Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Multimodal Deep Learning Method Based on Multiple Medical Images for Fuchs Endothelial Corneal Dystrophy Diagnosis

En-shuo Liu
Peking University Third Hospital
Li-li Cao
Institute of Automation
Jing-hao Qu
Peking University Third Hospital
Hao-ran Wu
Institute of Automation
Ge-ge Xiao
Peking University Third Hospital
Li-xue Shuai
Peking University Third Hospital
Shao-feng Gu
Peking University Third Hospital
Hai-kun Wang
Peking University Third Hospital
Ke-chao Zheng
Nanjing Artificial Intelligence Research of IA
Rong-mei Peng
Peking University Third Hospital
Jing Hong
hongjing196401@163.com
Peking University Third Hospital

Research Article

Keywords: Fuchs endothelial corneal dystrophy, multimodal, deep learning, artificial intelligence

Posted Date: January 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8473446/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Ophthalmology on September 9th, 2026. See the published version at https://doi.org/10.1186/s12886-026-05256-z.

A Multimodal Deep Learning Method Based on Multiple Medical Images for Fuchs Endothelial Corneal Dystrophy Diagnosis

En-shuo Liu,[1,2*] Li-li Cao,[3,4*] Jing-hao Qu, [1,2*] Hao-ran Wu,[3] Ge-ge Xiao,[1,2] Li-xue Shuai,[1,2] Shao-feng Gu,[1,2] Hai-kun Wang,[1,2] Ke-chao Zheng, [5] Rong-mei Peng, [1,2#,] Jing Hong,[1,2#]

[1]*Department of Ophthalmology, Peking University Third Hospital, Beijing, China;*

[2]*Beijing Key Laboratory of Restoration of Damaged Ocular Nerve, Peking University Third Hospital, Beijing, China.*

[3]*Institute of Automation, Chinese Academy of Sciences, Beijing, China.*

[4] *School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China.*

[5] *Nanjing Artificial Intelligence Research of IA, Nanjing, China.*

*These authors contributed equally to this work.

[#]Corresponding author

Rong-mei Peng (email: pengrongmei0217@aliyun.com); Department of Ophthalmology, Peking University Third Hospital, Beijing, China; No. 49 Garden North Road, Haidian, Beijing, 100191, China.

Jing Hong (email: hongjing196401@163.com); Department of Ophthalmology, Peking University Third Hospital, Beijing, China; No. 49 Garden North Road, Haidian, Beijing, 100191, China.

A Multimodal Deep Learning Method Based on Multiple Medical Images for Fuchs Endothelial Corneal Dystrophy Diagnosis

Abstract

Purpose: To establish a multimodal deep learning network for the fully automated diagnosis of Fuchs endothelial corneal dystrophy (FECD).

Methods: The ResNet-50 neural network was trained and validated using patients with FECD, patients with other anterior segment diseases, and healthy controls. Single-modal and multimodal models were developed on the basis of anterior segment photographs, anterior segment OCT images, and IVCM images. Independent test sets were employed to assess the diagnostic performance of the models, with evaluation metrics including precision, recall, F1 score, and accuracy.

Results: The multimodal model achieved a precision of 0.9663 and a recall of 0.971, with an F1 score of 0.9685, in distinguishing between FECD, other anterior segment diseases, and healthy eyes at the single-eye level. The diagnostic performance was significantly better than that of single-modal models based on anterior segment photographs (F1 score: 0.8664) and anterior segment OCT (F1 score: 0.8334) but slightly better than that of the single-modal model based on IVCM (F1 score: 0.9537).

Conclusion: This study presents the first multimodal deep learning model for diagnosing FECD, effectively distinguishing it from healthy corneas and various anterior segment disorders.

Translational Relevance: This multimodal AI integrates standard slit-lamp photo, OCT and IVCM inputs to deliver instant, technician-level FECD screening without extra hardware. Embedding the ResNet-50 classifier in EMR or cloud platforms could halve unnecessary corneal referrals and prioritise endothelial transplants before cataract surgery.

Keywords: Fuchs endothelial corneal dystrophy, multimodal, deep learning, artificial intelligence

INTRODUCTION

Fuchs endothelial corneal dystrophy (FECD) is a progressive, bilateral, and noninflammatory genetic disorder of the corneal endothelium that can be classified into early-onset and late-onset forms. In the early stages of the disease, patients typically exhibit no significant symptoms, with only drop-like extracellular matrix (ECM) excrescences, referred to as guttae, forming on the surface of the corneal endothelium. As the disease progresses, the guttae increase in number, fuse , and are accompanied by a reduction in corneal endothelial cell density, ultimately leading to corneal edema and decreased vision acuity [1]. Globally, nearly 300 million people are estimated to be affected by FECD, with an estimated prevalence of 7.33%. By 2050, the number of patients with FECD is predicted to further increase to over 400 million [2]. Despite the high prevalence of FECD, its insidious clinical manifestations contribute to a significant underdiagnosis rate in early-stage patients. Failure to diagnose FECD prior to intraocular surgeries, such as cataract procedures, significantly increases the risk of postoperative corneal endothelial decompensation. FECD represents the primary indication for corneal endothelial transplantation. A study conducted in 116 countries in 2012 demonstrated that FECD accounted for approximately 39% of all indications for corneal transplantation procedures [3]. These data highlight the substantial global healthcare burden imposed by FECD.

Slit-lamp biomicroscopy is the most commonly used method for the clinical diagnosis of FECD [4]. This device enables direct visualization of corneal guttae on the posterior corneal surface, and the Kachmer grading scale is used according to their quantity and degree of fusion [5]. In addition to slit-lamp biomicroscopy, multiple auxiliary imaging devices/methods can assist in the diagnosis of FECD, including anterior segment optical coherence tomography (AS-OCT), specular microscopy, and in vivo confocal microscopy (IVCM) [6]. AS-OCT enables noninvasive acquisition of high-resolution cross-sectional images across all corneal layers, including the epithelium, Bowman layer, stroma, Descemet membrane, and endothelium. In FECD, AS-OCT images show a thickened Descemet membrane and small nodular formations of endothelial cells; as the disease progresses, corneal stromal edema and epithelial bullous tissue can be observed by AS-OCT [7]. IVCM employs a 670-nm laser for

point illumination through an objective lens and conjugate pinhole, enabling optical sectioning of corneal structures. Compared with specular microscopy, IVCM provides superior corneal imaging clarity in patients with FECD, particularly in severe cases with corneal edema [8], confirming the greater clinical diagnostic value of IVCM. Furthermore, the critical value of IVCM lies in its ability to identify subtle pathological alterations undetectable by slit-lamp biomicroscopy, particularly in early-stage cases.

Artificial intelligence (AI) is a subfield of computer science that focuses on the development of computer systems capable of mimicking human intelligence to perform tasks, which include recognizing speech, identifying images and solving problems [9]. Deep learning (DL) is a vital branch of AI and is widely used in medical image recognition and classification. In 2020, a study demonstrated the utility of deep learning models in diagnosing FECD [10]. Another study introduced a VGG-19 neural network to differentiate patients with FECD from healthy subjects, patients with dry eye disease, and patients with keratoconus [7]. AS-OCT images were used to construct a deep learning model, which achieved superior diagnostic efficacy. Subsequently, the research team further optimized the model [11]. Another research team explored the use of deep learning models to differentiate FECD from abnormal corneal endothelial specular microscopy images caused by other etiologies, including pigment deposition, iatrogenic endothelial damage, and uveitis [12].

All the studies mentioned above have focused on the application of deep learning models in the diagnosis of FECD by only a single imaging method. Thus far, no studies have reported the application of multimodal deep learning models in the diagnosis of FECD. Furthermore, the aforementioned studies failed to incorporate corneal endothelial disorders with clinical manifestations similar to those of FECD, consequently compromising the extrapolation validity of these models in real-world applications. Our team previously developed fully automated deep learning systems for the segmentation and quantification of normal and abnormal corneal endothelial cell images obtained via IVCM [13, 14]. In the current study, we aimed to develop a multimodal deep learning diagnostic model for FECD, referred to as multimodal fusion framework with attention (MFFA), and compare its performance to that of its single-modal counterparts. Furthermore, by incorporating multiple similar corneal

endothelial diseases into our framework, we demonstrate the promising clinical application prospects of AI systems in the differential diagnosis of FECD.

MATERIALS AND METHODS

Study Population

From January 2022 to April 2025, subjects were consecutively recruited from the Ophthalmology Clinic of Peking University Third Hospital and its three affiliated hospitals (Peking University Third Hospital Airport Yard, Haidian Section of Peking University Third Hospital, and Peking University Third Hospital Northen Courtyard). FECD was diagnosed by the presence of corneal guttae in the central cornea (rather than the peripheral cornea) via slit-lamp examination, with or without corneal edema. Only subjects whose guttae were confirmed by IVCM or non-contact specular microscopy were included in our study. Patients with prior intraocular surgery except uncomplicated cataract procedures were excluded. Additional exclusion criteria included other anterior segment diseases, such as keratitis, ocular trauma, or uveitis. In addition, subjects with a history of systemic diseases that could affect the cornea or a history of topical medication use (with the exception of artificial tears) were excluded.

The control group comprised healthy individuals, defined as those who were not diagnosed with any ocular diseases upon examination or had only age-related cataracts, with no history of ocular trauma, topical or systemic medication use, or other systemic diseases that could affect the eyes.

Subjects with other corneal endothelial disorders, including posterior polymorphous corneal dystrophy (PPCD), congenital hereditary endothelial dystrophy (CHED), iridocorneal endothelial syndrome (ICE syndrome), corneal decompensation secondary to intraocular surgery, and corneal endothelitis, were also included in our study. The diagnosis of all subjects was determined by three experienced corneal specialists (Hong J., Peng R.M., and Xiao G.G.) on the basis of medical history, slit-lamp findings, and AS-OCT and IVCM examinations, and only those for whom all three reached a consensus were included in our study.

This study was approved by the Peking University Third Hospital Institutional Review Board and adhered to the tenets of the Declaration of Helsinki (IRB00006761-M2022834). Informed consent was obtained from the subjects.

Test Methods

Slit-lamp images were acquired using a standardized device model (BX-900, Haag-Streit, Switzerland) with diffuse white light, diffuse blue light combined with sodium fluorescein staining, and slit beam techniques; imaging was performed at multiple magnification settings (6.3×, 10×, 16×). Images meeting standardized inclusion criteria (complete corneal coverage with proper centration) were selected for subsequent analysis.

AS-OCT images were acquired using a commercially available swept-source OCT system (CASIA SS-1000; Tomey Corporation, Nagoya, Japan). All scans were performed with the "anterior segment" imaging mode built into commercial software. The subjects were instructed to fixate on the target, and scanning was initiated when the corneal reflex was in the center of the pupil. In each scan, sixteen images were obtained from each eye and stored for subsequent analysis.

IVCM images were acquired using a LASER IVCM system (HRT III Rostock Cornea Module [RCM]; Heidelberg Engineering GmbH, Heidelberg, Germany). Prior to examination, a drop of topical anesthetic (0.4% tetracaine hydrochloride) was administered into the conjunctival sac. Images of the central corneal endothelium were obtained by adjusting the fixation position of the subject's nonexamined eye. We selected and stored well-focused corneal endothelial images for subsequent analysis. Specifically, images selected for the healthy group were those with clearly identifiable corneal endothelial cell boundaries. For the FECD group and the disease group, images were chosen that exhibited characteristic manifestations of the respective conditions (e.g., guttae for FECD, epithelioid cells for iridocorneal endothelial [ICE] syndrome). We excluded endothelial images with artifacts or distortions, images where severe corneal edema precluded the identification of corneal endothelial cells or lesion details, images that did not capture the lesion area, and images belonging to other corneal layers rather than the corneal endothelium.

Datasets

All the subjects were stratified into three groups: The FECD group, the other corneal endothelial disorders group (see 'Study Population' for specific composition), and the healthy control group. Each included patient underwent 1 to 3 types of ophthalmic examinations. Samples with ≥2 modalities at the monocular level were classified as multimodal samples. We randomly selected 70% and 15% of the total subjects for the development and validation of the algorithm, and the residual 15% was reserved as an independent test set for the performance of the algorithm. Multiple diagnostically annotated images from an individual eye were integrated into a single training unit. The output of the model is one of three predefined clinical groups.

Problem Formulation

Precise diagnosis of anterior segment disorders necessitates the comprehensive analysis of heterogeneous information derived from disparate imaging modalities, including anterior segment photography (ASP), AS-OCT and IVCM. These multimodal data exhibit substantial variations in spatial resolution, imaging principles, and information content: Anterior segment photography presents macroscopic morphological characteristics, and AS-OCT reveals hierarchical structural information of tissues, whereas IVCM provides microscopic pathological features at the cellular level. Nevertheless, the effective fusion of such heterogeneous data and the handling of the absence of modality, which is prevalent in clinical scenarios, remain critical challenges in contemporary research.

To address these challenges, this study introduced a multimodal fusion framework with attention (MFFA) (Figure 1). This framework employs an end-to-end learning strategy to automatically acquire deep feature representations from multimodal images and dynamically evaluates the contribution of different modalities to specific diagnostic tasks through attention mechanisms. The overall methodological architecture encompasses four core components: Multimodal feature encoders, cross-modal attention adaptive fusion, missing modality handling strategies, and classification prediction with loss optimization.

We formalized the ophthalmic disease classification task as a multimodal learning problem.

Given a dataset $\mathcal{D} = \{(X_i, y_i)\}_{i=1}^{N}$ containing N patient samples, where each patient sample X_i comprises image data from M distinct modalities, $X_i = \left\{X_i^{(1)}, X_i^{(2)}, \ldots, X_i^{(M)}\right\}$.

For each modality m, the image collection of patient i is represented as follows:

$$X_i^{(m)} = \left\{x_{i,j}^{(m)}\right\}_{j=1}^{K_i^{(m)}}, x_{i,j}^{(m)} \in \mathbb{R}^{H \times W \times C}$$

where $K_i^{(m)}$ denotes the number of images for patient i under modality m.

In practical clinical environments, data from certain modalities may be absent because of factors such as equipment availability, examination protocols, or patient compliance. We employed a binary mask vector $m_i \in \{0, 1\}^M$ to represent the modality availability for patient i, where $m_i^{(k)} = 1$ indicates that modality k is available, and $m_i^{(k)} = 0$ denotes absence.

Our objective was to learn a robust mapping function $f: \mathcal{X} \times \mathcal{M} \to \mathcal{Y}$ that can accurately predict disease categories $\hat{y}_i$ given multimodal input X_i and modality mask m_i:

$$\hat{y}_i = f(X_i, m_i; \theta)$$

where θ represents the learnable parameters of the model.

Multimodal Feature Encoders

Given the heterogeneous characteristics of medical imaging modalities, where each modality generates images with distinct visual features and information content, we designed specialized feature encoders for each modality. Our approach was built upon the ImageNet pretrained ResNet-50[15] architecture, with specific improvements tailored for multimodal medical image analysis.

For each modality m, we designed a dedicated encoder $\mathcal{E}_m: \mathbb{R}^{H \times W \times C} \to \mathbb{R}^D$ that transformed the raw image data into unified feature representations. The encoder architecture removed the original adaptive average pooling layer and fully connected layer from ResNet-50, preserving all the convolutional layers from the first convolution to the fourth residual block. This yielded feature maps of dimension $\mathbb{R}^{H' \times W' \times 2048}$, where $H' = H/32$ and $W' = W/32$.

To obtain patient-level representations from the extracted feature maps, we applied adaptive global average pooling followed by feature transformation:

$$\mathrm{h}^{(\mathrm{m})} = \mathrm{AdaptiveAvgPool2d}\left(\mathrm{ResNet50}^{(\mathrm{m})}\left(\mathrm{x}^{(\mathrm{m})}\right)\right) \in \mathrm{R}^{2048}.$$

The pooled features were processed through a multilayer perceptron incorporating batch normalization, ReLU activation, and dropout regularization, yielding the final modality-specific feature representation:

$$\mathrm{f}^{(\mathrm{m})} = \mathrm{Dropout}\left(\mathrm{ReLU}\left(\mathrm{BatchNorm}\left(\mathrm{Linear}\left(\mathrm{h}^{(\mathrm{m})}\right)\right)\right)\right),$$

resulting in $\mathrm{f}^{(\mathrm{m})} = \mathrm{Dropout}\left(\mathrm{ReLU}\left(\mathrm{BatchNorm}\left(\mathrm{Linear}\left(\mathrm{h}^{(\mathrm{m})}\right)\right)\right)\right)$, where $D = 2048$ represents the unified feature dimension across all the modalities.

For patients with multiple images per modality, we independently extracted features from each image and subsequently aggregated them through simple averaging to obtain eye-level modality representations:

$$f_{i,j}^{(m)} = \mathcal{E}^{(m)}\left(x_{i,j}^{(m)}\right), F_i^{(m)} = \frac{1}{K_i^{(m)}} \sum_{j=1}^{K_i^{(m)}} f_{i,j}^{(m)}.$$

Specifically, each imaging modality (anterior segment photography, AS-OCT, and IVCM) was first processed independently by a ResNet-50 backbone to learn high-level semantic representations. These representations capture modality-specific pathological patterns and are mapped into a comparable latent feature space. Adaptive global average pooling was then applied as a simple and stable fusion strategy to integrate complementary information across modalities.

Cross-modal Attention Adaptive Fusion

To effectively integrate complementary information from multiple imaging modalities while addressing data absence scenarios, we designed a cross-modal attention fusion mechanism. This approach modeled intermodal relationships through a simplified self-attention architecture, dynamically weighing the contributions of available modalities.

Given the modality feature set $F_i = \left\{F_i^{(1)}, F_i^{(2)}, \ldots, F_i^{(M)}\right\}$ for patient i, we constructed the feature matrix $S_i = \left[F_i^{(1)}; F_i^{(2)}; \ldots; F_i^{(M)}\right] \in \mathbb{R}^{M \times D}$ as input to the attention mechanism.

The attention computation followed the standard query-key-value paradigm, utilizing learnable projection matrices $W^Q, W^K, W^V \in \mathbb{R}^{D \times D}$:

$$Q = S_i W^Q, K = S_i W^K, V = S_i W^V.$$

Attention weights were computed through scaled dot-product attention:

$$\mathrm{A} = \mathrm{softmax}\left(\frac{\mathrm{QK^T}}{\sqrt{\mathrm{D}}}\right).$$

Attention features were obtained through weighted combination and linear projection:

$$\mathrm{Z} = \mathrm{AV}, \mathrm{F_{attention}} = \mathrm{Linear(Z)}.$$

Finally, we aggregated cross-modal attention-weighted features to produce the following fused representation:

$$\mathrm{F_{fused}} = \sum_{m=1}^{M} w_m \cdot \mathrm{F}_{\mathrm{attention}}^{(m)},$$

where the weights w_m were dynamically adjusted according to the modality availability, ensuring robust fusion under data absence conditions.

Missing Modality Handling Strategy

Clinical deployment scenarios frequently encounter incomplete modality sets, necessitating robust handling of missing data. Our framework addressed this challenge through a comprehensive masking strategy that operates at both the feature and attention levels.

For missing modalities, we adopted a zero-padding strategy, replacing unavailable modality features with zero vectors:

$$\mathrm{F}_i^{(m)} = \begin{cases} \mathcal{E}_m(\mathrm{X}_i^{(m)}) & \text{if } m_i^{(m)} = 1 \\ 0 \in \mathbb{R}^D & \text{if } m_i^{(m)} = 0 \end{cases}$$

To prevent missing modalities from affecting attention computation, we constructed a two-dimensional attention mask $\mathrm{M_{attn}} = \mathrm{m_i m_i^T} \in \{0,1\}^{\mathrm{M \times M}}$ to mask attention scores

corresponding to missing modalities. The masked attention scores were computed as follows:

$$\mathrm{A}_{\text{masked}}[j,k] = \begin{cases} \mathrm{A}[j,k] & \text{if } \mathrm{M}_{\text{attn}}[j,k] = 1 \\ -10^{10} & \text{if } \mathrm{M}_{\text{attn}}[j,k] = 0 \end{cases}$$

Subsequent softmax normalization ensured that the attention weights were summed to one over the available modalities: $A_{norm} = \text{softmax}(A_{masked})$.

Classification Prediction and Loss Optimization

The fused multimodal representation was ultimately classified through a two-layer multilayer perceptron incorporating batch normalization and dropout regularization:

$$\mathrm{h}_1 = \text{DropoutReLU}\left(\text{BatchNorm}\left(\text{Linear}_1(\mathrm{F}_{\text{fsd}})\right)\right),$$

$$\mathrm{z} = \text{Linear}_2(\mathrm{h}_1),$$

where $z \in \mathbb{R}^L$ represents unnormalized class scores (logits) for L disease categories. The final class probabilities were obtained through softmax normalization:

$$p(y = c|X_i, m_i) = \frac{\exp(z_c)}{\sum_{j=1}^{L} \exp(z_j)}.$$

To optimize model training, we designed two primary loss functions to address class imbalance commonly encountered in medical imaging scenarios, such as disease distribution disparities in ophthalmic datasets. First, we employed standard cross-entropy loss to ensure global optimization:

$$\mathrm{L}_{\text{ce}} = -\frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \sum_{\mathrm{c}=1}^{\mathrm{L}} \mathrm{y}_{\mathrm{i,c}} \log\left(\widehat{\mathrm{y}_{\mathrm{i,c}}}\right).$$

Second, to address class imbalance and emphasize the learning of difficult samples, we introduce focal loss:

$$\mathrm{L}_{\text{focal}} = -\frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \alpha_{\mathrm{t_i}} \left(1 - \mathrm{p}_{\mathrm{t_i}}\right)^{\gamma} \log\left(\mathrm{p}_{\mathrm{t_i}}\right),$$

where α_{t_i} is the class balancing parameter, p_{t_i} is the predicted probability for the true class, and $\gamma = 2.0$ serves as the focusing parameter.

Baselines

To validate the advantages of multimodal fusion, we conducted direct comparisons between the MFFA model and multiple single-modal baseline models.

For single-modality models, we trained independent classification models for each modality (ASP-only, OCT-only, IVCM-only). These models maintained architectural consistency with the single-modal feature extractors within our multimodal framework, employing ResNet-50 as the backbone network and performing ensemble predictions across multiple images from the same patient. This approach ensured fair comparison, attributing any performance differences to modality combination and fusion strategies rather than architectural variations.

Evaluation Metrics

To comprehensively assess model performance, we employed standard classification evaluation metrics, including accuracy, precision, recall, and F1 score. Given that the F1 score provides a balanced reflection of both precision and recall, we adopted it as the primary metric for model comparison. All the metrics were computed on the respective independent test sets of each model.

Implementation Details

For anterior segment photographs (ASPs), original color images were resized to 224×224pixels and retained as three-channel (RGB) inputs. Anterior segment OCT images and IVCM images were grayscale by nature and were resized to 224×224pixels; these images were duplicated across three channels to match the input requirements of the ResNet-50 architecture pretrained on ImageNet.

All the models were implemented within the PyTorch framework and trained/evaluated using NVIDIA GeForce RTX 4090 GPUs. To ensure experimental reproducibility and fair comparison, we standardized critical implementation details and hyperparameter configurations.

Adaptive Optimization and Regularization: We employed the AdamW optimizer, which

decouples weight decay from gradient updates, providing superior regularization effects. We implemented cosine annealing learning rate scheduling with warm restarts, enabling periodic learning rate decay and reset to facilitate escaping local optima and exploring broader parameter spaces. Additionally, we extensively applied dropout techniques in the model's fully connected layers, randomly dropping neuron connections to prevent overfitting and ensure generalizability when complex multimodal learning tasks were performed.

Training Parameters: The initial learning rate was set to 1×10^{-4} with a weight decay coefficient of 1×10^{-2}. All the models received training for 50 epochs with a batch size of 8. For multimodal models employing focal loss, the focusing parameter γ was set to 2.0.

RESULTS

Participants & Datasets

Our study included 494 eyes from 312 subjects in the FECD group, 422 eyes from 406 subjects in the other corneal endothelial diseases group, and 450 eyes from 236 subjects in the healthy group (Table 1). Specifically, the other corneal endothelial disorders group comprised 63 subjects with PPCD, 12 subjects with CHED, 121 subjects with ICE syndrome, 139 subjects with corneal decompensation secondary to intraocular surgery, and 71 subjects with corneal endothelitis. The number of samples and images for each diagnosis in each modality is detailed in Table 2.

At the single-eye level, the number of included subjects was essentially balanced across the three groups. The mean age was similar between the FECD group and the healthy control group, while patients in the other corneal endothelial diseases group were the youngest. This is attributed to the fact that many corneal endothelial diseases have a peak incidence at an earlier age compared to FECD, which typically progresses with age. In terms of sex distribution, the male-to-female ratio in the FECD group was approximately 1:3, consistent with the higher prevalence of FECD in females. The sex distribution in the remaining two groups was largely balanced.

Diagnostic Efficacy of Single-Modal Models

First, we examined the diagnostic efficacy of three single-modal models. As previously mentioned, 15% of the overall sample was randomly selected to test the diagnostic performance of each model. In the test set, the distribution of test samples for each category is presented in the "support" column. The anterior segment photography-only (ASP-only) model achieved an F1 score of 0.7544 for differentiating FECD from other diagnoses, with a precision of 0.7234 and a recall of 0.8357 (Table 3). The weighted-average F1 score reached 0.8664, and the accuracy was 0.8539.

The anterior segment optical coherence tomography-based single-modal model (OCT-only) achieved an F1 score of 0.5567 for differentiating FECD from other diagnoses, with a precision of 0.4523 and a recall of 0.7243 (Table 4). The weighted-average F1 score reached 0.8334, and the accuracy was 0.8094.

The in vivo confocal microscopy-based single-modal model (IVCM-only) achieved an F1 score of 0.9189 for differentiating FECD from other diagnoses, with a precision of 1.00 and a recall of 0.85 (Table 5). The weighted-average F1 score reached 0.9537, and the accuracy was 0.9518.

Diagnostic Efficacy of the Multimodal Fusion Framework with Attention（MFFA）

We demonstrated the diagnostic efficacy of the multimodal fusion framework with attention (MFFA) in the test set. The MFFA achieved an F1 score of 0.9714 for differentiating FECD from other diagnoses, with a precision of 0.9444 and a recall of 1.00 (Table 6). The weighted-average F1 score reached 0.9685, and the accuracy was 0.971. Overall, compared with the ASP-only and OCT-only models, the MFFA demonstrated significantly superior diagnostic efficacy but marginally outperformed the IVCM-only model.

Attention Visualization

To gain deeper insight into model decision-making mechanisms, we employed gradient-weighted class activation mapping (Grad-CAM) techniques to visualize multimodal fusion framework with attention. By overlaying attention heatmaps onto original images, red regions indicate areas of high model attention, whereas blue regions represent areas of lower attention.

We present the heatmap results for two FECD subjects with varying disease severity. The selected anterior segment photography, AS-OCT, and IVCM images were all obtained from the same eye. The first subject, a 47-year-old female with mild FECD, showed a small number of central corneal guttae identified by MFFA in the anterior segment photograph (Fig. 2a, b). In the AS-OCT image, MFFA sensitively detected mild thickening of Descemet's membrane in the central cornea (Fig. 2c, d), which was consistent with the central-to-peripheral progression pattern of FECD. On the IVCM image, MFFA accurately localized the guttae within an otherwise normal endothelial background (Fig. 2e, f).

The second subject was a 70-year-old female with more severe FECD than the first, and her results are presented in Figure 3. In the anterior segment photograph, the directly illuminated corneal area and the posterior corneal guttae adjacent to the light band were identified by MFFA (Fig. 3a, b). On AS-OCT images, MFFA revealed thickening of Descemet's membrane in the central cornea and a wavy appearance of the posterior corneal surface due to guttae formation (Fig. 3c, d). Finally, in endothelial images nearly covered with guttae, MFFA primarily focused on the peripheral, near-annular areas of high-density guttae, while assigning less attention to the central region where guttae were relatively sparse (Fig. 3e, f). Our results demonstrated that MFFA can recognize the characteristic clinical features of FECD across various modalities.

DISCUSSION

In this study, we developed a fully automated multimodal deep learning model demonstrating high diagnostic accuracy for Fuchs' Endothelial Corneal Dystrophy (FECD. The trained model achieved an F1 score of 0.9685 and an accuracy of 0.971, indicating significant clinical applicability. Notably, we independently trained and evaluated three single-modal neural networks, among which the IVCM-based model demonstrated the best performance, which was comparable to that of the multimodal model. To our knowledge, this represents the first demonstration of a multimodal AI model for diagnosing corneal endothelial diseases.

Several early studies have reported deep learning systems for FECD diagnosis. However, they were restricted to single-modal architectures based on either AS-OCT or specular

microscopy. More importantly, their study failed to incorporate adequate disease training models with clinical manifestations resembling those of FECD. Several studies by Elsawy et al. included dry eye syndrome and keratoconus as control groups [7, 10, 11], whereas a study by Foo et al. included pigment deposition, iatrogenic endothelial damage, and uveitis as control groups [12]. In contrast to these prior studies, one of the main highlights of our study is the inclusion of multiple corneal endothelial disorders presenting clinical manifestations resembling FECD, including PPCD, CHED, ICE syndrome, surgically induced corneal decompensation, and corneal endothelitis [16]. For instance, IVCM images of some PPCD patients may demonstrate characteristic features, including nucleated endothelial cells, guttae-like hyporeflective lesions, and endothelial deposits, which have a high degree of similarity to FECD images [17]. Slit-lamp examination of ICE syndrome typically reveals an orange skin or hammered silver appearance caused by abnormal endothelial cells, which is highly similar to the appearance of guttae formed by abnormal extracellular matrix aggregates in FECD [18]. Visually similar images present challenges for deep learning models in classification tasks. By introducing these “real-world” pathological corneal images to train the model, we aimed to make the model more promising for clinical applications. Despite the inclusion of a variety of similar diseases, which significantly increases the difficulty of algorithm classification, our model demonstrated robust discriminative capacity in differentiating FECD from other corneal disorders and normal eyes.

An imaging modality alone often cannot provide all the information necessary to ensure an accurate clinical diagnosis. Consequently, clinicians are increasingly relying on images obtained from a variety of sources to formulate a diagnostic conclusion [19]. To capture multiperspective disease information, multimodal deep learning diagnostic models have emerged. In the field of ophthalmology, the application of multimodal modeling to glaucoma and fundus diseases has been widely reported. Xiong et al. developed a FusionNet framework integrating visual field and peripapillary circular OCT for diagnosing glaucomatous optic neuropathy (GON), achieving a diagnostic performance superior to that of visual field or OCT single-modal neural networks [20]. Li et al. constructed a multimodal deep learning model integrating electronic medical records, fundus photographs, OCT, ultrasound biomicroscopy

(UBM), and retinal nerve fiber layer (RNFL) thickness for glaucoma diagnosis and classification [21]. While substantial research has reported AI implementations in corneal diseases—including infectious keratitis [22, 23], keratoconus [24, 25], and pterygium [26]—almost all of them were based on a single modality: Anterior segment photography, AS-OCT, Pentacam, biomechanical data (Corvis ST), or IVCM.

Slit-lamp examination and photography provide a comprehensive morphological evaluation of the anterior segment. AS-OCT demonstrates cross-sectional visualization of corneal structures and detects abnormal alterations in the anterior chamber, iris, and lens, offering important contributions to differential diagnosis. IVCM specifically examines the microarchitectural features of the corneal endothelium, effectively identifying characteristic guttae lesions in FECD eyes. Owing to the complementary nature of these three diagnostic modalities, we developed a multimodal deep learning algorithm through systematic integration. Our findings demonstrate that the optimized multimodal model (MFFA) achieved superior performance (F1 score: 0.9685; accuracy: 0.971) compared with models using anterior segment photography or AS-OCT alone, with corresponding F1 scores of 0.8664 and 0.8334 and accuracy rates of 0.8539 and 0.8094, respectively. While the IVCM modality exhibited outstanding performance among the single-modal models (F1 score: 0.9537; accuracy: 0.9518), MFFA further enhanced overall performance through effective fusion of AS-OCT and anterior segment photography modalities. This not only demonstrates performance enhancement capabilities, particularly when complex cases are being handled but also highlights clinical robustness: The model continues to operate and provide reliable results even under modality absence conditions, thereby strengthening the generalization capability against data incompleteness. Furthermore, the potential of attention mechanisms in information fusion is evident, as they integrate rich information from dominant modalities (such as IVCM) with complementary details from auxiliary modalities while avoiding noise interference, thereby improving quantitative metrics and providing higher-value support for practical clinical decision-making.

IVCM is a contact-based, noninvasive diagnostic and imaging technique that enables high-magnification imaging of corneal and ocular surface structures. The Rostock Cornea

Module/Heidelberg Retina Tomograph III (RCM/HRT III), a widely employed IVCM imaging system utilized in our present study, achieves a tissue structural resolution of 1 micrometer [27]. Compared with specular microscopy, IVCM produces clearer corneal images in patients with FECD, especially when severe corneal edema is present [8]. This is attributed mainly to the distinct imaging principles between the two techniques. Specular microscopy projects light onto the cornea at a specific incident angle and captures the reflected light from the interface between the endothelium and aqueous humor [28]. In contrast, the IVCM employs a laser beam with a 670-nm wavelength for image acquisition. The characteristic lesion of FECD, guttae, manifests as nonreflective dark areas on specular microscopy, which may be similar to endothelial cell abnormalities of multiple other etiologies. IVCM enables clear visualization of guttae morphology and density, facilitating differentiation between FECD, other corneal endothelial disorders, and normal endothelium. Thus, among the three single-modality networks, the model trained on IVCM images demonstrated superior diagnostic performance.

CONCLUSION

This is the first study that presents a multimodal deep learning diagnostic model for FECD capable of distinguishing FECD from healthy corneas and various corneal endothelial disorders, including PPCD, CHED, ICE syndrome, corneal decompensation secondary to intraocular surgery, and corneal endothelitis. The multimodal model and the single-modal model trained using IVCM images demonstrated excellent accuracy. The algorithm we developed demonstrates potential for fully automated diagnosis of FECD, although we acknowledge that further work is needed to validate its generalizability.

LIMITATIONS

Despite the promising results, this study had limitations. First, although our dataset was derived from four different centers, all four hospitals are located in Beijing, which may have constrained the generalizability of the model. Future investigations should focus on incorporating multicenter datasets from diverse geographic regions and ethnic populations and employing external validation sets to verify generalizability across real-world clinical settings.

Second, to ensure the quality and consistency of the dataset, we exclusively utilized a single type of device (although multiple units were employed) to capture anterior segment photography, AS-OCT, and IVCM images. Prior to formal deployment, the algorithm's compatibility with different types of devices must be validated. Third, the main aim of our study was to construct a fully automated diagnostic model for FECD that was capable of distinguishing between normal individuals and those with other corneal endothelial disorders. Therefore, we did not grade the severity of FECD (e.g., distinguishing FECD with/without clinical edema or employing the Kachmer grading scale) in our analysis.

CONTRIBUTORS

Study design: GG Xiao, J Hong and RM Peng; study execution: ES Liu, JH Qu and LL Cao; collection and management of the data: ES Liu, JH Qu and LX Shuai; image capture: SF Gu and HK Wang; analysis and interpretation of the data: LL Cao; writing of the manuscript: ES Liu, JH Qu and LL Cao; and review or approval of the manuscript: J Hong and RM Peng.

FUNDING

This work was supported by the Natural Science Foundation of Beijing–Changping Innovation Joint Fund (Key Program) under Grant Nos. L254041, National Natural Science Foundation of China under Grant Nos. 82371027 and the Natural Science Foundation of Beijing under Grant Nos. Z20240017.

COMPETING INTERESTS

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

AVAILABILITY OF DATA AND MATERIAL

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ACKNOWLEDGMENTS

We are very grateful to all the patients for their cooperation in the study.

REFERENCES

1. Weiss JS, Rapuano CJ, Seitz B, Busin M, Kivela TT, Bouheraoua N, et al. IC3D Classification of Corneal Dystrophies-Edition 3. Cornea. 2024;43(4):466-527.
2. Aiello F, Gallo Afflitto G, Ceccarelli F, Cesareo M, Nucci C. Global Prevalence of Fuchs Endothelial Corneal Dystrophy (FECD) in Adult Population: A Systematic Review and Meta-Analysis. J Ophthalmol. 2022;2022:3091695.
3. Gain P, Jullienne R, He Z, Aldossary M, Acquart S, Cognasse F, et al. Global Survey of Corneal Transplantation and Eye Banking. JAMA Ophthalmol. 2016;134(2):167-73.
4. Patel SV. Imaging Fuchs Endothelial Corneal Dystrophy in Clinical Practice and Clinical Trials. Cornea. 2021;40(12):1505-11.
5. Adamis AP, Filatov V, Tripathi BJ, Tripathi RC. Fuchs' endothelial dystrophy of the cornea. Surv Ophthalmol. 1993;38(2):149-68.
6. Oie Y, Yamaguchi T, Nishida N, Okumura N, Maeno S, Kawasaki R, et al. Systematic Review of the Diagnostic Criteria and Severity Classification for Fuchs Endothelial Corneal Dystrophy. Cornea. 2023;42(12):1590-600.
7. Elsawy A, Eleiwa T, Chase C, Ozcan E, Tolba M, Feuer W, et al. Multidisease Deep Learning Neural Network for the Diagnosis of Corneal Diseases. Am J Ophthalmol. 2021;226:252-61.
8. Huang J, Maram J, Tepelus TC, Sadda SR, Chopra V, Lee OL. Comparison of Noncontact Specular and Confocal Microscopy for Evaluation of Corneal Endothelium. Eye Contact Lens. 2018;44 Suppl 1:S144-S50.
9. Hogarty DT, Mackey DA, Hewitt AW. Current state and future prospects of artificial intelligence in ophthalmology: a review. Clin Exp Ophthalmol. 2019;47(1):128-39.
10. Eleiwa T, Elsawy A, Ozcan E, Abou Shousha M. Automated diagnosis and staging of Fuchs' endothelial cell corneal dystrophy using deep learning. Eye Vis (Lond). 2020;7:44.
11. Elsawy A, Abdel-Mottaleb M. A Novel Network With Parallel Resolution Encoders for the Diagnosis of Corneal Diseases. IEEE Trans Biomed Eng. 2021;68(12):3671-80.
12. Foo VHX, Lim GYS, Liu YC, Ong HS, Wong E, Chan S, et al. Deep learning for detection of Fuchs endothelial dystrophy from widefield specular microscopy imaging: a pilot study. Eye Vis (Lond). 2024;11(1):11.
13. Qu JH, Qin XR, Peng RM, Xiao GG, Cheng J, Gu SF, et al. A Fully Automated Segmentation and Morphometric Parameter Estimation System for Assessing Corneal Endothelial Cell Images. Am J Ophthalmol. 2022;239:142-53.
14. Qu J, Qin X, Peng R, Xiao G, Gu S, Wang H, et al. Assessing abnormal corneal endothelial cells from in vivo confocal microscopy images using a fully automated deep learning system. Eye Vis (Lond). 2023;10(1):20.
15. He KM, Zhang XY, Ren SQ, Sun J. Deep Residual Learning for Image Recognition. Proc Cvpr Ieee. 2016:770-8.
16. Matthaei M, Hribek A, Clahsen T, Bachmann B, Cursiefen C, Jun AS. Fuchs Endothelial Corneal Dystrophy: Clinical, Genetic, Pathophysiologic, and Therapeutic Aspects. Annu Rev Vis Sci. 2019;5:151-75.
17. Bozkurt B, Irkec M, Mocan MC. In vivo confocal microscopic findings in posterior polymorphous corneal dystrophy. Cornea. 2013;32(9):1237-42.
18. Silva L, Najafi A, Suwan Y, Teekhasaenee C, Ritch R. The iridocorneal endothelial syndrome. Surv

Ophthalmol. 2018;63(5):665-76.
19. Li Y, El Habib Daho M, Conze PH, Zeghlache R, Le Boite H, Tadayoni R, et al. A review of deep learning-based information fusion techniques for multimodal medical image classification. Comput Biol Med. 2024;177:108635.
20. Xiong J, Li F, Song D, Tang G, He J, Gao K, et al. Multimodal Machine Learning Using Visual Fields and Peripapillary Circular OCT Scans in Detection of Glaucomatous Optic Neuropathy. Ophthalmology. 2022;129(2):171-80.
21. Li Y, Han Y, Li Z, Zhong Y, Guo Z. A transfer learning-based multimodal neural network combining metadata and multiple medical images for glaucoma type diagnosis. Sci Rep. 2023;13(1):12076.
22. Wu J, Yuan Z, Fang Z, Huang Z, Xu Y, Xie W, et al. A knowledge-enhanced transform-based multimodal classifier for microbial keratitis identification. Sci Rep. 2023;13(1):9003.
23. Essalat M, Abolhosseini M, Le TH, Moshtaghion SM, Kanavi MR. Interpretable deep learning for diagnosis of fungal and acanthamoeba keratitis using in vivo confocal microscopy images. Sci Rep. 2023;13(1):8953.
24. Abdelmotaal H, Hazarbassanov RM, Salouti R, Nowroozzadeh MH, Taneri S, Al-Timemy AH, et al. Keratoconus Detection-based on Dynamic Corneal Deformation Videos Using Deep Learning. Ophthalmol Sci. 2024;4(2):100380.
25. Kundu G, Shetty R, Khamar P, Mullick R, Gupta S, Nuijts R, et al. Universal architecture of corneal segmental tomography biomarkers for artificial intelligence-driven diagnosis of early keratoconus. Br J Ophthalmol. 2023;107(5):635-43.
26. Liu Y, Xu C, Wang S, Chen Y, Lin X, Guo S, et al. Accurate detection and grading of pterygium through smartphone by a fusion training model. Br J Ophthalmol. 2024;108(3):336-42.
27. Bohnke M, Masters BR. Confocal microscopy of the cornea. Prog Retin Eye Res. 1999;18(5):553-628.
28. McCarey BE, Edelhauser HF, Lynn MJ. Review of corneal endothelial specular microscopy for FDA clinical trials of refractive procedures, surgical devices, and new intraocular drugs and solutions. Cornea. 2008;27(1):1-16.

Figure 1. Multimodal Fusion Framework with Attention (MFFA) for Accurate Disease Diagnosis

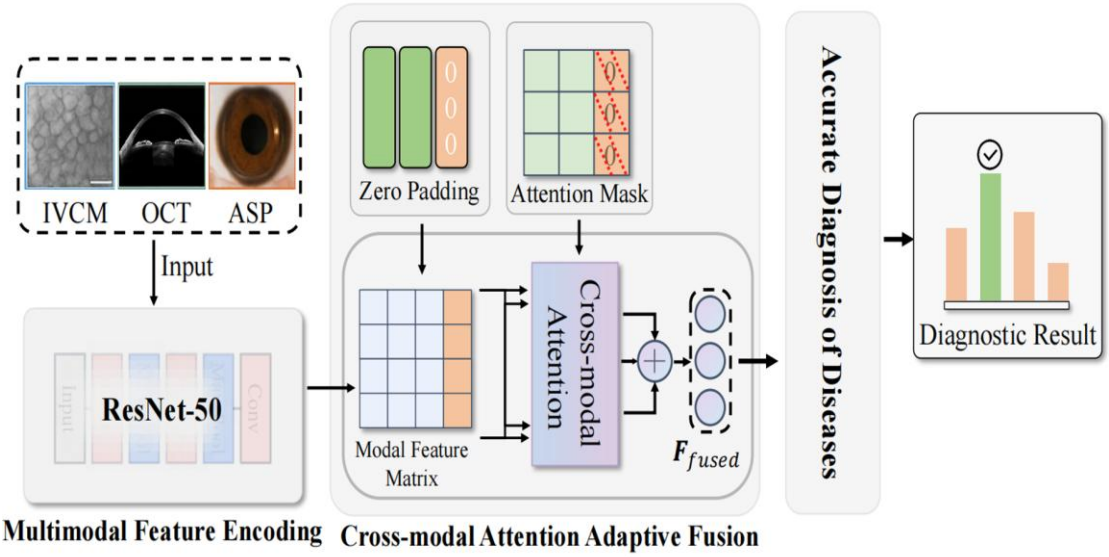


Figure 2. Grad-CAM analysis for the first FECD patient in the multimodal fusion model. (a, b): Input of ASP images and multimodal model attention. (c, d): Input of AS-OCT images and multimodal model attention. (e, f): Input of IVCM images and multimodal model attention.

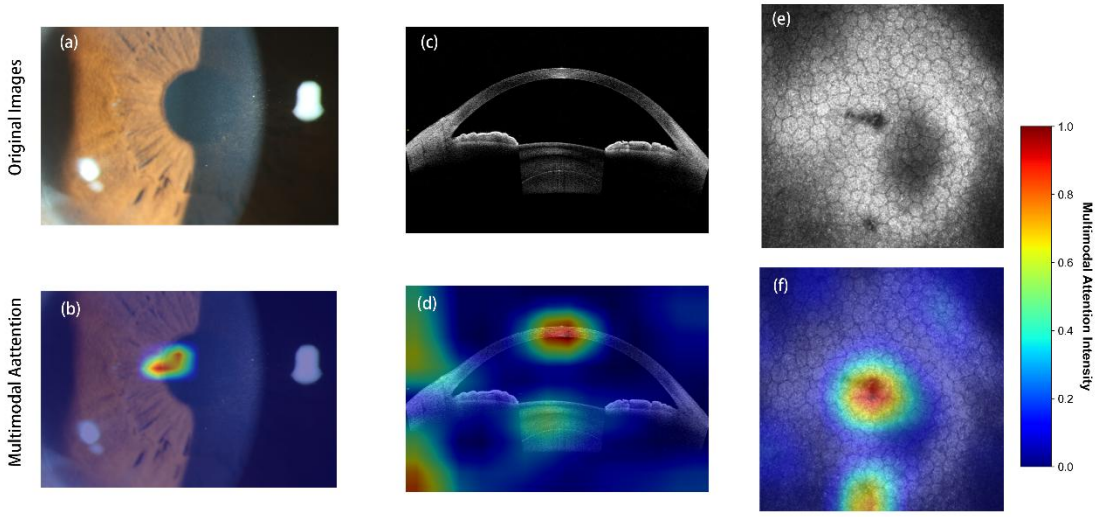


Figure 3. Grad-CAM analysis for the second FECD patient in the multimodal fusion model. (a, b): Input of ASP images and multimodal model attention. (c, d): Input of AS-OCT images and multimodal model attention. (e, f): Input of IVCM images and multimodal model attention.

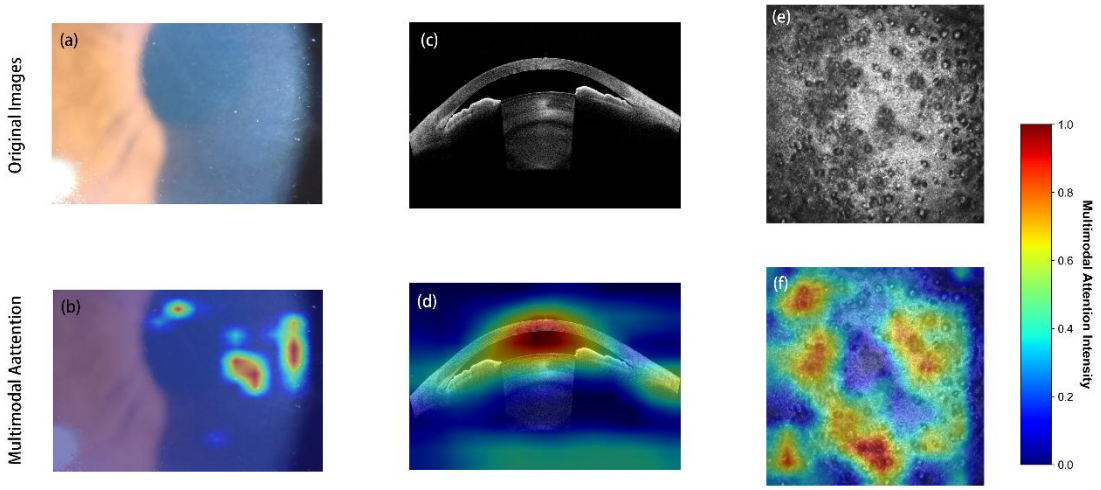


Table 1. Characteristics of study groups.

	FECD	Other	Normal
Number of subjects	312	406	236
Age, mean (range)	61.5 (31-86)	53.5 (3-88)	65.3 (25-87)
Female, number (%)	235 (75.3%)	212 (52.2%)	141 (59.7%)
Number of eyes	494	422	450
Eye, OD (%)	263 (53.2%)	210 (49.8%)	229 (50.9%)

Table 2. Dataset distribution across modalities and diagnostic categories.

Modality	Total samples (images)	FECD	Other	Normal
ASP	1311 (6078)	484 (2305)	389 (2159)	438 (1614)
OCT	740 (13509)	240 (4176)	281 (5829)	219 (3504)
IVCM	980 (22375)	265 (10099)	274 (8620)	441(3656)
Multimodal	455 (39889)	128 (15653)	118 (15860)	209 (8376)

Table 3. Performance of the single-modal model (ASP-only) based on the test set

Class	Precision	Recall	F1-Score	Accuracy	Support
FECD	0.7234	0.8357	0.7784	—	50
Other	0.9487	0.8156	0.8812	—	50
Normal	0.9012	0.9012	0.9012	—	81
Weighted avg	0.8823	0.8539	0.8664	0.8539	181

Table 4. Performance of the single-modal model (OCT-only) based on the test set

Class	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Support**
FECD	0.4523	0.7143	0.5567	—	14
Other	0.9184	0.8167	0.8651	—	60
Normal	0.9643	0.8438	0.9	—	32
Weighted avg	0.8637	0.8094	0.8334	0.8094	106

Table 5. Performance of the single-modal model (IVCM-only) based on the test set

Class	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Support**
FECD	1	0.85	0.9189	—	40
Other	0.8547	0.9821	0.9139	—	25
Normal	0.9868	0.9868	0.9868	—	76
Weighted avg	0.9614	0.9518	0.9537	0.9518	141

635 Table 6. Performance of the multimodal model (MFFA) based on the test set.

636

Class	Precision	Recall	F1-Score	Accuracy	Support
FECD	0.9444	1	0.9714	—	16
Other	0.9762	0.9524	0.9641	—	20
Normal	0.9706	1	0.9851	—	33
Weighted avg	0.9663	0.971	0.9685	0.971	69

637