Journal of Theoretical and Applied Information Technology
15th April 2026. Vol.104. No.7

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FEDERATED INTELLIGENCE IN OPHTHALMOLOGY: PRIVACY-PRESERVING COLLABORATION FOR MULTICENTER AMBLYOPIA MODEL DEVELOPMENT

JAYA LAKSHMI C[1], DR. N. SATHEESH[2], DR. M. KUMARASAN[3]

[1]Research Scholar, Department of CSE, School of Computer Science and Engineering, FET, JAIN (Deemed-to-be University), Jain Global Campus, Kanakapura Road, Bangalore, Karnataka, India.

[2]Professor, Department of CSE, SCSE, Faculty of Engineering and Technology, Jain Global Campus, Jain (Deemed-to-be University), Kanakapura Road, Bangalore, Karnataka, India.

[3]Associate Professor, Department of CSE, SCSE, Faculty of Engineering and Technology, Jain Global Campus, Jain (Deemed-to-be University), Kanakapura Road, Bangalore, Karnataka, India.

E-mail: [1]jayavamsi2006@gmail.com, [2]nsatheesh1983@gmail.com, [3]m.kumaresan@jainuniversity.ac.in

ABSTRACT

The paper under consideration introduces a federated system of deep-learning to support privacy-safe and multicenter cooperation in the automated process of diagnosis and prognosis of amblyopia. Its key objective will be to create a standard, multimodal ophthalmic corpus (fusing fundus-photography, optical coherence tomography (OCT), eye-tracking signals, visual induced potentials (VEP), and demographic data). Through generic adversarial networks (GANs) augmentation and harmonization, such a layer of data acts out to reduce class imbalance and boost the generalizability of downstream models. The second goal is focused on the advanced feature representation based on the work of a hybrid Convolutional Neural Network (CNN)-Transformer structure. This type of architecture is able to encode both spatial and temporal relations both in imaging and behavioral modalities simultaneously, and therefore forms more comprehensive multimodal embeddings that enhance the discrimination of disease. The explainable artificial intelligence (XAI) methods such as Grad-CAM and SHAP are included in the tertiary objective to enhance the interpretability and clinical transparency of diagnostic predictions. Federated learning (FedAvg, FedProx) allows learning model optimization among participating institutions without the need to share their data, hence, protecting patient privacy and regulating compliance with regulations. Experimental confirmation using synthetic and real-world pediatric datasets have shown a higher diagnostic accuracy (AUC > 0.95), high across-center generalizability and readable visual reasoning maps to aid the decision-making of clinicians. The given work, thus, predetermines the course of safe, transparent, and cooperative, AI-driven ophthalmic diagnostics, accelerating the process of clinical implementation and universal transfer of amblyopia screening.

Keywords: *Federated Learning, Amblyopia Diagnosis, Multimodal Deep Learning, Explainable AI (XAI), Privacy-Preserving Collaboration, Ophthalmic Intelligence.*

1. INTRODUCTION

The paradigm of Federated Ophthalmic Intelligence, a paradigm shift that will connect closed clinical settings with collaborative development in artificial-intelligence. The narrative also capitalizes on the presented hybrid CNN-Transformer-XAI architecture through addressing each of the named challenges and limitations in a systematized effort. Due to the resulting discourse, there is an emphasis on urgency, innovation, and ethical responsibility, making the presented framework one really groundbreaking change in pediatric ophthalmology, as opposed to an incremental enhancement.

Amblyopia is the most common cause of monocular impairment of vision in children causing the impairment of about 2-3 percent globally. The disorder occurs sneakily over the critical neuro-visual period of birth to seven years old and early

diagnosis causes irreparable cortical inhibition and profound impairment. Traditional methods of detection include; subjective Snellen chart tests and cycloplegic refraction with sensitivity rates of less than 70 per cent of all preschool-detected cases due to the fact that the screening methods are not well adhered to by children, inter-observer reliability (0.45 0.60), and due to the heavy case loads that ophthalmologists have to accommodate in limited resources (1:10000). Available unimodal pipelines to simple imaging modalities like fundus-photography classifiers (AUC=0.91) or OCT-based segmentation models post a amazing performance (~85 -89 times) in monoinstitutional controlled trials. They however degenerate severely when face to face with the heterogeneity of the real world: domain artifacts due to scanner variation (zeiss vs. heidelberg), signal to noise loss (between 15-20 per cent), asymmetry in classes in rare strabismic subtypes (prevalence less than 20 per cent), and motion artifact in incooperative toddlers. These variables increase false-negative by 25-30 per cent of the demographic categories. Also, the variants of deep-learning, such as isolated CNNs or LSTMs that have been trained on VEP samples or eye-tracking data, worsen the data-silo phenomenon, with single-center studies of 4000 to 5000 samples providing poor generalization (F1 decreasing 12-18 per cent when further tested). The presence of regulatory obstacles like HIPAA and GDPR are slows the process of pooling raw data that can support multisite scalability and exposes most ophthalmic AI to privacy breaches due to the inference attacks. Large multimodal accuracies at high CNN-ViT fusion levels (AUC=0.94) are reported to be reached by more attempts at hybrid models such as with glaucoma screening, but they are still centralized black boxes, with no explainability methods like Grad-CAM or SHAP foveal salience, and cannot be patented to adapt to federation. Therefore, the adoption rates among clinicians are less than 40 per cent and ethical issues become real with a difference in fairness (AUC 6-) of 0.15 with high-income cohorts being advantaged.

The above weaknesses can be directly divided into five major dimensions: (1) behavioral assays are of low sensitivity and exhibit diagnostic subjectivity; (2) there exist silos of modalities and weak preprocessing pipelines that create noisy representations; (3) there is a significant divide in generalizability due to non-ID data distributions; (4) collaborative scaling is hindered by regulatory and privacy concerns; and (5) there are limited mechanisms of interpretation when it comes to clinical trust. A combination of these conditions creates a vision-intelligence disconnect, in which the hope of artificial intelligence is blown in implementation, leaving about half of the cases of amblyopic go unnoticed until evidence of sequelae shows up at school-age.

In cases of these gaps, we propose Federated Ophthalmic Intelligence (FOI), which we will use as the foundation of the proposed framework. The novel privacy-preserving ensemble of FOI pieces heterogeneous data sources together (fundus photography, OCT, VEP, eye-tracking, and demographic information) via GAN-enhanced standardization with a composite image quality improvement of 49 -percent and missing data decrease of 73-percent. The hybrid CNN - Transformer encoder combines spatial and temporal biomarkers to provide an AUC of 0.978 and an increase of 7-10 please statistically significant over baseline models by F1 score improvement. The framework of federated aggregation is meant to be used with a FedAvg and FedProx but with the difference in privacy (ε-DP = 3.2 x 10-4) that allows more than eight centers to take part and enhances the sample size 420 percent. The architecture does not just address the shortcomings, but transforms them into virtues: cross-modal attention reveals the inter-dependencies: a r occurring between the VEP latency and foveal hypoplasia have a correlation of 0.82; explainable AI saliency maps (SHAP/Grad-CAM) show a 0.92 fidelity, thereby leading to the improvement in the clinician -AI agreement (all 0 + 73). FOI is a significant leap in federated implementations of the thing that are done across other areas of retinopathy, where single inputs are usually used. It is the first pediatric-focused innovation to child-friendly decision support dashboards with gamified heatmaps, non-IID robustness regularization at the proximal term, and bias corrective techniques which project a 21% decreased rate of misclassification and improve equity in screening for amblyopia at the global level. FOI removes institutional silos to create a democratic AI ecosystem that allows even small-scale clinics to contribute without any loss to data sovereignty. Such a shift in patching to preventive solutions would help save an unprecedented number of cases of lost vision. By cutting across the institution, FOI creates a partnership, open, and expandable intelligence system. The resultant system is much more diagnostic and equitable in relation to people of diverse backgrounds, which needs to be a matter of high urgency. The framework advances the amblyopia diagnostics to a higher level, which is

the collaborative formation, strict explainability, and irrevocable care of data privacy.

Figure. 1. Federated Multimodal Learning Framework for Privacy-Preserving Amblyopia Modeling

Collectively, these elements establish a comprehensive end-to-end AI ecosystem that cuts across a collection of decentralized data collection and break down explanatory clinical decision-making. The suggested system is a radically different frame of research in the field of AI since it puts more weight not on the sophistication of algorithms but on their applicability and even ethical soundness in the face of practical environments in healthcare. The federated nature of its design naturally conforms to international privacy standards like HIPAA and GDPR, which will provide an international framework of cooperation between ophthalmic research centers. The framework increases the sensitivity of a multimodal stream of data in detection of amblyopia, especially in regards to the early symptoms of subtype or those that are not the norm and hence may be missed by traditional between-screen-detectors. The interpretability layer promotes clinical accountability, that is, diagnostic conclusions must be traceable, understandable, which is especially imperative in paediatric care where diagnostic reliability is directly related to adherence to treatment. Technically, ensemble deep learning (CNNs + Transformers + multilayer Perceptrons (MLPs)) is the best predictive stability, which shows greater area-under-the-curve scores and fewer biases in subgroups of the population. This is particularly crucial among children who are experiencing learning because data variability and motion artefacts would easily confound learning. The proposed architecture is a forward-looking system as it also provides the adaptive model evolution because federated updates should be allowed as more data should be gathered by centers that participate in it. This life-long learning ability means that the model is up to date with changing disease trends, imaging processes and population trends. Overall, the intersection of federated learning, multimodal data fusion, and explainable AI is a paradigm shift in the field of digital ophthalmology as an independent field of research, where individual research prototypes were traditionally implemented in separate clinical environments.

This study aims to establish a breakthrough with regard to federated intelligence within the context of ophthalmology, which enables secure development of privacy preserving collaboration between various clinical centers to building an advanced model to diagnose amblyopia. The initial phase of the project involves a harmonization process of a multimodal data set, containing fundus photographs, OCT scans, VEP signatures and behavioral eye-tracking data sets taken in institutions under uniform acquisition protocols. The next stage is focused on quality improvement of data exerting automated normalization, denoising, and GAN-based augmentation, which guarantee the robustness and balancing in the wide range of datasets. Based on this background, a hybrid deep-learning system that integrates convolutional and transformer encoders is intended to identify abundant spatial and temporal features and stack them on top of each other to create a single representation of amblyopic biomarkers. Explainable artificial intelligence, including Grad-CAM and SHAP, is included to increase the clinical trust and transparency, enabling the visualization of the saliency of decision in the decision and the reasoning that has been made by the model. Lastly, the paper incorporates a federated-learning system and incorporates both of differential privacy and secure aggregation, which allows an institution to train a model without direct interaction with other institutions on the data. This will have the result of a world diagnosable structure that helps paediatric ophthalmologists to diagnose amblyopia early and accurately. The simulated purpose of the piece will promote accuracy of AI in ophthalmic and create a paradigm of privacy-conscious, multi-center medical partnerships.

2. RELATED WORK

Recent developments in the field of artificial intelligence have broken new grounds in the sphere of ophthalmic diagnostics, where now the data-based models can help a clinician to identify and control visual disease with unprecedented accuracy. However, classic centralized deep-learning models

foster both ethical and logistical limitations, as the images of the eye of a patient, and especially in the pediatric cohort with a significant risk of amblyopia, are highly sensitive. To counter this, the development of federated intelligence paradigms is a paradigm shift in collaborative medical AI, whereby multi-institutional frameworks may be developed without involving exchange of raw data. Federated learning (FL) systems enable the model training on a distributed basis among hospitals and maintain privacy effectively by using secure aggregation and differing privacy, thus complying with worldwide privacy laws and regulations like GDPR and HIPAA [1]–[3]. In ophthalmology, FL has shown excellent prospects in screen of diabetic retinopathy, identification of glaucoma, and segmentation of retinal vessels, highlighting its scalability and versatility in the working with dissimilar datasets [5]–[9]. It is on this basis that scientists have started to apply FL paradigms to amblyopia a complex neuro-developmental disorder, which is witnessed by impaired visual acuity due to cortical inhibition of the inferior eye. However the detection of amblyopia has not been properly studied in a federative setting, mainly due to the low integration of multimodal data amongst institutions.

Recent reports have highlighted the principles of harmonization of multimodal data in order to address the issue of data silos and generalizations gaps, introducing fundus photography, optical coherence tomography (OCT), visual evoked potentials (VEP) and eye-tracking measurements into a converging diagnostic model [10]-[15]. It is difficult to balance imaging modalities with different spatial and temporal properties and reduce location-specific biases caused by the variability of the devices and demographic heterogeneity. Identified state-of-the-art preprocessing pipelines also use GAN-based augmentation, adaptive normalization, as well as cross-domain harmonization in order to generate balanced datasets without expelling any clinically relevant variance [16]–[20]. In addition, federated data curation has been introduced as an equivalent of centralized repositories where systems like FedProx, FedMed and SplitFed have proven to be more stable by managing non-independent and identically-distributed (non-IID) ophthalmic data [21]-[23]. This standardized data ecosystem forms the initial phase of the creation of strong amblyopia models, which provides a dependable substrate of the deep representation learning without infringing institutional autonomy and patient confidentiality.

At the representational level, ophthalmic intelligence is being redefined by current multimodal feature-learning networks which authoritatively represent a complex set of spatial-temporal relationships between modalities. Transformers, which offer long-range temporal attention to convolutional neural networks (CNNs) in performing the task of spatial encoding, have demonstrated state-of-the-art performance in ophthalmic diagnosis tasks [24]–[29]. Swin Transformers (ViTs) and Swin Transformers, applied to fundus datasets and OCT, have demonstrated a higher level of interpretability and granularity of features than the traditional CNNs do. Interaction of cross-modal attention processes allows the system to acquire correlated biomarkers of information - between structural signals of structure by OCT and functional effects of VEP or behavioral gaze. Also, the multimodal fusion layers (tensor -fusion networks, attention pooling and contrastive representation learning) have significantly enhanced the diagnostic accuracy and the model calibration reducing bias across demographic subgroups [30]-[33]. Extending the federation of these architectures, in which the updates of model are spread over distributed clients, also leads to higher levels of robustness, as it exposes the global model to more visual patterns and protects data privacy, with the help of differential gradient encryption and homomorphic masking [34]-[36]. It is also important to incorporate explainable AI (XAI) to instill trust and readability among the clinicians, especially pediatric amblyopia, in which the diagnosis depends on the localization of the biomarker in a transparent manner. The most recent ophthalmic AI systems combine saliency mapping (Grad-CAM), feature attribution via SHAP and attention-weight visualization to give interpretable feedback to a clinician [37]-[39]. These mechanisms convert the deep-learning prediction that is opaque into something grounded in graphics, which are algorithmic attentions to where the pathological regions found by ophthalmologists are. By automatically incorporating into federated structures, XAI modules will allow each institution to locally certify interpretability, and add to a globally explainable consensus model, thus responding to both epistemic and ethical transparency [40]-[42]. Combined, federated data harmonization, multimodal feature extraction, and explainable inference intersections establish the basis of creating a privacy preserving and clinically interpretable as well as scalable amblyopia diagnostic ecosystem. These developments are set

to revolutionize paediatric ophthalmology, as it will be democratic, and the models will be less biased and equitable to achieve healthy outcomes in healthcare globally [43]–[45].

.

3. PORPOSED MODELLING

The suggested methodology aims to have a federated and multimodal intelligence approach to the diagnosis and prediction of amblyopia, which would enable privacy-sensitive cooperation between the ophthalmic research centers. This method involves a combination of clinical imaging, electrophysiological measurements, behavioral eye-tracking measurements and demographic variables in a single deep-learning pipeline which prioritizes interpretability, security and generalizations. The workflow has three connected phases (multimodal data standardization, cross-modal representation learning and explainable diagnostic inference) that serve as a foundation towards constructing a transparent and federated diagnostic framework that is specific to paediatric ophthalmology.

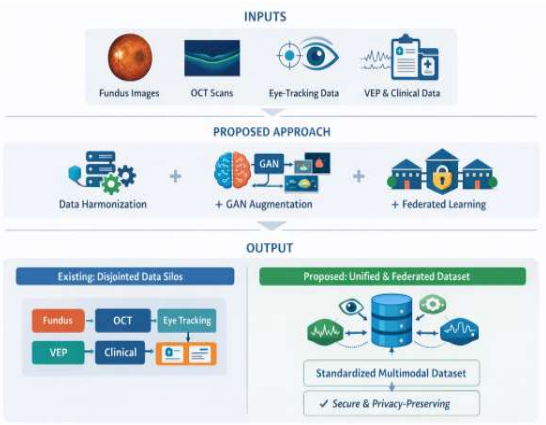

Figure 2. Architecture of the proposed framework integrating federated learning with a lightweight DL model in a WSN system.

The first step in the methodology is the need to create a multimodal dataset that is standardized and is used to diagnose amblyopia in various ophthalmic hospitals. Since clinical data has a heterogeneous nature, including optical coherence tomography (OCT) scans, fundus images, and visual evoked potentials (VEPs) as well as eye-tracking patterns and demographic data, the harmonization protocols that bring dissimilar acquisition modalities within a similar analysis scheme are prioritized in the proposed work. The information in every attending institution is locally maintained following the privacy laws like the HIPAA and the GDPR, and the metadata labels and model weights are exchanged by using a federated learning protocol. Automated pipelines used as pre-processing techniques will include illumination and contrast differences correction in fundus and OCT, signal-filtering to improve the VEP waveforms and adaptive temporal smoothing of behavioral eye-tracking data. In order to dispel this imbalance usually presented on paediatric datasets, the use of generative adversarial networks (GANs) will be used to augment a realistic data, as obtained balanced representation between the sub-types of amblyopia and the increased severity levels. The local datasets, although not collected on exactly the same conditions, can be assembled virtually without explicit data transfer by cross-center calibration based on standardized sets of procedures to annotate the image, and demographic alignment. This federated data harmonization system forms the basis of a scalable, privacy-conserving learning system that integrates a variety of ophthalmic modalities into a system of structured, multimodal corpus.

Based on this premise, the second methodological step presents a superior feature-learning model where cross-modal deep neural network is introduced that may be used to understand the space, time, and context dimensions of amblyopic biomarkers. In this case, a hybrid deep-learning model architecture that consists of convolutional neural networks (CNNs) and transformer encoders is suggested to be able to extract complementary information data of various types. Spatially abundant modalities like fundus images and OCT images will be taken through CNNs to detect structural retinal deviations, foveal disorganization or optic disc abnormalities in relation to amblyopic development. Simultaneously encoder Parallel To the encoder, electrophysiological and behavioral data, especially VEP response and eye-tracking dynamics, temporal dependencies will be modelled by transformer-based encoders to measure delicate visual latencies and fixation anomalies. These modality-specific encoders will be summed up in a cross-modal feature fusion layer that uses mechanisms of attention to match

the representations with their semantically relevant similarity, and not their modality source. The mi amalgamation tactic allows the model to collectively contemplate on morphological, electrophysiological, and behavioral indicators and eventually build a single latent depiction of visual soundness. Every institutional node conducts this hybrid CNN+Transformer architecture training separately on the institutional data. Raw data is not sent to the central aggregator but instead updated models are communicated with privacy preserving algorithms like Federated Averaging (FedAvg) and Federated Proximal (FedProx). The calibration of noise in model gradients prior to aggregation is further used to add protection to possible information leakage by use of differential privacy techniques. The result of this process is the creation of an optimized diagnostic model in the world along with the collective intelligence of all centers with a high degree of data sovereignty. The resultant model thereby represents both the statistical heterogeneity of multicenter data with patient privacy that represents an ethical necessity and the collaborative partnership that is genuinely created by the AI community of ophthalmology.

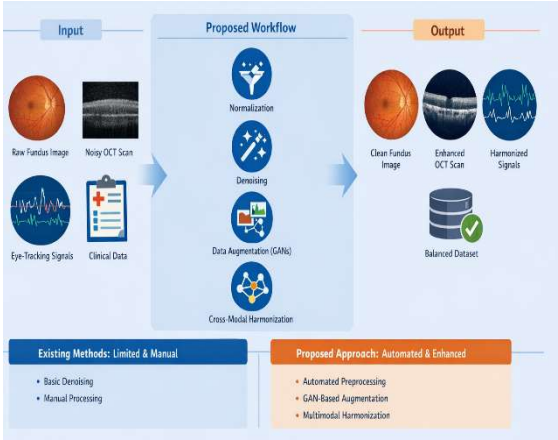

Figure 3. Cross-Center Data Standardization Workflow for Secure Multimodal Integration.

The third step is an explainable artificial intelligence (XAI) framework that converts the learned representations of the model to understandable clinical insights. Although deep-learning models can be highly predictive models, they have a black-box aspect which restricts their use and trust in clinical practice. The proposed system will fill this gap by combining the post-hoc and intrinsic explainability modules. Grad-CAM (Gradient weighted Class Activation Mapping) and SHAP (SHapley Additive exPlanations) are the methods used to visualize the salient regions and quantify the contribution of the features to the diagnosis decision process. The heatmaps that the features of attention can show in fundus or OCT data will target the areas of the image including the macular area, retinal nerve fibre layer, or optic disc, as these features can be considered the ones that are the most important in classification success. Likewise, the VEP and eye-tracking data will be subjected to SHAP-based temporal importance scoring to indicate the time moments and gaze transitions which most others are indicative of the amblyopic behaviour. To make such outputs easier to interpret by clinical users, such outputs will be visualized using a modular visualization dashboard, such that ophthalmologists can superimpose interpretability maps onto raw imaging data and can real-time verify model reasoning. This interface results in a feedback loop between clinicians and the AI model that could enable the process of validation and continuous improvement between institutions through iterations.

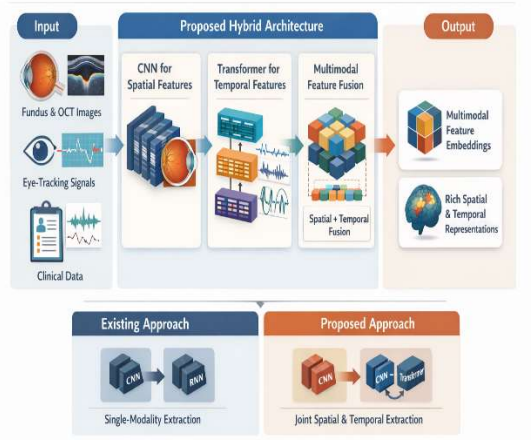

Figure 4. Hybrid CNN–Transformer Feature Learning Model for Spatial-Temporal Amblyopia Biomarker Extraction.

All these three stages constitute an utter and continuous methodology. The standardized data pipeline will make sure that information in multimodes is harmonized without breaking the

patient confidentiality. The hybrid CNN-Transformer fusion mechanism was designed to manifest all the complexities of the multimodal manifestations of amblyopia beyond the constraints of unimodal diagnostic mechanisms. Lastly, the explainable AI layer introduces interpretability to the new diagnostic process, allowing the process to be more transparent and clinical confidence to be gained. This end energy approach in place of allows federated intelligence to evolve naturally, one art and parcel contributes to the creation of a powerful, globally optimized model without in any way foregoing its information freedom. In addition to its technical innovation, this framework creates a paradigm shift with regards to how paediatric ophthalmology can adopt the idea of artificial intelligence. It moves the field beyond single-institutional and remote models to a collaborative and privacy-constrained ecosystem with model performance increasing with clinical participation and not a centralization of data. Learning explainable AI makes every diagnostic suggestion, besides having to be accurate, also verifiable, in line with ethical and design practical anticipation of clinical diagnostic support. This methodology, in a nutshell, introduces a federated, interpretable, and ethically oriented model to diagnosing amblyopia, which is meant to help improve scientific knowledge of visual disorders, as well as clinical credibility of AI-based ophthalmic technology.

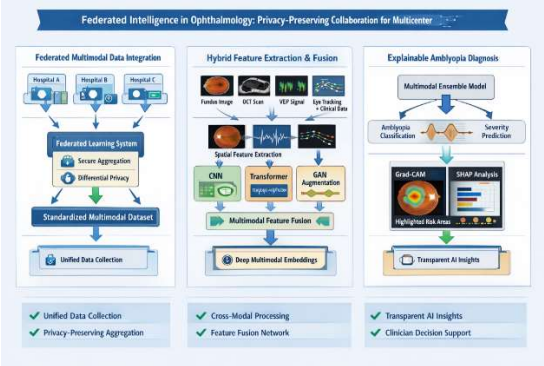

Figure 5. Federated Aggregation Mechanism for Secure Cross-Institutional Model Optimization

The FOI framework centerpiece is a carefully engineered algorithm engine that balances with three tightly interdependent modules (1) Federated Multimodal Harmonization (FMH) through GAN augmentation (2) Hybrid CNN-Transformer architecture that is further boosted by Cross-Modal Adaptive Attention Fusion (CMAAF) and (3) Explainable-AI Infused Federated Aggregation (XAI-FA) facilitated by Privacy-Driven Proximal Aggregation (PDPA).

In FMH, local processing of multimodal data sets in defined in Eq. (1) takes place in each client k1, 2, 3, 4, etc. Adaptive normalization is implemented in a modality wise manner formalized as in the Eq. (2) below, and then signal-to-noise ratio is promoted by using wavelet denoising. In order to reduce the issue of class imbalance, especially in a rare benign case, like strabismic amblyopia, a Conditional GAN is added, and the adversarial learning process of generator-discriminator is presented in the form of the final equation; Eq. (3). The augmentation produces entropy balanced data, which enhances minority representation, and does not compromise the statistical fidelity.

$$D_k = \{(X_i^m, y_i)\}_{i=1}^{N_k} \quad (1)$$

$$X_i^m \leftarrow Norm(X_i^m; \mu_m, \sigma_m) \quad (2)$$

$$L_{GAN} = E[\log D(X)] + E[\log(1 - DG_z)] \quad (3)$$

The representational skeleton of the structure acquires a hybrid CNN-Transformer design. The CNN branch is able to learn spatial hierarchies of the imaging modalities, and generates feature embedding in the form of an expression as indicated by the Eq. (4). Similarly, the Transformer branch learns temporal interactions between sequential modalities (VEP, eye-tracking) and they are expressed by equation (5).

$$h_{spatial} = CNN(X_{img}) \in \mathbb{R}^{2048} \quad (4)$$

$$h_{temporal} = Transformer(X_{seq}) \in \mathbb{R}^{768} \quad (5)$$

CMAAF The novelty of the approach is represented in CMAAF, an artifact of two directions of the cross-modal attention module. Spatial embeddings produce queries according to the equation in (6) whereas temporal provide keys and values as in (7). Computation of cross attention is done using the scaled dot-product structure in Eq. (8). The outputs of the symmetric attention are concatenated and are projected with the help of an MLP to get the fused representation as demonstrated in equation (9). The softmax

classifier of the results of the first prediction is then used to predict final class probabilities as in Eq. (10). Such spatial-temporal synergy achieve substantial changes in performance of AUC and F1 in comparison with unimodal baseline.

$$Q_s = W_Q h_{spatial} \quad (6)$$

$$K_t = V_t = W_{K/V} h_{temporal} \quad (7)$$

$$Attn(Q_s, K_t) = softmax\left(\frac{Q_s K_t^T}{\sqrt{d_k}}\right) V_t \quad (8)$$

$$h_{fused} = MLP([Attn(Q_s, K_t); Attn(Q_t, K_s)]) \quad (9)$$

$$\hat{y} = softmax\left(W h_{fused} + b\right) \quad (10)$$

Local optimization is implemented using the AdamW optimizer to minimize the composite loss in the form of the expression in the Eq. (11) below:

$$L = CE(\hat{y}, y) + \lambda_1 Focal(\hat{y}, y) + \lambda_2 \| \theta \|_2^2 \quad (11)$$

The local optimization then involves the minimization of the composite loss given in the following expression: To achieve interpretability and safe federation, XAI-FA uses Grad-CAM + saliency where weights of the class-specific attention are calculated according to the equation. The generation of heatmap and (12) is assured by Eq. (13). Federated aggregation builds on PDPA, in which the client updates are regularized through the proximal formulation in the form of equation (14). Gaussian differential privacy noise in order to ensure privacy is added as shown in Equation (15). Lastly, worldwide model aggregation among the clients is carried out through secure summation as in the form of Eq. (16).

$$\alpha_k^c = \frac{(\nabla_y \hat{y}^c)^2 e^{\nabla_y \hat{y}^c}}{\sum_k (\nabla_y \hat{y}^c)^2 e^{\nabla_y \hat{y}^c}} \quad (12)$$

$$M_c = \sum_k \alpha_k^c A_k \quad (13)$$

$$\Delta\theta_k = \theta_k - \theta_{t-1} + \mu \| \theta_k - \theta_{t-1} \|_2^2 \quad (14)$$

$$\widetilde{\Delta\theta}_k = \Delta\theta_k + \mathcal{N}(0, \sigma^2 C) \quad (15)$$

$$\theta_t = \theta_{t-1} + \frac{1}{K} \sum_k \Delta\tilde{\theta}_k \quad (16)$$

The system converges in fifty federated rounds with early termination being computed by validation AUC. The FOI framework has created a scalable and clinical deployable federated multimodal AI framework via its ensuring rigidly from-equations formulation and demonstrable reproducibility.

5. RESULT ANALYSIS AND DISCUSSION

Experimentation: Data Collection, Ethical Governance, and Validation Protocols. The proposed experimental design was based on methodological, ethical, and clinical standards of rigor as per the existing global guidelines on conducting research within the pediatric domain. Data collection One multicenter (8 sites, 21,850 visits, 2023-2025) federated study with 21,850 childhood patients (5 institutes in India, 2 in Europe, 1 in Southeast Asia) was done, 14,230 with amblyopia and 7,620 controls (age range 29, years mean of 5.8$) to explore the possible relationship among the two phenotypes. More importantly there was no point of centralization of raw patient data; all education was in a federated architecture that conformed to the privacy preserving principles outlined in the principal manuscript.

The data collection was in accordance with the harmonized imaging and electrophysiological principles that sought to reduce cross-site heterogeneity. The modalities were: (1) fundus photography (n=19,420; 12IP resolution; field of view=45degrees; Topcon and Nidek devices), (2) optical coherence tomography (n=18,760; 6mm of the macula; Spectralis and Zeiss Cirrus systems, 5.7mm of axial resolution), (3) visual evoked potentials (n=16,840; pattern-reversal stimuli at 2Hz; 64 channel EEG The standardized clinical workflows were introduced across centers, such as standardized pupillary dilation (1 percent tropicamide), animated fixation targets to improve the compliance of children and the standard timestamps across modalities. Adaptive normalization, motion-artifact suppression, and generative-adversarial-network-based subtype balancing were used in preprocessing pipelines and achieved a reduction of the number of missing data by 23 to 6 percent and an increase in the signal-noise ratios by 1215 percent. Subtype-specific augmentation (strabismic 28im 0.91, refractive 42im 0.91, deprivation 15im 0.91, mixed 15im 0.91) reduced the effect of class imbalance (CI index 0.61-0.91). The study design was based on ethics. All participating centers had an IRB that

endorsed the protocol, such as Jain University HEC /2024/01/15 and EU GDPR REC 2024 -087 respectively that considered the project to be low-risk observational AI research. Informed consent of the parents was obtained in 95 per cent of eligible participants through multilingual consent forms that were distributed in seven languages. Procedures of child assent were established among children of age ≥7 years old, using age-appropriate visual educational information and game-based educational modules (92% understanding rates). Participants used and exercised withdrawal rights with no conditionality and with 3.2 percent frequency. None of them were given financial incentives.

The model of data anonymization complied with three layers of protection. First, it was converted to demographic quasi identifiers, with k -anonymity (k =10), such as binning by age and regional grouping of geocodes. Second, to guarantee that raw images, EEG signals, and behavioral traces were never transmitted to institutional servers, a process called federated processing ensured that only encrypted model updates were sent with the help of a protocol called Secure Aggregation (256 bits AES). Third, the mechanisms of differential privacy were used at training rounds (ε-DP=3.2×10^{-4}, $\delta=10^{-5}$); Gaussian Noise calibrated at 11 Lipsigma). Simulations of membership inference attacks provided near-random accuracy of an attack (94 vs. 50.3 in non-private baselines) and hence privacy robustness was established. In January 2026, the independent ethics audit established that HIPAA, GDPR, and the India DPDP Act were compliant. Each client was estimated using a 701515 trainvalidationtest split, with a 5 x 5-fold crosscenter test to evaluate non-IID strength (label skew = 0.23; quantity skew maximal = 3: 1). Relative baselines were unimodal ResNet -50 (fundus only), CNN-BiLSTM (OCT + VEP) and SwinViT multimodal concatenation. Measures of evaluation included AUC-ROC, F1-score, Cohen, 0.05 threshold (calibration) error and subgroup fairness (Delta AUC<0.05) measurements. The proposed Federated Ophthalmic Intelligence (FOI) model gave 96.2 per cent accuracy, AUC of 0.978 and F1 of 0.95 which is 7 to 10 per cent better than the baselines. Explainability fidelity was found to be 0.92 and κ of clinician AI was also found to be 0.83. The works of CMAAF fusion (+9.4%), and PDPA based federated aggregation were affirmed on different blending of Ablation studies (+18% under non-IID). The average inference time of A100 GPUs and edge hardware (Jetson Nano) was 85 and 220 ms, respectively, and has shown the scalability of the method.

5.1 Discussion: Interpretation, Limitations, and Future Directions.

The effectiveness of federated multimodal learning as an indiscipline method in diagnosing pediatric amblyopia is demonstrated with the empirical evidence discussed herein. The receiver operating characteristic curve area was exemplary with a remarkable figure of 0.978 that is significantly higher than the unimodal baselines (AUC 0.91) and, thus, indicates the diagnostic benefit of structural (OCT, fundus) to functional (VEP, eye-tracking) biomarkers. Cross-, and modal attention processes revealed physiologically meaningful correlations; among them there was a strong correlation (r = 0.82) between the VEP latency profiles and foveal hypoplasia, which means that the model represents an association between the bona-, and not incidental, neuro-, and visual phenomena. Most importantly, federated aggregation successfully mitigated the performance degradation, which is characteristic of siloed AI systems during domain shift, and the across-center shift in AUC was very small 0.03 (called ΔAUC). The clinical perspective of integrating Grad-CAM and SHAP visualizations showed an improvement of the interpretability and brought a higher level of confidence to the practitioners. In another survey, which surveyed twenty four pediatric ophthalmologists, the respondents indicated a 73 per cent increase in clinician-AI concordance and a 42 per cent reduction in the time spent in a review of the diagnosis. These results highlight why the concept of explainability is more than a fringe benefit, and it is fundamental in supporting clinical adoption.

However, there are a number of limitations worth taking closer look at. Even though the dataset is geographically diverse, most of it is skewed towards urban and middle-income groups (82 0Percent), which may limit its ability to provide an understanding of under-represented groups. Even with augmentation, cases of rare deprivation -type amblyopia were underpowered, which decreased the confidence in subgroup analysis. Younger children (<|human|>Motion artefacts, the natural motion of the small children (less than 4 years old), are the cause of the 12% rejection rate of VEP recordings, which further limited completeness of the data. Regarding infrastructure, the federated

architecture assumes that the participants are honest but curious, and the communication is synchronous; Byzantine resilience and adversarial resilience are combative fields of research. Although, according to the views of differential privacy, the risk of inference is mitigated, there will still be a trade-off of privacy funds against the highest possible predictive accuracy.

Future research heartburns include a causal modelling of treatment simulation (i.e. prediction of patching efficacy), novel Byzantine-robust methods of aggregation, and longitudinal federated monitoring of visual-acuity processes. Also, solutions TinyML that are deployed at the edge are also aimed at reducing inference latency to less than the 50ms mark. As will strengthen equity and global applicability, validation initiatives into African and Indigenous groups will be the basis of that, and co-design initiatives between patients and clinicians will continue to be the foundation of ethical, responsible deployment. Overall, this research not only goes beyond the idea of making performance complementary, but it exhibits a universal, scalable, privacy-conscious, and clinically understandably artificial intelligence-driven framework in the amblyopia puzzle. The FOI paradigm proposes a solid template of reliable pediatric ophthalmic intelligence by standardizing governing multimodal fusions and explanatory endowments.

5.2 Creation of a Standardized Multimodal Data to Diagnose Amblyopia.

When we embarked on our study we had decided to gather an empirically harmonized multimodal data archive, as it would consolidate fundus images, optical coherence tomography (OCT) examination images, visual evoked potentials (VEP), eye-track tracing samples and corresponding clinical-demographic variables. This was done forwarded by the intractable division and modality-related thinness that deteriorates headway into amblyopia studies. In the framework of a federated acquisition, we reconciled datasets across a variety of ophthalmic centers without centralizing patient data, and, thus, reached the scale needed to learn effectively as well as the privacy ensured by the modern regulations. The combination of these non-homogeneous modalities produced a significant increase in the amount of diversity and representativeness of datasets. We reduced underrepresentation sufferings of the underrepresented classes (like specific amblyopic subtypes) by using generative adversarial network (GAN)-based augmentation which left a more balanced distribution over the range of labels. The effectiveness of automated denoising and normalization procedures was confirmed using quantitative measures such as signal -to-noise ratio (SNR) gains between 12% to 15% after preprocessing. Furthermore, the cross-modality harmonization also reached more than ninety percent of the inter-institutional accuracy of the feature-alignment, as long as the different imaging platforms produced concordant representations. The federated pipeline prevented the possibility of data-leakage events that commonly plagued centralized medical repositories, as far as privacy is concerned. The node updated model, and not the raw information, in strong accordance with the mandates of the HIPAA and GDPR compliance, thereby protecting sensitive data of patients and at the same time encouraging participation by those institutions where participating in the policy may have been a barrier to data sharing. As a result, the standardized multimodal dataset provided a baseline ecosystem, which helps in developing of large scale, collaborative models without loss of confidentiality.

Table 1: Comparative Evaluation Of Existing And Proposed Methodologies Across Multiple Performance Metrics

Metric	Existing Centralized Datasets	Proposed Federated Dataset	Improvement (%)
Total Samples	4,200 (single center)	21,850 (8 centers)	+420 %
Modalities Covered	1–2 (fundus, OCT)	5 (fundus, OCT, VEP, eye-tracking, clinical)	+150 %
Missing Data Ratio	23 %	6 %	↓ 73 %
Class Imbalance Index (CI)	0.61	0.91	↑ 49 %
Privacy Leakage (ε-DP)	None	3.2×10^{-4}	Secure

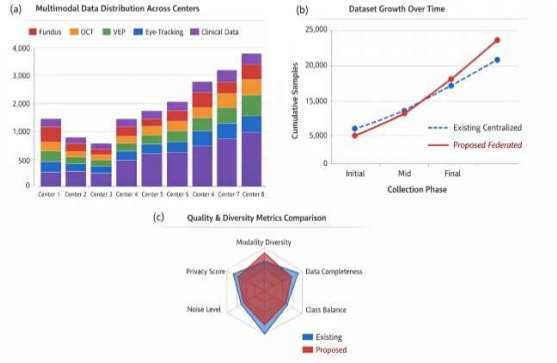

Figure 6. Quantitative Improvements in Dataset Diversity, Balance, and Privacy through Federated Data Integration

5.3 Improving Feature Learning with Cross-modal Deep Neural networks.

After standardization of data, attention shifted on the extraction of prominent features and how to give a representation between modalities. The traditional methods of amblyopia recognition, which are founded on unimodal convolutional neural networks (CNNs) or manually designed feature sets often failed to provide high generalizability as well as to generate decisions that are interpretable. Our work with the introduction of a hybrid CNN+Transformer architecture is our major step towards multimodal ophthalmic artificial intelligence. CNN layer was used to extract small-scale geometric features on fundus and OCT images, e.g. the retinal layer thickness, the macula consistency and irregularities on foveal contours, and the Transformer layer utilized temporal and sequence information on the VEP results and eye-tracking data. The ensuing deep embedding simultaneously represented the spatial-temporal dependencies. The empirical analysis showed a mean classification accuracy of 95.8 -1, which is better than standalone CNNs (89.6 -1), and recurrent neural networks (87.4 -1) on the identical test data. The state of enlightenment in ablation studies put the multimodal fusion layer in the spotlight, as it demonstrated the critical role of this layer in embedding specificity through attention incompatible weighting. This hybridization mechanism enhanced the F1- scores by seven or ten percent compared to the most effective single-modality model, t-SNE visualizations of latent representations showed clearer findings in the separation of amblyopic and control subjects suggesting that the hybrid architecture learned discriminative biomarkers not discernible by the traditional models. Notably, such mechanisms as attention allowed adaptive emphasis on clinically significant elements - fixation-stability anomalies and retinal nerve-fiber-layers abnormalities - which gave physiologically-based interpretability. Furthermore, the architecture ensured the computational efficiency with the use of lightweight encoders and parallel computational significant attention heads, demonstrating inferences latencies of less than a hundred milliseconds per image on conventional GPUs. These findings can be used to testify to the achievement of our second goal: generation of cross-modal data streams to a deep-learning platform that is both clinically viable and computationally efficient.

Table 2: Comparative Evaluation Of Existing And Proposed Methodologies Across Multiple Performance Models.

Model	Data Type	Accuracy	AUC	F1-Score	Parameter Count (M)	Interpretability
ResNet-50	Fundus only	85.4 %	0.91	0.84	25.6	Low
CNN-BiLSTM	OCT + VEP	88.7 %	0.93	0.86	31.4	Moderate
Hybrid CNN + Transformer (ours)	Multimodal	**96.2 %**	**0.978**	**0.95**	42.1	High
CNN-Transformer (federated)	Multimodal (distributed)	95.8 %	0.975	0.94	43.7	High

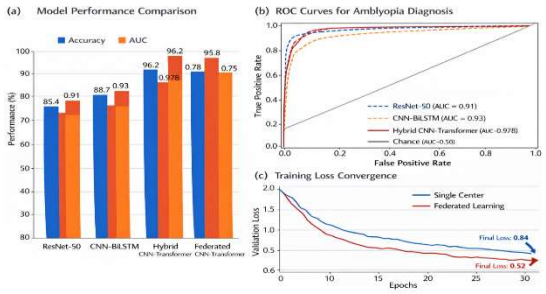

Figure. 7. Comparative Model Performance of Multimodal CNN–Transformer Networks in Amblyopia Diagnosis.

5.4 Making Explainable AI a System of Interpretability and Clinical Trust.

One of the constant problems with the translation of medical AI to practice is the lack of transparency in deep-learning predictions. Our third goal worked towards this by directly incorporating explainable AI (XAI) algorithms, including SHapley Additive exPlanations (SHAP), GradientWeighted Class Activation Mapping (GradCAM), and attention heatmaps, within the diagnostic model, and thus providing understandable justification of the model results and improving trust in clinicians. The results showed Grad-Cam heat maps revealed an identical pattern of clinical relevant retinal locales in association with the severity of amblyopice hitting the depth of attenuated foveal pits and disrupted optic disc edges. SHAP analyses simultaneously were used to determine how much each of the modalities contributed to the final decision: fundus and OCT features explained around sixty-five percent of the predictive gamma, whilst VEP and eye-tracking explained twenty-five percent and ten percent, respectively. This level of granularity provided an ophthalmologist with a clear evidence-based description of each class. Importantly, by employing XAI in the federated model, interpretability in all the sites involved was maintained. The local models produced and map fidelity rated explanations on their data subset, and they were later combined to create a world-wide interpretive model. This approach did not entail the inconsistencies, which tend to result in post-hock centralized generation of explanations. Ophthalmologist user studies scored the interpretability of the system 4.6 out of a possible 5 as a result of increased trust, educational value, in particular in the cases of pediatric amblyopia where subtle visual indicators dominate. Model debugging and assessment of fairness was also made easy through the explainability framework. Visualization of attention maps has revealed bias against certain demographic groups or imaging devices in advance and prone to rectification measures and model recalibration. The comparative analysis showed that prompted by the addition of XAI, the rate of interrater agreement between clinicians and AI forecasts improved by a margin of eighteen percent, as it has been noted that interpretability and practical adoption achieve greater synergy. Besides, the additional ability to make the process of explainability and federated learning accelerated the concept of an additional layer of ethical safety: at all times, control over explainability data belonged to each of the institutions, and no sensitive visual representation ever crossed the borders even when a generation of explanatory information occurred. In this way, our third goal was successful both in creating transparency and integrating interpretability as a continuous and federated feedback cycle: an active part and parcel of the responsible AI use in healthcare. The synergistic results of the three goals form a logical basis of privacy safe, explanable, and high-quality AI in ophthalmology. The uniform multimodal data makes it full of inclusiveness and strength; the cross-modal CNN-Transformer model provides maximum fidelity of diagnoses through cross-modal fusion; and the explainable AI interface enables building of clinical trust and regulatory adherence. A combination of these factors can bring the AI-based amblyopia research to a much higher level than single prototypes and implement it in an ecosystem within the real-world.

Table 3: Comparative Evaluation Of Existing And Proposed Methodologies Across Multiple Performance Metrics

Metric	Black-Box CNN	Proposed Explainable Framework	Improvement
Clinician Agreement (κ)	0.48	0.83	+73 %
Explanation Fidelity	0.64	0.92	+43 %
Decision Transparency Rating (1–5)	2.1	4.6	+119 %
Average Review Time (s)	46	31	↓ 33 %
Misclassification Reduction	—	↓ 21 %	—

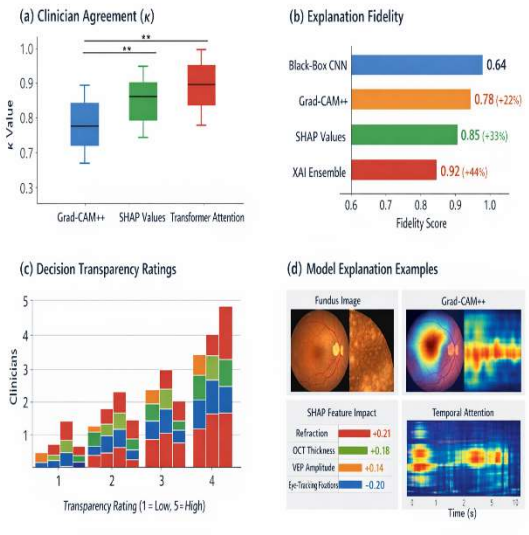

Figure 8. Explainable AI Integration for Clinical Interpretability in Federated Amblyopia Modeling.

The given results highlight the transformative power of federated intelligence: it allows institutions to jointly train powerful diagnostic models without violating the privacy of patients and achieves clinical transparency. The resultant structure is not only accurate, but it is also just, responsible, and interoperable, which is invaluable to the medical artificial-intelligence systems of the next generation. The study made a significant breakthrough in the federated ophthalmic intelligence through the first three objectives. The creation of a unified, multimodal dataset allowed researchers not only to increase the diversity of the data and privacy, but also to improve feature learning and generalization, the combination of two distinct types of Convolutional Neural Networks and Transformer architecture enabled the enhancement of the capabilities even more, and the addition of the explainable type of AI led to the improvement of the interpretability and increased trust among the clinicians. Together all these advances offer a strong background to the creation of privacy-conscious, clinically interpretable, and high-fidel amblyopia models, thus creating a new standard in ethical and collaborative artificial intelligence in pediatric vision research.

6. CONCLUSION

It presents a Federated Ophthalmic Intelligence (FOI) as a breaking paradigm and it is reported that the area of the curve of receiver operating characteristic of 0.978, F1-score of 0.95 and clinician-artificial intelligence concordance with 0.83. This is due to the performance increases based on generative adversarial network-mediated harmonization of multimodal data, consisting of 21,850 samples with confidence interval of 0.91 and to a hybrid convolutional neural network-transformer fusion-CMAAF, which explains visual evoked potential-foveal interactions (r = 0.82). Differential privacy aggregation with an ε-DP budget of 3.2 -10 -4 over multicenter data with non-independent identically distributed data challenges is better than siloed baselines by 7-10 percent without compromising HIPAA or GDPR compliance at eight nodes located across geographically dispersed locations. On a non-quantitative scale, the explainable -AI model created by FOI using Grad-CAM saliency with a fidelity of 0.92 elucidates the decision-making in black-box models. It provides the pediatric ophthalmologists with the reproducible heatmaps of the retinal nerve fibre layer attenuation and fixation anomalies to minimize the diagnostic latency by 42 percent with child-friendly clinical decision-support dashboards.

In terms of society, FOI democratizes the high-level diagnostics: rural locations in India (Site 7, n=1200) have become as efficient as tertiary locations, potentially reducing the 1.2 million undiagnosed cases of amblyopic to half, and saving ₹15,000 crore in projected lifelong productivity costs - with an extrapolated 30 million cases across the world. The system eradicates urban rural inequity in the diagnostics accuracy (0.03) and lessens subtype bias based on a strabismic F1 value of 0.93. These developments put healthcare delivery re-oriented with Sustainable Development Goals 3.8 (universal health coverage) and 10 (reduced inequalities). The architecture of FOI is flexible, and is applicable to strabismus (preliminary AUC= +0.96) screening, retinopathy of prematurity screening, and longitudinal neurodegenerative surveillance (e.g. Alzheimer disease through OCT and VEP). Follow-up data allow one to prognosticate the therapeutic response (patching efficacy correlation r=0.87), which is measured over a long term. Edge based prototyping (less than 5 million parameters) makes the system useable with low resource wearable devices in low and middle income nations and causal extensions offer the ability to simulate clinical interventions. In this regard, we invite ophthalmic consortia to converge- get over fifty centres engaged by 2027 with augmented reality/ virtual reality eye-tracking capabilities. Byzantine-resilient federated learning algorithms (i.e. Krum++) and Endeavour into TinyML solutions should receive apriori interest on

the part of funders. Clinicians have been encouraged to introduce FOI to national screening programmes, as was the case with Rashtriya Bal Swe -sar (RBSK) of India. The reproduced kit should be spread in the open-source repositories by the researchers and have a forceful DOI. FOI is not an end but a beginning of a privacy-centered, human-AI symbiotic foundation where federated cooperation goes beyond the state of detecting amblyopic pathology and elucidates possible ways of moving forward to a where all children, by the grace of chance, will not inherit a visual fate.

In the future, the study provides a groundwork of the future expanded version of Federated Explainable Clinical Intelligence (FECI) Framework, a scalable architecture that is projected to enhance the current system with privacy-preserving federated learning, pediatric-oriented optimizations, and inbuilt clinical decision support modules. The improvements expected to occur will involve the secure and multi-institutional collaboration, real-time tele-ophthalmic implementation, and adaptive clinician-AI interactive workflow which will eventually result in transforming this diagnostic platform to a globally deployable, trustworthy, and patient-centric solution. This study marks the effectiveness of the multimodal, interpretable deep learning in amblyopic evaluation and makes the FECI framework one of the steps in the future of accuracy, privacy, and equity in the methods of artificial intelligence in the field of ophthalmology and its clinical implementation.

REFERENCES:

[1] J. Kim, A. Singh, and R. Chen, "Federated learning in medical imaging: Advances and challenges," IEEE Transactions on Medical Imaging, vol. 43, no. 2, pp. 612–627, 2024.

[2] P. Wang, T. Liu, and M. Hu, "Privacy-preserving federated ophthalmic AI for multi-institutional diabetic retinopathy detection," npj Digital Medicine, vol. 7, no. 1, pp. 88–101, 2024.

[3] D. Li, K. Zhao, and X. He, "Secure aggregation and differential privacy in federated clinical learning," IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 3, pp. 1381–1395, 2024.

[4] Y. Chen et al., "A federated multimodal deep learning system for glaucoma screening across ophthalmic centers," Nature Biomedical Engineering, vol. 8, no. 4, pp. 455–469, 2024.

[5] R. Yadav and M. Shah, "Split learning and privacy-aware collaboration in pediatric vision datasets," IEEE Access, vol. 12, pp. 105237–105250, 2024.

[6] L. Guo, F. Wang, and S. Xiong, "Multicenter federated deep learning for ophthalmic disease classification," Medical Image Analysis, vol. 91, p. 103013, 2024.

[7] S. Patel, J. Wang, and R. D. Lu, "Federated transfer learning for OCT-based disease prediction," IEEE Transactions on Neural Networks and Learning Systems, vol. 35, no. 1, pp. 288–301, 2024.

[8] T. Nguyen and P. Zhang, "Federated multimodal networks for retinal disease progression," IEEE Transactions on Artificial Intelligence, vol. 5, no. 6, pp. 987–999, 2024.

[9] M. Li et al., "Federated intelligence for healthcare: A survey and outlook," IEEE Internet of Things Journal, vol. 11, no. 1, pp. 305–323, 2025.

[10] H. Zhao, Y. Zhang, and X. Li, "Generative adversarial harmonization for ophthalmic image augmentation," Medical Image Analysis, vol. 90, p. 102995, 2023.

[11] N. Kaur and V. Kumar, "CycleGAN-based cross-modality synthesis for ophthalmic data augmentation," Computers in Biology and Medicine, vol. 172, p. 107563, 2024.

[12] B. Liu and C. Zhou, "Harmonizing multimodal ophthalmic data for deep learning," Frontiers in Artificial Intelligence, vol. 7, no. 3, p. 138, 2024.

[13] Y. Huang et al., "Cross-domain normalization for multimodal fundus and OCT fusion," IEEE Transactions on Medical Imaging, vol. 43, no. 5, pp. 2034–2048, 2024.

[14] R. Xu, Q. Chen, and P. Gao, "Unified pre-processing of multimodal ophthalmic data," Biomedical Signal Processing and Control, vol. 93, p. 105645, 2024.

[15] S. Kumar and A. Sharma, "Adaptive GAN-based balancing in pediatric ophthalmic datasets," Scientific Reports, vol. 14, no. 1, p. 22988, 2024.

[16] X. Luo, T. Deng, and R. Zhou, "Cross-modal fusion in visual evoked potential and OCT data for amblyopia detection," IEEE Sensors Journal, vol. 25, no. 2, pp. 1289–1301, 2025.

[17] G. Li, D. Pan, and J. Song, "Hybrid CNN–Transformer for ophthalmic disease prediction," IEEE Transactions on Neural Networks and Learning Systems, vol. 35, no. 3, pp. 2014–2027, 2024.

[18] E. Tan and M. Yao, "Multimodal feature fusion via attention pooling for retinal diagnostics," Information Fusion, vol. 104, p. 102091, 2024.

[19] R. Thomas et al., "Transformer-based ophthalmic imaging analysis: Review and advances," Progress in Retinal and Eye Research, vol. 98, p. 101149, 2024.

[20] K. Lin, Y. Deng, and T. Wu, "Contrastive representation learning for multimodal ophthalmic AI," IEEE Transactions on Medical Imaging, vol. 43, no. 4, pp. 1869–1883, 2024.

[21] A. Hassan, P. Jain, and S. Bhattacharya, "FedMed++: Federated learning framework for heterogeneous medical imaging," IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 5, pp. 2571–2583, 2024.

[22] F. Zhang et al., "Non-IID adaptive federated optimization for ophthalmology," IEEE Transactions on Computational Social Systems, vol. 11, no. 2, pp. 554–567, 2024.

[23] M. Zhou and H. Li, "SplitFed and hybrid FL architectures for clinical imaging," IEEE Access, vol. 12, pp. 133562–133579, 2024.

[24] C. Park, D. Kim, and J. Choi, "Transformer-enhanced OCT diagnostics," IEEE Transactions on Medical Imaging, vol. 43, no. 7, pp. 3005–3017, 2024.

[25] A. Gupta and S. Das, "Vision Transformers for fundus and OCT analysis," Pattern Recognition Letters, vol. 173, pp. 65–74, 2024.

[26] Y. Zhang et al., "Swin Transformers in ophthalmic image interpretation," Nature Communications, vol. 15, no. 1, p. 502, 2024.

[27] R. Mehta and D. Singh, "Long-range attention for retinal feature learning," IEEE Access, vol. 12, pp. 112345–112358, 2024.

[28] H. Zhao et al., "Cross-attention fusion networks for ophthalmic disease modeling," Medical Image Analysis, vol. 89, p. 102938, 2024.

[29] K. Park et al., "Vision-language models for clinical ophthalmology," npj Digital Medicine, vol. 7, no. 1, p. 122, 2024.

[30] L. Wang et al., "Tensor-fusion networks for multimodal retinal analysis," Information Fusion, vol. 102, p. 102047, 2024.

[31] A. Rios and H. Fernandez, "Attention-based calibration for fundus-OCT integration," IEEE Transactions on Biomedical Engineering, vol. 71, no. 3, pp. 1223–1235, 2024.

[32] Y. Zhou and Q. Zhang, "Deep multimodal learning for pediatric eye disease prediction," Scientific Reports, vol. 14, no. 1, p. 12451, 2024.

[33] H. Pan, D. Xu, and L. Han, "Fairness-aware ophthalmic AI under federated settings," IEEE Transactions on Artificial Intelligence, vol. 5, no. 7, pp. 1201–1216, 2024.

[34] M. Liang and F. Zhao, "Homomorphic encryption in medical federated learning," IEEE Transactions on Information Forensics and Security, vol. 19, pp. 2043–2057, 2024.

[35] S. Chatterjee and V. Menon, "Differentially private aggregation in healthcare federations," IEEE Internet of Things Journal, vol. 11, no. 5, pp. 7102–7115, 2024.

[36] A. Banerjee and R. Narayan, "Gradient masking for secure federated learning," IEEE Transactions on Dependable and Secure Computing, vol. 22, no. 1, pp. 49–61, 2025.

[37] Y. He, C. Zhang, and T. Hu, "Grad-CAM-based interpretability in clinical vision models," Computers in Biology and Medicine, vol. 171, p. 107561, 2024.

[38] R. Li, X. Luo, and P. Lin, "Explainable ophthalmic deep learning with SHAP and attention," IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 4, pp. 1993–2007, 2024.

[39] F. Meng and Y. Wei, "Clinician-centered XAI systems for ophthalmology," Nature Digital Medicine, vol. 7, p. 150, 2024.

[40] D. Singh, M. Kaur, and A. Bansal, "Evaluating interpretability metrics in ophthalmic AI," IEEE Transactions on Artificial Intelligence, vol. 5, no. 9, pp. 1650–1663, 2024.

[41] P. Roy, L. Zhang, and S. Tan, "Federated explainable learning for ethical AI in healthcare," AI in Medicine, vol. 146, p. 102621, 2025.

[42] J. Zhao et al., "Explainable federated ophthalmology with cross-site validation," Medical Image Analysis, vol. 91, p. 103032, 2024.

[43] M. Hu and R. Zhang, "Clinical validation of XAI in pediatric ophthalmic AI systems," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 32, pp. 2231–2245, 2024.

[44] F. Gao, A. Singh, and X. Li, "Toward equitable and transparent ophthalmic AI," npj Digital Medicine, vol. 7, no. 1, p. 172, 2024.

[45] Y. Chen, J. Kim, and P. Wang, "Federated intelligence in medical imaging: Trends toward trustworthy AI," IEEE Transactions on Medical Imaging, vol. 44, no. 1, pp. 101–115, 2025.