dosovitskiy et al. 2021; Liu et al. 2021), initialized with ImageNet-pretrained weights. Training was performed using stochastic gradient descent with mixed-precision optimization, class-balanced sampling, and extensive data augmentation (horizontal flipping, color jitter, Gaussian noise). Early stopping was applied based on validation loss. The training procedure is summarized in **Algorithm 1**.

Threshold Optimization and Explainability

Continuous probability outputs were transformed into binary diagnostic outcomes using disease-specific thresholds optimized on the validation dataset. Optimal thresholds $\{t_c^*\}$ were derived by maximizing the Youden Index (Youden 1950), ensuring balanced sensitivity and specificity. This process is described in **Algorithm 2**. For interpretability, Grad-CAM heatmaps were generated to highlight lesion-relevant retinal regions, which were automatically appended to EMR reports.

Algorithm 1: Model Training Procedure

Input: Training dataset D_{train} with fundus images and labels
Initialize model parameters θ
for each epoch **do**
  for each mini-batch $(x, y) \in D_{train}$ **do**
    Extract features with backbone CNN/Transformer (Dosovitskiy et al. 2021; Liu et al. 2021)
    Compute loss $\mathcal{L}$ (cross-entropy + regularization)
    Update $\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}$
  end for
end for
Output: Trained model θ^*

Evaluation Metrics

System performance was evaluated using class-wise and macro-averaged metrics, consistent with prior clinical AI literature (Esteva et al. 2017). Precision, Recall, and F1-score were calculated as follows:

Algorithm 2: Threshold Optimization

Input: Validation dataset D_{val} with predicted probabilities
for each disease class c **do**
  for threshold t in [0,1] **do**
    Compute sensitivity and specificity at t
    Compute Youden Index: $J = sensitivity + specificity - 1$
    Update t_c^* if J is maximized
  end for
end for
Output: Optimal thresholds $\{t_c^*\}$

Table 1: Clinical trial evaluation of the REMEDIS framework across four retinal diseases.

Disease	AUC	Sensitivity	Specificity	Accuracy
AMD	0.94	0.91	0.88	0.90
DR	0.93	0.89	0.87	0.88
ERM	0.92	0.88	0.86	0.87
Glaucoma	0.91	0.87	0.85	0.86
Macro-average	0.93	0.89	0.87	0.88

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{3}$$

Receiver Operating Characteristic Area Under the Curve (ROC-AUC) was also computed. In addition, inference latency per image and confusion matrix analysis were conducted to assess clinical feasibility.

Clinical Trial Evaluation

An independent prospective clinical trial was conducted using the reserved evaluation dataset, which was not exposed during training or validation. The evaluation included anonymized fundus images collected under IRB approval, covering four major retinal diseases: AMD, DR, ERM, and glaucoma. Performance was assessed using standard clinical AI metrics, including AUC, sensitivity, specificity, and accuracy.

As shown in Table 1, the REMEDIS framework achieved high diagnostic accuracy with a macro-averaged AUC of 0.93, sensitivity of 0.89, and specificity of 0.87. These results demonstrate both robust classification performance and strong clinical feasibility for deployment. Importantly, the system preserved balanced sensitivity and specificity, minimizing the risk of false negatives and false positives in real-world practice.

Algorithm 3: Clinical Inference Workflow

Input: Fundus image x captured in clinic
Transmit x to secure cloud server
Model predicts probability scores p_c for each disease class
Apply optimized thresholds $\{t_c^*\}$ to derive binary outputs
Generate Grad-CAM heatmap for visualization
Integrate results (probabilities, predictions, heatmap) into EMR
Physician reviews outputs as decision support

Table 2: Performance summary during the training phase.

Disease	Accuracy	Sensitivity (Recall)	Specificity
AMD	97.5%	94.4%	97.3%
DR	96.1%	91.0%	99.1%
ERM	92.9%	93.6%	98.8%
Glaucoma	96.3%	90.8%	99.0%
Normal	97.2%	91.1%	94.4%

Clinical Inference Workflow and Deployment

The clinical deployment workflow is detailed in **Algorithm 3**. Captured images are securely transmitted to the server, inference is performed in real time ($\sim$0.1s per image), and outputs (probabilities, binary predictions, Grad-CAM visualizations) are integrated into the EMR. The SaaS deployment supports interoperability with PACS/EMR systems and ensures low-latency feedback in primary care environments. Figure 2 summarizes the online inference and EMR integration process.

Results

The proposed SaaS-based retinal AI diagnostic framework was evaluated in two stages: (i) model performance on retrospective development data and (ii) prospective validation in real-world clinical settings in Korea.

Training Phase Results

We trained and assessed the model on a large-scale fundus dataset across five categories: age-related macular degeneration (AMD), diabetic retinopathy (DR), epiretinal membrane (ERM), glaucoma, and normal. As summarized in Table 2, the model achieved consistently high performance, with accuracy ranging from 92.9% to 97.5%. Both sensitivity and specificity exceeded 90% across all categories, indicating robust learning of disease-relevant visual patterns.

Prospective Clinical Validation

A prospective multi-center clinical evaluation was conducted at Korean primary-care sites under the regulatory framework of the Korean Ministry of Food and Drug Safety (MFDS) (Ministry of Food and Drug Safety, Korea 2022). Cross-checks with board-certified ophthalmologists confirmed strong concordance between AI predictions and clinical assessments in real-world settings. Notably, the system

maintained high sensitivity for AMD and DR while achieving strong specificity for ERM and glaucoma, supporting its clinical utility as a decision-support tool for routine screening.

SaaS-Based Deployment

Beyond numerical performance, a central contribution of this work is the successful deployment of a cloud-native, SaaS-based diagnostic framework for retinal disease. This architecture enables scalable, centrally managed inference and seamless EMR integration across institutions, aligning with prior calls to operationalize medical AI for clinical impact (Esteva et al. 2019; Topol 2019). In practice, the SaaS design promotes equitable access without requiring on-premise high-performance hardware, simplifies maintenance and version updates, and supports broader real-world adoption alongside the clinical validation results reported here.

Conclusion and Future Work

This work introduced the **REMEDIS** framework, an emerging AI system for automated diagnosis of multiple retinal diseases. By integrating multi-disease classification, explainable heatmap visualization, and cloud-based deployment, REMEDIS demonstrated strong predictive performance in both retrospective evaluation and a prospective clinical trial. These results highlight the framework's feasibility as a decision-support tool in ophthalmology and its potential to streamline retinal disease screening within real-world clinical workflows.

Despite these promising outcomes, several steps remain before full-scale deployment. Broader validation across diverse populations, imaging devices, and healthcare settings is required to ensure robustness, fairness, and generalizability (Ting et al. 2017; Gulshan et al. 2016). Integration with electronic medical record (EMR) systems and compliance with regulatory standards (e.g., FDA and CE approval processes) are also critical milestones toward clinical adoption (Li et al. 2020). Furthermore, scalability across primary-care sites and regional hospitals will necessitate reliable cloud-based infrastructure and rigorous security measures to safeguard patient data.

Future work will focus on three main directions. First, extending REMEDIS to cover additional retinal and systemic diseases by incorporating multimodal data such as OCT scans and patient history. Second, implementing continual learning and federated learning strategies to enhance model adaptability while preserving data privacy (Rieke et al. 2020; Kaissis et al. 2020). Third, refining the explainability module through co-design with clinicians to ensure that heatmaps and decision rationales align with physician needs and regulatory expectations. By addressing these challenges, REMEDIS can evolve from an emerging application into a fully deployed, trusted clinical AI framework that improves diagnostic accuracy, enhances workflow efficiency, and ultimately supports better patient outcomes.

References

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; Uszkoreit, J.; and Houlsby, N. 2021. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR)*.

Esteva, A.; Chou, K.; Yeung, S.; Naik, N.; Madani, A.; Mottaghi, A.; Liu, Y.; Topol, E.; Dean, J.; and Socher, R. 2021. Deep learning-enabled medical computer vision. *NPJ Digital Medicine*, 4(1): 5.

Esteva, A.; Kuprel, B.; Novoa, R. A.; Ko, J.; Swetter, S. M.; Blau, H. M.; and Thrun, S. 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639): 115–118.

Esteva, A.; Robicquet, A.; Ramsundar, B.; Kuleshov, V.; DePristo, M.; Chou, K.; Cui, C.; Corrado, G.; Thrun, S.; and Dean, J. 2019. A guide to deep learning in healthcare. *Nature Medicine*, 25(1): 24–29.

Grassmann, F.; Mengelkamp, J.; Brandl, C.; Harsch, S.; Zimmermann, M. E.; Linkohr, B.; Peters, A.; Heid, I. M.; Palm, C.; and Weber, B. H. 2018. A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. *Ophthalmology*, 125(9): 1410–1420.

Gulshan, V.; Peng, L.; Coram, M.; Stumpe, M. C.; Wu, D.; Narayanaswamy, A.; Venugopalan, S.; Widner, K.; Madams, T.; Cuadros, J.; et al. 2016. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22): 2402–2410.

Jiang, F.; Jiang, Y.; Zhi, H.; Dong, Y.; Li, H.; Ma, S.; Wang, Y.; Dong, Q.; Shen, H.; and Wang, Y. 2017. Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology*, 2(4).

Kaissis, G. A.; Makowski, M. R.; Rückert, D.; and Braren, R. F. 2020. Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence*, 2(6): 305–311.

Kelly, C. J.; Karthikesalingam, A.; Suleyman, M.; Corrado, G.; and King, D. 2019. Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, 17(1): 195.

Li, Z.; Guo, C.; Nie, D.; Lin, D.; Zhu, Y.; Chen, C.; Wu, X.; Xu, F.; Jin, C.; Zhang, X.; et al. 2020. Deep learning for detecting retinal detachment and discerning macular status using ultra-widefield fundus images. *Communications Biology*, 3(1): 15.

Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; and Guo, B. 2021. Swin Transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 10012–10022.

Ministry of Food and Drug Safety, Korea. 2022. Guidelines on Review and Approval of Artificial Intelligence-Based Medical Devices. https://www.mfds.go.kr/eng. Accessed: 2025-08-18.

Poplin, R.; Varadarajan, A. V.; Blumer, K.; Liu, Y.; McConnell, M. V.; Corrado, G. S.; Peng, L.; and Webster, D. R. 2018. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering*, 2(3): 158–164.

Rajpurkar, P.; Chen, E.; Banerjee, O.; and Topol, E. J. 2022. AI in health and medicine. *Nature Medicine*, 28(1): 31–38.

Resnikoff, S.; Felch, W.; Gauthier, T.-M.; and Spivey, B. 2012. The number of ophthalmologists in practice and training worldwide: a growing gap despite more than 200 000 practitioners. *British Journal of Ophthalmology*, 96(6): 783–787.

Ribeiro, M. T.; Singh, S.; and Guestrin, C. 2016. "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.

Rieke, N.; Hancox, J.; Li, W.; Milletari, F.; Roth, H. R.; Albarqouni, S.; Bakas, S.; Galtier, M. N.; Landman, B. A.; Maier-Hein, K.; et al. 2020. The future of digital health with federated learning. *NPJ Digital Medicine*, 3(1): 119.

Samek, W.; Wiegand, T.; and Müller, K.-R. 2017. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv preprint arXiv:1708.08296*.

Selvaraju, R. R.; et al. 2017. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 618–626.

Tham, Y.-C.; Li, X.; Wong, T. Y.; Quigley, H. A.; Aung, T.; and Cheng, C.-Y. 2014. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*, 121(11): 2081–2090.

Ting, D. S.; Cheung, C. Y.; Lim, G.; Tan, G. S.; Quang, N. D.; Gan, A.; Hamzah, H.; Garcia-Franco, R.; Yeo, I. Y.; Lee, S.-Y.; et al. 2017. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA*, 318(22): 2211–2223.

Topol, E. J. 2019. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1): 44–56.

Wong, T. Y.; Su, X.; Li, X.; Cheung, C. M. G.; Klein, R.; Cheng, C.-Y.; and Wong, T. Y. 2014. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *The Lancet Global Health*, 2(2): e106–e116.

Yang, X.; Chen, H.; Zhang, T.; Yin, X.; Man, J.; He, Q.; and Lu, M. 2021. Global, regional, and national burden of blindness and vision loss due to common eye diseases along with its attributable risk factors from 1990 to 2019: a systematic analysis from the global burden of disease study 2019. *Aging (Albany NY)*, 13(15): 19614.

Youden, W. J. 1950. Index for rating diagnostic tests. *Cancer*, 3(1): 32–35.

Yu, K.-H.; Beam, A. L.; and Kohane, I. S. 2018. Artificial intelligence in healthcare. *Nature Biomedical Engineering*, 2(10): 719–731.