

Development of a Multimodal Large Language Model for Early Warning and Diagnosis of Dry Eye Disease

Mingyan He

Peking University Third Hospital

Lili cao

Chinese Academy of Sciences

Kechao Zheng

Nanjing Artificial Intelligence Research of IA Nanjing

Zhiping Lin

Peking University

Haoran Wu

Chinese Academy of Sciences

Shuzhen Wang

Zhangjiakou First Hospital

Rui Tian

The Second Hospital of Jilin University

Yu Wang

Fushun Eye Hospital

Tiehong Chen

Liaoning Aier Eye Hospital

Wei Wang

Zhangjiakou First Hospital

Liang Xiao

The Second Hospital of Jilin University

Tingting Yang

Liaoning Aier Eye Hospital

Jing Hong

hongjing1964010163.com

Peking University Third Hospital

Research Article

Keywords: Dry Eye Disease, Multimodal Large Language Model, Early Warning, Diagnosis

Posted Date: October 11th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11121960/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

This study aimed to develop and externally validate an instruction-tuned multimodal large language model (MLLM) for early warning and early diagnosis of dry eye disease (DED) using routinely obtainable clinical data.

Methods

A total of 874 participants were enrolled, including an internal development cohort ($n = 743$) from Peking University Third Hospital (October 2024 to August 2025) and an independent external validation cohort ($n = 131$) from three additional centers. Participants included both DED and non-DED individuals, diagnosed according to the Chinese Expert Consensus (2020) and TFOS DEWS II criteria. We developed DED-MLLM on the basis of Qwen2-VL and constructed a domain-specific instruction-tuning dataset integrating Ocular Surface Disease Index (OSDI), Standard Patient Evaluation of Eye Dryness (SPEED), self-reported questionnaires, ophthalmic imaging, and clinical examination data. Model performance was evaluated for risk stratification and diagnosis under both diagnostic frameworks, with ablation analyses used to examine the relative contributions of visual and structured clinical modalities.

Results

Based on the international diagnostic criteria for DED, the model achieved an accuracy of 0.8232 for early warning (specificity: 0.6873, sensitivity: 0.9558) and an accuracy of 0.9203 for diagnosis (specificity: 0.8776, sensitivity: 0.9720). According to the Chinese diagnostic criteria for DED, the model achieved an accuracy of 0.9584 for early warning (specificity: 0.9112, sensitivity: 0.9927) and an accuracy of 0.9327 for diagnosis (specificity: 0.8742, sensitivity: 0.9851). External validation showed AUCs of 0.910 and 0.957 for early warning and 0.960 and 0.893 for diagnosis. Ablation analyses indicated that structured clinical information contributed most strongly to diagnostic performance, whereas image-based information provided complementary, interpretable ocular-surface evidence.

Conclusion

The proposed DED-MLLM demonstrates excellent performance in both early warning and diagnosis of DED.

Key points

Problem: Can a multimodal large language model (MLLM) integrate subjective patient-reported outcomes and objective clinical and imaging data to enable comprehensive early warning and diagnosis of dry eye disease (DED) across diverse healthcare settings?

Findings: In a multicenter study of 874 participants (internal cohort $n=743$; external validation $n=131$ across three independent centers), DED-MLLM achieved robust discriminative performance for both early warning (AUCs 0.91–0.96; sensitivity >95%) and diagnosis (AUCs 0.89–0.96) under international and Chinese diagnostic criteria. Ablation studies confirmed that the integration of structured tabular data—including subjective questionnaires (OSDI/SPEED) and objective clinical metrics—was essential for diagnostic accuracy (AUC decreased by 0.3 when

removed). The model demonstrated strong generalizability across external validation sites with varying patient demographics and clinical workflows.

Significance: These findings suggest that instruction-tuned multimodal large language models can effectively synthesize heterogeneous subjective and objective data sources for comprehensive DED assessment, providing a scalable and clinically robust solution to enhance early detection and diagnostic accuracy in both specialized and community-based ophthalmic care.

1. Introduction

Dry eye disease (DED) is a multifactorial, chronic disorder of the ocular surface that manifests as a spectrum of symptoms, including dryness, burning sensation, and blurred vision(1, 2). Beyond physical discomfort, DED substantially impairs visual function, quality of life, work productivity, and psychological well-being(3–6). Its etiology is complex and involves a wide range of risk factors, including advanced age, systemic comorbidities (e.g., diabetes, thyroid disorders, and hypertension), as well as lifestyle-related factors such as poor sleep quality and depression(3, 5–7). Consequently, DED assessment often depends on the combined interpretation of patient-reported outcomes, clinical history, and objective ocular surface examinations, including the Ocular Surface Disease Index (OSDI)(8) and the Standard Patient Evaluation of Eye Dryness (SPEED) questionnaire(9).

The high prevalence and variable presentation of DED create a practical challenge for early recognition and standardized assessment. Epidemiological studies have reported a wide global prevalence of DED, ranging from 5.8% to 85.6% (10–12), with a pooled prevalence of 17.0% in China (13). In many regions, high-quality diagnostic resources, including ocular surface analyzers and tear-related testing devices, remain concentrated in tertiary centers. Low disease awareness may further delay clinical consultation(14). A body of economic research further verifies that DED brings heavy financial burdens to patients and healthcare systems, with average annual direct costs over 400 USD per patient worldwide(15–20). Combined, these clinical, resource and economic barriers emphasize the critical necessity of efficient early screening and diagnostic tools to deliver timely intervention.

AI-based approaches have been applied to multiple ophthalmic tasks,, including keratitis (21), keratoconus (22) and glaucoma(23). In DED, AI-based approaches have been developed for automated analysis of tear film breakup time (24), lipid layer classification (25), meibomian gland morphology(26), and corneal nerve structure(27). Although these methods are valuable for targeted image or signal analysis, their clinical use may be limited when DED evaluation requires simultaneous consideration of symptoms, risk factors, examination results, and images. Therefore, a clinically useful AI framework for DED should not only classify ocular images, but also integrate structured and patient-reported data that clinicians routinely use in decision-making.

Recent advances in Multimodal Large Language Models (MLLMs) have demonstrated strong reasoning capabilities in complex medical tasks (28–31). However, conventional fixed-resolution visual encoders (e.g., CLIP and Vision Transformer (ViT) (32)) may fail to capture fine-grained ocular features, such as corneal punctate staining and subtle meibomian gland alterations. While domain-specific models such as VisionFM are effective for image analysis, they lack the capacity to process and integrate clinical text data(33). In contrast, Qwen2-VL addresses these limitations through its 2D Rotary Position Embedding (RoPE), which enhances spatial awareness for precise visual interpretation (34), and its strong language understanding abilities for processing unstructured clinical information. By effectively aligning visual and textual modalities for comprehensive analysis(35), Qwen2-

VL provides an ideal foundation for multimodal medical applications. Accordingly, this study utilizes Qwen2-VL to develop a unified framework for the early warning and diagnosis of DED.

In this study, we developed DED-MLLM, an instruction-tuned model designed to integrate ophthalmic imaging with structured and unstructured clinical information for DED risk stratification and diagnostic support. The study focused on two practical questions: whether heterogeneous clinical inputs can be organized within a unified MLLM framework, and whether the resulting model maintains performance across internal and external cohorts under both Chinese and international diagnostic criteria.

2. Methods

2.1. Ethics

This study was approved by the Medical Science Research Ethics Committee of Peking University Third Hospital (Protocol Number: M20241030) and was conducted in accordance with the principles of the Declaration of Helsinki. All applicable ethical guidelines were strictly followed. Written informed consent was obtained from all patients after a full explanation of the study's purpose and potential implications.

2.2. Patient Selection and Data Collection

Clinical data were collected from four hospitals and organized into two cohorts. The training internal development cohort comprised 743 patients (1,486 eyes) from Peking University Third Hospital, which was used for model training and testing. An independent external validation cohort included 131 patients recruited from three centers: The Second Hospital of Jilin University, Fushun Eye Hospital, and Zhangjiakou First Hospital. Baseline characteristics were summarized separately for the internal development cohort and the external validation cohort.

Based on the *Chinese Expert Consensus on Dry Eye: Examination and Diagnosis (2020)* (36) and the *TFOS DEWS II Diagnostic Methodology report* (37), patients were classified as non-DED or DED by experienced ophthalmologists using clinical history and ophthalmic imaging findings. Each case was independently reviewed and confirmed. The ophthalmologists were blinded to the outputs of the deep learning model, including both early warning and diagnostic results.

Exclusion criteria included active bacterial or viral ocular infection; active or chronic allergic, bacterial, or viral ocular adnexal infection; infectious keratitis; blepharitis; lagophthalmos; any history of ocular surgery within the preceding 6 months; and long-term use of ocular medications.

During data collection for the early warning model, all patients attending the dry eye specialist clinic were required to complete a self-assessment questionnaire, the OSDI questionnaire, and the SPEED questionnaire. The self-assessment questionnaire collected demographic information, medical history, symptom profiles, and lifestyle factors (see Supplementary Table S1). A secondary verification step was performed to ensure diagnostic accuracy. For minor participants, questionnaires were completed with assistance from a parent or legal guardian.

Diagnostic data were derived from OSDI and SPEED questionnaire scores in conjunction with clinical examination findings. The following objective ophthalmic parameters were collected: morphology of the upper and lower meibomian glands, anterior segment photography, slit-lamp examination, upper and lower eyelid margin assessment, anterior segment staining, and other relevant clinical features.

Ocular images were acquired using a photo slit-lamp system (BX 900, Haag-Streit, Bern, Switzerland) and a corneal topographer (OCULUS Optikgeräte GmbH, Germany). Corneal staining was performed using sodium fluorescein ophthalmic strips (Meizilin Pharmaceutical Co, Liaoning, China). All images were initially captured in RAW format to preserve maximal information, then transferred to a personal computer and converted to JPG format (3456×2304 pixels, 24-bit, RGB). To assess model generalizability, external validation datasets were collected from three representative clinical centers using heterogeneous imaging systems. Fushun Eye Hospital and Zhangjiakou First Hospital used slit lamps (Model SLM-6E) and dry eye analyzers (Models DED-2M and DED-1L, Chongqing Kanghua Ruiming, China), while the Second Hospital of Jilin University employed a Takagi anterior segment imaging system (Takagi Seiko, Japan; equipped with iki-image software at 20-megapixel resolution) and a corneal topographer (Model 77000, OCULUS, Germany). To ensure consistency of model input, all external images underwent standardized preprocessing and were resized to match the resolution and RGB format of the internal set.

2.3. Construction of the Multimodal Deep Learning Model

The MLLM developed in this study comprises three primary feature extraction modules: visual, tabular, and language embedding modules. Features from these modalities are fused through a Transformer backbone(38). We adopted the open-source Qwen2-VL as the base architecture, inheriting its advanced visual and language encoding modules. To accommodate structured medical data (e.g., patient questionnaires and ophthalmic examination records), we designed a dedicated tabular embedding module. Structurally aligned with the language module, this component processes tabular inputs that are first converted into JSON format. However, raw structured clinical data containing specialized terms and numerical values (e.g., 'OSDI': 34.09, 'CFS': 1) present challenges for direct semantic interpretation. To address this, tabular entries were transformed into natural language descriptions (see Supplementary Materials), thereby aligning domain-specific medical information with the model's input space.

Following feature extraction, embeddings from each modality are converted into tokens and concatenated into a unified multimodal input sequence. The visual embedding module, based on ViT, employs self-attention mechanisms to capture long-range spatial dependencies in ocular surface images. These image tokens are then jointly processed with tokenized textual and tabular inputs within the shared Transformer backbone. Through cross-modal attention, the model learns fine-grained associations between localized image features and corresponding clinical indicators. This design enables task-oriented feature fusion and decision-making under the guidance of task-specific prompts, thereby facilitating the joint optimization of heterogeneous data sources. The resulting model, termed DED-MLLM (Dry Eye Disease-Multimodal Large Language Model), utilizes this integrated representation to perform multimodal reasoning for both early warning and diagnosis of DED. The overall model construction workflow is illustrated in Fig. 1.

2.4. Model Tasks

Four tasks were defined: early warning and diagnosis of DED under both international and Chinese diagnostic criteria. For the early-warning tasks, questionnaire-based and patient-level risk-factor data were used as inputs. For the early-diagnosis tasks, ophthalmic examination findings and imaging data were additionally incorporated. All samples were randomly divided into development and test sets at a ratio of 8:2. The overall workflow of model development is presented in Fig. 2. In addition, ablation studies were conducted to quantify the contribution of each modality, and external validation was performed to assess model robustness and generalizability. External validation was conducted after completion of model development, validation, and internal testing. The external cohort was not used for training, model selection, or threshold determination.

2.5. Statistical Analysis

Model performance was evaluated using standard metrics derived from the confusion matrix including sensitivity, specificity, accuracy, and F1-score. All analyses were conducted in Python 3.12. Data processing was primarily performed using pandas and NumPy, model performance metrics were computed using scikit-learn, and data visualization was carried out with Matplotlib and Seaborn.

3. Results

We evaluated the performance of our DED-MLLM framework in early warning and DED diagnosis based on both international (TFOS DEWS II) and Chinese (2020 Chinese Expert Consensus) diagnostic criteria. Following a summary of the cohort characteristics (Table 1), the diagnostic efficacy of the model was analyzed, with results and ROC curves detailed in Table 2 and Fig. 3.

Table 1

Baseline demographic and ocular surface characteristics of the internal and external validation cohorts. Data are presented as median (interquartile range) or number (percentage), unless otherwise indicated. Patient-level variables include age, sex, and revisit status. Eye-level clinical variables were analyzed separately for the internal cohort from Peking University Third Hospital and the external validation cohort. Abbreviations: DED, dry eye disease; NIBUT, non-invasive breakup time; FBUT, fluorescein breakup time; IOP, intraocular pressure. Meibomian gland loss was graded based on the proportion of gland dropout at the eyelid margin: grade 1, loss < 1/3; grade 2, loss > 1/3; grade 3, loss > 2/3.

Characteristic	Peking University Third Hospital	External validation
Demographics		
Age, years	53 (32–69); n = 743	60 (52–67); n = 131
Sex	Female: 502 (67.6%); Male: 241 (32.4%)	Female: 94 (71.8%); Male: 37 (28.2%)
Revisit	Yes: 101 (13.6%); No: 642 (86.4%)	Yes: 13 (9.9%); No: 118 (90.1%)
Clinical examination	Eye-level variables; total n = 1,486 eyes vs 262 eyes	
Current DED severity	Mild: 837 (56.3%); Moderate: 271 (18.2%); Severe: 342 (23.0%); None: 4 (0.3%); Not recorded: 32 (2.2%)	Mild: 128 (48.9%); Moderate: 80 (30.5%); Severe: 30 (11.5%); None: 1 (0.4%); Not recorded: 23 (8.8%)
Corneal fluorescein staining score	Right: 1 (0–2), n = 731; Left: 1 (0–2), n = 729	Right: 1.0 (0–3), n = 109; Left: 0.5 (0–3), n = 108
Non-invasive breakup time (NIBUT), s	Right: 6.72 (4.59–10.71), n = 582; Left: 6.66 (4.76–10.92), n = 580	Right: 1.0 (1.0–1.0), n = 124; Left: 1.0 (1.0–1.0), n = 125
Fluorescein breakup time (FBUT), s	Right: 3.00 (3.00–4.00), n = 717; Left: 3.00 (3.00–4.00), n = 717	Right: 3.0 (2.0–4.0), n = 56; Left: 3.0 (2.0–4.0), n = 56
Tear meniscus height (NIBUT), mm	Right: 0.18 (0.15–0.22), n = 713; Left: 0.18 (0.14–0.22), n = 714	Right: 0.20 (0.16–0.25), n = 123; Left: 0.21 (0.16–0.25), n = 123
Non-contact intraocular pressure, mmHg	Right: 14.90 (13.00–16.00), n = 699; Left: 15.00 (13.00–16.00), n = 699	Right: 16 (15–18), n = 125; Left: 17 (15–18), n = 125
Meibomian gland loss at the eyelid margin	Right: grade 1 (< 1/3): 524 (70.5%); grade 2 (> 1/3): 156 (21.0%); grade 3 (> 2/3): 42 (5.7%); no loss: 4 (0.5%); not recorded: 17 (2.3%) Left: grade 1 (< 1/3): 521 (70.1%); grade 2 (> 1/3): 159 (21.4%); grade 3 (> 2/3): 43 (5.8%); no loss: 4 (0.5%); not recorded: 16 (2.2%)	Right: grade 1 (< 1/3): 50 (38.2%); grade 2 (> 1/3): 38 (29.0%); grade 3 (> 2/3): 6 (4.6%); no loss: 27 (20.6%); not recorded: 10 (7.6%) Left: grade 1 (< 1/3): 50 (38.2%); grade 2 (> 1/3): 40 (30.5%); grade 3 (> 2/3): 4 (3.1%); no loss: 27 (20.6%); not recorded: 10 (7.6%)

Table 2

The discriminative performance of DED-MLLM was evaluated on tasks of early warning and diagnosis for dry eye disease (DED) according to both the international diagnostic criteria (TFOS DEWS II) and the Chinese diagnostic criteria (2020 Chinese Expert Consensus). The case proportions are as follows: in the early warning task, 50.87% (378/743 patients) meet the international criteria and 58.82% (437/743 patients) meet the Chinese criteria; in the diagnosis task, 43.07% (640/1486 eyes) comply with the international criteria and 52.62% (782/1486 eyes) comply with the Chinese criteria.

Task	Criteria	Sample (n)	Cases, n (%)	AUC	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)	F1-score (95% CI)
Early Warning	International	743 patients	378 (50.87%)	0.929	95.58% (90.61–100.00)	68.73% (44.50–92.96)	82.32% (72.70–91.95)	84.36% (78.48–90.78)
	Chinese	743 patients	437 (58.82%)	0.953	99.27% (96.16–100.00)	91.12% (89.97–92.28)	95.84% (94.08–97.60)	96.72% (95.40–98.04)
Diagnosis	International	1486 eyes	640 (43.07%)	0.925	97.20% (94.71–99.68)	87.76% (85.55–89.96)	92.03% (90.51–93.54)	91.49% (89.88–93.10)
	Chinese	1486 eyes	782 (52.62%)	0.944	98.51% (97.69–99.34)	87.42% (85.66–89.18)	93.27% (93.27–93.27)	93.89% (93.06–94.72)

3.1. Performance in Early Warning and Diagnosis Tasks

In the early-warning task (n = 743 patients), the model demonstrated promising performance based on questionnaire and risk-factor information. Under the Chinese criteria, DED-MLLM achieved an AUC of 0.953, with a sensitivity of 99.27% (95% CI: 96.16–100.00%), a specificity of 91.12% (95% CI: 89.97–92.28%), and an overall accuracy of 95.84% (Table 2). Under the international criteria, the model achieved an AUC of 0.929, with a sensitivity of 95.58% (95% CI: 90.61–100.00%) and an accuracy of 82.32%.

For the early-diagnosis task (n = 1,486 eyes), the model additionally incorporated ophthalmic examination and imaging data and maintained robust performance across both standards. According to the international criteria, DED-MLLM achieved an AUC of 0.925, with a sensitivity of 97.20% (95% CI: 94.71–99.68%), a specificity of 87.76% (95% CI: 85.55–89.96%) and an accuracy of 92.03%. Under the Chinese criteria, the model achieved an AUC of 0.944, with a sensitivity of 98.51% (95% CI: 97.69–99.34%) and an accuracy of 93.27% for distinguishing DED from non-DED eyes.

3.2. Interpretability of DED-MLLM

To assess model interpretability, we assessed attention weights assigned to different input features (Fig. 4). In the early-warning task, "patient age" and "history of ocular diseases" emerged as the most influential predictors under both diagnostic frameworks. The model also demonstrated adaptive behavior by weighting additional risk factors differently depending on the diagnostic criteria. For example, under the international criteria, greater importance was assigned to "occupational environment" and "short-term electronic device use," whereas under the Chinese criteria, the model emphasized "chronic comorbidities" and "contact lens wear". In addition, attention heatmaps were generated to visualize the model's focus on ophthalmic images. As shown in Fig. 4c-f, regions with close attention weights (highlighted in warm colors) corresponded closely to clinically relevant pathological features.

The model accurately localized hallmark signs of DED, including conjunctival vascular dilation, meibomian gland obstruction or atrophy, and corneal punctate staining.

3.3. Ablation Studies and External Validation

To further investigate the contribution of different modalities and evaluate model generalizability, we conducted ablation studies and external validation experiments. The ablation analysis focused on the diagnostic task (Fig. 5 and Table S1). Removal of the image module resulted in only marginal decreases in AUC (0.021 and 0.004 under the international and Chinese criteria, respectively), with overall accuracy remaining above 91%. In contrast, removal of the tabular module led to a substantial decline in performance, with AUC values decreasing by approximately 0.3, highlighting the critical importance of structured clinical data.

For external validation, data were collected from three independent hospitals using heterogeneous imaging systems. Utilizing the dynamic resolution capability of the Qwen2 backbone, images were processed without additional preprocessing. As shown in Table S2 and Fig. 6, the model demonstrated strong and consistent performance across all tasks. In early warning tasks, AUCs reached 0.910 (international criteria) and 0.957 (Chinese criteria). In diagnostic tasks, the model maintained high discriminative ability, achieving AUCs of 0.960 (international criteria) and 0.893 (Chinese criteria), confirming its robustness and effectiveness in diverse real-world clinical settings. These results describe the relative contribution of each modality and do not independently assess the relationship between model inputs and the diagnostic reference criteria.

Discussion

In this study, we propose an innovative tabular embedding module specifically designed to integrate structured clinical data with MLLMs such as Qwen2-VL. This module effectively "translates" structured clinical variables into natural language representations that are interpretable by the MLLM, enabling semantic-level fusion with ophthalmic imaging data. The resulting multimodal framework demonstrated excellent performance in both early-warning and diagnostic tasks for DED, achieving AUC values exceeding 0.92 under both international and Chinese diagnostic criteria. Moreover, evaluation on an independent external validation cohort confirmed the robustness and generalizability of the model. By addressing the heterogeneity inherent in multimodal clinical data and utilizing the strong cross-modal reasoning capabilities of Qwen2-VL, the proposed DED-MLLM provides a viable technical pathway toward accurate and interpretable AI-assisted diagnosis of DED.

Previous AI research in DED has primarily focused on unimodal analysis, such as blink patterns, corneal staining images, or tear film breakup time^(39–43). Although these approaches perform well in specific settings, they may not fully reflect real-world clinical reasoning, where physicians integrate symptoms, risk factors, examination findings, and imaging features. In contrast, DED-MLLM was developed under two widely used diagnostic frameworks: the TFOS DEWS II Diagnostic Methodology report³⁷ and the Chinese Expert Consensus on Dry Eye: Examination and Diagnosis (2020)³⁶. Both emphasize that DED diagnosis requires multimodal assessment rather than reliance on isolated indicators. Consistent with this principle, our results indicate that integrating complementary clinical information improves disease representation, in line with prior evidence supporting multimodal learning in medical AI (44, 45).

The model showed particularly strong performance for early warning under the Chinese diagnostic criteria, with an internal AUC of 0.953 and an external AUC of 0.957. This is clinically relevant because DED is highly prevalent,

especially among older adults (10, 46). In addition, discordance between symptoms and signs may delay consultation and intervention(47). Interpretability analysis identified age, ocular disease history, recent electronic device use, and occupational environment as influential variables. These factors are consistent with known epidemiological and behavioral risks for DED(48). Therefore, DED-MLLM may help use patient-reported information, including OSDI, SPEED, and self-assessment questionnaires, to support low-barrier risk stratification before comprehensive hospital-based imaging is available.

The ablation results deserve careful interpretation. Removing the tabular module caused a substantial performance decline, whereas removing the image module produced only a modest decrease. This pattern is consistent with the clinical nature of DED, which is strongly influenced by subjective symptoms and structured measures such as OSDI and SPEED. It also indicates that, in the current dataset, the value of DED-MLLM lies primarily in integrating and interpreting structured clinical information, with imaging acting as a complementary source for diagnostic confirmation and interpretability. This result should not be viewed as weakening the multimodal framework, but rather as clarifying which modality carries the greatest diagnostic signal for DED.

Despite these limitations, attention heatmaps indicate that the model successfully focuses on clinically relevant regions, including scleral vasculature, meibomian gland obstruction, and corneal staining areas, thereby enhancing its diagnostic capability. Interestingly, the model also identified additional features not traditionally emphasized in clinical assessment, such as eyelash texture and density, iris pigmentation patterns, and limbal vasculature, suggesting the potential to uncover novel imaging biomarkers for DED.

Several limitations of this study should be acknowledged. First, the dataset lacks sufficient geographical and racial diversity, as it was derived from a limited number of centers. Second, due to regional characteristics and the demographic composition of the study population (predominantly older adults)(49, 50), moderate-to-severe DED cases were overrepresented, while mild cases were relatively underrepresented. This imbalance may limit the model's sensitivity to early-stage disease and its ability to capture features across the full severity spectrum. Third, further improvements in image modality representation and recognition are needed. Future work will focus on expanding multi-center collaborations to increase data diversity and optimize the model, thereby enhancing its generalizability and enabling more personalized diagnostic applications. Fourth, Because OSDI, SPEED, and some ocular-surface findings may overlap with components of the diagnostic criteria, the reported results should be interpreted as agreement with predefined clinical criteria rather than independent diagnostic superiority.

Conclusion

The proposed DED-MLLM demonstrated promising performance for early warning and early diagnosis of DED. The model integrated questionnaire and risk-factor information for early warning and incorporated ophthalmic examination and imaging data for early diagnosis.

Abbreviations

AI

artificial intelligence

AUC

area under the receiver operating characteristic curve

CFS

corneal fluorescein staining
CI
confidence interval
DED
dry eye disease
DED-MLLM
dry eye disease multimodal large language model
FBUT
fluorescein breakup time
IOP
intraocular pressure
JSON
JavaScript Object Notation
MLLM
multimodal large language model
NIBUT
non-invasive breakup time
OSDI
Ocular Surface Disease Index
RGB
red-green-blue
ROC
receiver operating characteristic
RoPE
rotary position embedding
SPEED
Standard Patient Evaluation of Eye Dryness
TFOS DEWS II
Tear Film and Ocular Surface Society Dry Eye Workshop II
USD
United States dollars
ViT
Vision Transformer.

Declarations

Ethics approval and consent to participate

This study was approved by the Medical Science Research Ethics Committee of Peking University Third Hospital (Protocol No. M20241030) and was conducted in accordance with the principles of the Declaration of Helsinki. Written informed consent was obtained from all participants after a full explanation of the study purpose and potential implications. For participants younger than 18 years, written informed consent was obtained from a parent or legal guardian.

Consent for publication

Written informed consent for publication of de-identified ocular images and related clinical information was obtained from all participants whose data or images are included in this article. For participants younger than 18 years, consent for publication was obtained from a parent or legal guardian.

Competing interests

The authors declare that they have no competing interests.

Authors' information

The authors' institutional affiliations and corresponding author information are provided on the title page.

Funding

This study was supported by the National Health and Family Planning Commission of the People's Republic of China (project number 2022YB57). The funder had no role in the design of the study; collection, analysis, and interpretation of the data; or writing of the manuscript.

Author Contribution

Mingyan He contributed to project design and data collection, data organization, data analysis, manuscript writing and revision. Lili Cao was responsible for project design, model construction, data analysis, manuscript writing and revision. Kechao Zheng was responsible for project design, model construction and data analysis. Zhiping Lin contributed to data collection, and manuscript revision. Haoran Wu was involved in project design ,manuscript revision.Shuzhen Wang, Rui Tian, Yu Wang,Tiehong Chen,Wei Wang,Liang Xiao,Tingting Yang were responsible for data collection. Jing Hong contributed to project design and manuscript revision.

Acknowledgements

Not applicable.

Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available because access is restricted by hospital regulations concerning patient privacy and data protection. De-identified data may be made available from the corresponding author upon reasonable request, subject to approval by the relevant institutional ethics committee and compliance with applicable hospital policies.

References

1. Aragona P, Barabino S, Di Zazzo A, Giannaccare G, Villani E, Aiello F, et al. Dry Eye Disease: From Causes to Patient Care and Clinical Collaboration-A Narrative Review. *Ophthalmol Ther.* 2025;14(7):1411–28.

2. García-Marqués JV, Talens-Estarellles C, García-Lázaro S, Wolffsohn JS, Cerviño A. Systemic, environmental and lifestyle risk factors for dry eye disease in a mediterranean caucasian population. *Cont Lens Anterior Eye*. 2022;45(5):101539.
3. Jareebi MA, Akkur AA, Otayf DAH, Najmi AY, Mobarki OA, Omar EZ et al. Exploring the Prevalence of Dry Eye Disease and Its Impact on Quality of Life in Saudi Adults: A Cross-Sectional Investigation. *Med (Kaunas)*. 2025;61(6).
4. Grubbs JR Jr., Tolleson-Rinehart S, Huynh K, Davis RM. A review of quality of life measures in dry eye questionnaires. *Cornea*. 2014;33(2):215–8.
5. Friedman NJ. Impact of dry eye disease and treatment on quality of life. *Curr Opin Ophthalmol*. 2010;21(4):310–6.
6. He Q, Chen Z, Xie C, Liu L, Yang H, Wei R. Relationship Between Dry Eye Disease and Emotional Disorder: The Mediating Effect of Health Anxiety. *Front Public Health*. 2022;10:771554.
7. Jongkhajornpong P, Lekhanont K, Anothaisintawee T, Rattanasiri S, McKay G, Attia J, et al. Prevalence of dry eye disease and its association with sleep quality and depression: a hospital-based survey in Thai population. *BMJ Open*. 2025;15(6):e094046.
8. Schiffman RM, Christianson MD, Jacobsen G, Hirsch JD, Reis BL. Reliability and validity of the Ocular Surface Disease Index. *Arch Ophthalmol*. 2000;118(5):615–21.
9. Asiedu K. Rasch Analysis of the Standard Patient Evaluation of Eye Dryness Questionnaire. *Eye Contact Lens*. 2017;43(6):394–8.
10. Stapleton F, Alves M, Bunya VY, Jalbert I, Lekhanont K, Malet F, et al. TFOS DEWS II Epidemiology Report. *Ocul Surf*. 2017;15(3):334–65.
11. The epidemiology of dry eye disease: report of the Epidemiology Subcommittee of the International Dry Eye WorkShop. (2007). *Ocul Surf*. 2007;5(2):93–107.
12. Papas EB. The global prevalence of dry eye disease: A Bayesian view. *Ophthalmic Physiol Opt*. 2021;41(6):1254–66.
13. Liu NN, Liu L, Li J, Sun YZ. Prevalence of and risk factors for dry eye symptom in mainland china: a systematic review and meta-analysis. *J Ophthalmol*. 2014;2014:748654.
14. Fan Z, Du Y, Tang C, Tian R, Lu X, Zheng L, et al. Awareness, Prevalence, and Knowledge of Dry Eye Among Internet Professionals: A Cross-sectional Study in China. *Eye Contact Lens*. 2023;49(3):92–7.
15. Luo Y, Yang W, Qi M, Wang Y, Li S, Wang M, et al. Annual Direct Economic Burden and Influencing Factors of Dry Eye Diseases in Central China. *Ophthalmic Epidemiol*. 2023;30(2):121–8.
16. Chan C, Ziai S, Myageri V, Burns JG, Prokopich CL. Economic burden and loss of quality of life from dry eye disease in Canada. *BMJ Open Ophthalmol*. 2021;6(1):e000709.
17. Yu J, Asche CV, Fairchild CJ. The economic burden of dry eye disease in the United States: a decision tree analysis. *Cornea*. 2011;30(4):379–87.
18. McDonald M, Patel DA, Keith MS, Snedecor SJ. Economic and Humanistic Burden of Dry Eye Disease in Europe, North America, and Asia: A Systematic Literature Review. *Ocul Surf*. 2016;14(2):144–67.
19. Clegg JP, Guest JF, Lehman A, Smith AF. The annual cost of dry eye syndrome in France, Germany, Italy, Spain, Sweden and the United Kingdom among patients managed by ophthalmologists. *Ophthalmic Epidemiol*. 2006;13(4):263–74.

20. Waduthantri S, Yong SS, Tan CH, Shen L, Lee MX, Nagarajan S, et al. Cost of dry eye treatment in an Asian clinic setting. PLoS ONE. 2012;7(6):e37711.
21. Lv J, Zhang K, Chen Q, Chen Q, Huang W, Cui L, et al. Deep learning-based automated diagnosis of fungal keratitis with in vivo confocal microscopy images. Ann Transl Med. 2020;8(11):706.
22. Tan Z, Chen X, Li K, Liu Y, Cao H, Li J, et al. Artificial Intelligence-Based Diagnostic Model for Detecting Keratoconus Using Videos of Corneal Force Deformation. Transl Vis Sci Technol. 2022;11(9):32.
23. Hood DC, De Moraes CG. Efficacy of a Deep Learning System for Detecting Glaucomatous Optic Neuropathy Based on Color Fundus Photographs. Ophthalmology. 2018;125(8):1207–8.
24. Shimizu E, Ishikawa T, Tanji M, Agata N, Nakayama S, Nakahara Y, et al. Artificial intelligence to estimate the tear film breakup time and diagnose dry eye disease. Sci Rep. 2023;13(1):5822.
25. Hwang H, Jeon HJ, Yow KC, Hwang HS, Chung E. Image-based quantitative analysis of tear film lipid layer thickness for meibomian gland evaluation. Biomed Eng Online. 2017;16(1):135.
26. Yeh CH, Yu SX, Lin MC. Meibography Phenotyping and Classification From Unsupervised Discriminative Feature Learning. Transl Vis Sci Technol. 2021;10(2):4.
27. Aggarwal S, Kheirkhah A, Cavalcanti BM, Cruzat A, Jamali A, Hamrah P. Correlation of corneal immune cell changes with clinical severity in dry eye disease: An in vivo confocal microscopy study. Ocul Surf. 2021;19:183–9.
28. Tsai YH, Bai S, Pu Liang P, Kolter JZ, Morency LP, Salakhutdinov R. Multimodal Transformer for Unaligned Multimodal Language Sequences. Proc Conf Assoc Comput Linguist Meet. 2019;2019:6558–69.
29. Chen Y-C, Li L, Yu L, El Kholy A, Ahmed F, Gan Z, et al. editors. Uniter: Universal image-text representation learning. European conference on computer vision; 2020: Springer.
30. Shao J, Ma J, Yu Y, Zhang S, Wang W, Li W, et al. A multimodal integration pipeline for accurate diagnosis, pathogen identification, and prognosis prediction of pulmonary infections. Innov (Camb). 2024;5(4):100648.
31. Zarei O, Talebi Moghaddam M, Moradi Vastegani S. Machine learning and deep learning in clinical practice: Advancing neurodegenerative disease diagnosis with multimodal markers. Brain Res Bull. 2025;234:111667.
32. Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, et al. editors. Learning transferable visual models from natural language supervision. International conference on machine learning; 2021: PmLR.
33. Qiu J, Wu J, Wei H, Shi P, Zhang M, Sun Y et al. Visionfm: a multi-modal multi-task vision foundation model for generalist ophthalmic artificial intelligence. arXiv preprint arXiv:231004992. 2023.
34. Su J, Ahmed M, Lu Y, Pan S, Bo W, Liu Y. Roformer: Enhanced transformer with rotary position embedding. Neurocomputing. 2024;568:127063.
35. Li J, Guan Z, Wang J, Cheung CY, Zheng Y, Lim LL, et al. Integrated image-based deep learning and language models for primary diabetes care. Nat Med. 2024;30(10):2886–96.
36. 王宇, 李俊, 关卓, 王健, 郑颖, 林立立, et al. 基于深度学习和语言模型的糖尿病视网膜病变早期筛查模型构建与验证。中华医学杂志. 2020;99(10):741–7.
37. Wolffsohn JS, Arita R, Chalmers R, Djalilian A, Dogru M, Dumbleton K, et al. TFOS DEWS II Diagnostic Methodology report. Ocul Surf. 2017;15(3):539–74.
38. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN et al. Attention is all you need. Adv Neural Inf Process Syst. 2017;30.
39. Zheng Q, Wang L, Wen H, Ren Y, Huang S, Bai F, et al. Impact of Incomplete Blinking Analyzed Using a Deep Learning Model With the Keratograph 5M in Dry Eye Disease. Transl Vis Sci Technol. 2022;11(3):38.

40. Ren Y, Wen H, Bai F, Huang B, Wang Z, Zhang S, et al. Comparison of deep learning-assisted blinking analysis system and Lipiview interferometer in dry eye patients: a cross-sectional study. *Eye Vis (Lond)*. 2024;11(1):7.
41. El Barche FZ, Benyoussef AA, El Habib Daho M, Lamard A, Quellec G, Cochener B, et al. Automated tear film break-up time measurement for dry eye diagnosis using deep learning. *Sci Rep*. 2024;14(1):11723.
42. Nejat F, Eghtedari S, Alimoradi F. Next-Generation Tear Meniscus Height Detecting and Measuring Smartphone-Based Deep Learning Algorithm Leads in Dry Eye Management. *Ophthalmol Sci*. 2024;4(5):100546.
43. Elsayy 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.
44. Syed MG, Trucco E, Mookiah MRK, Lang CC, McCrimmon RJ, Palmer CNA, et al. Deep-learning prediction of cardiovascular outcomes from routine retinal images in individuals with type 2 diabetes. *Cardiovasc Diabetol*. 2025;24(1):3.
45. Yun JS, Kim J, Jung SH, Cha SA, Ko SH, Ahn YB, et al. A deep learning model for screening type 2 diabetes from retinal photographs. *Nutr Metab Cardiovasc Dis*. 2022;32(5):1218–26.
46. Cai Y, Wei J, Zhou J, Zou W. Prevalence and Incidence of Dry Eye Disease in Asia: A Systematic Review and Meta-Analysis. *Ophthalmic Res*. 2022;65(6):647–58.
47. Bartlett JD, Keith MS, Sudharshan L, Snedecor SJ. Associations between signs and symptoms of dry eye disease: a systematic review. *Clin Ophthalmol*. 2015;9:1719–30.
48. Britten-Jones AC, Wang MTM, Samuels I, Jennings C, Stapleton F, Craig JP. Epidemiology and Risk Factors of Dry Eye Disease: Considerations for Clinical Management. *Med (Kaunas)*. 2024;60(9).
49. Um SB, Kim NH, Lee HK, Song JS, Kim HC. Spatial epidemiology of dry eye disease: findings from South Korea. *Int J Health Geogr*. 2014;13:31.
50. Yang W, Wang H, Li J, Chen Y, Zhang Y, Niu X, et al. Demographic and lifestyle factors associated with dry eye disease in China: A cross-sectional multi-center study. *Ocul Surf*. 2024;34:146–55.

Figures

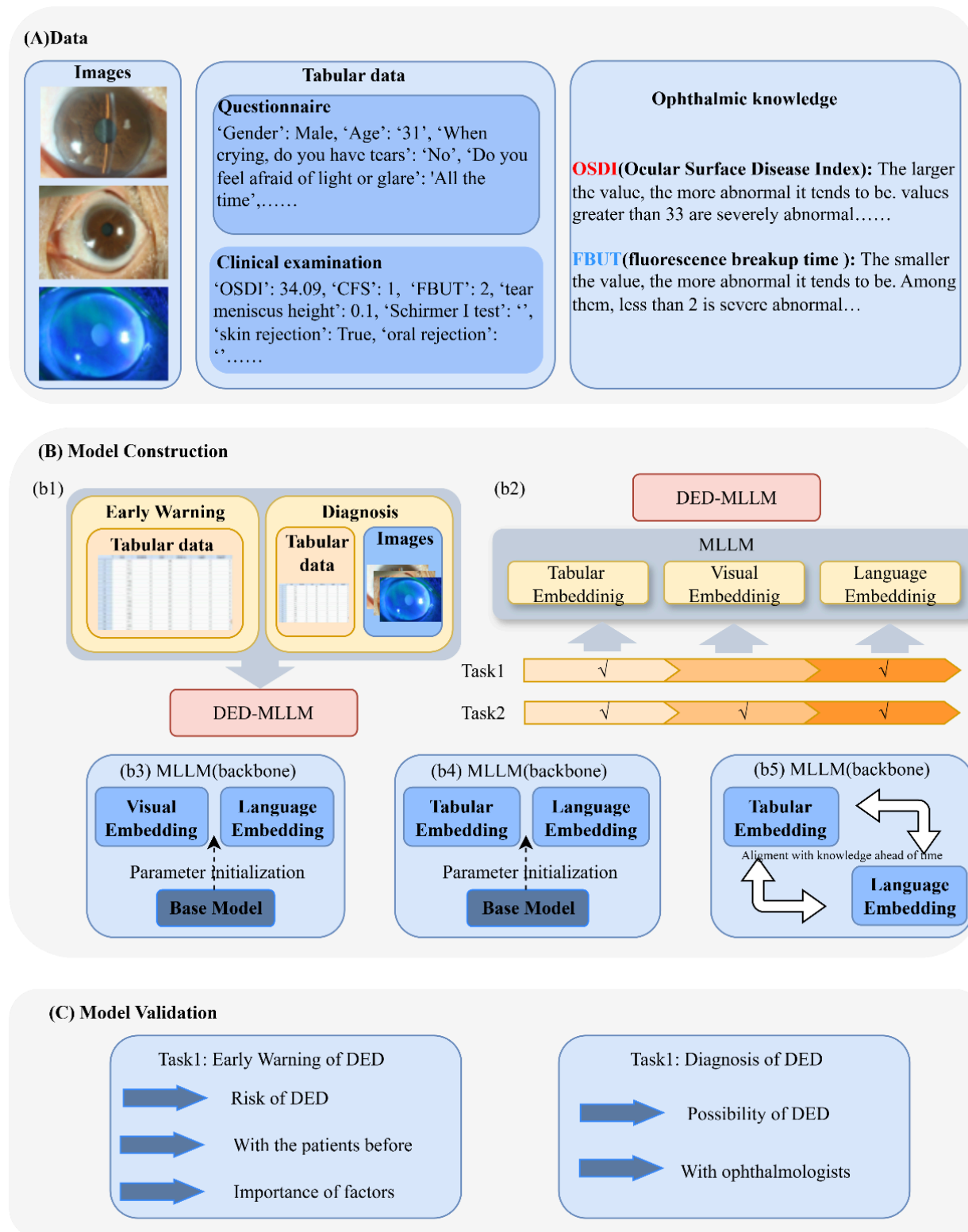


Figure 1

(A) This study integrated multimodal data, including images, text, tabular data, and ophthalmological expertise, to construct a large language model named DED-MLLM. (B) The construction process of the DED-MLLM model is illustrated. (b1) The modeling strategy and workflow are described. Multimodal fusion techniques were first employed, followed by multi-task learning methods to further optimize model performance. (b2) The architecture of the DED-MLLM model for dry eye disease (DED) early warning and diagnosis is shown. DED-MLLM comprises three feature extraction modules, each specifically designed for a particular data type. (b3) The backbone, visual

embedding module, and language embedding module of the MLLM were directly inherited from the base model (Qwen2VL), ensuring consistency and efficiency in processing multimodal data. (b4) (b5) The initial structure and parameters of the tabular embedding module were derived from the language embedding module. However, due to the high specialization and complexity of tabular data, fine-tuning was required to ensure alignment between the tabular data and the knowledge-enhanced text data. (C) Model validation encompassed two primary tasks: DED early warning and diagnosis. The accuracy of the model in these two tasks was evaluated, and attention values were utilized to analyze the key influencing factors in the early warning task.

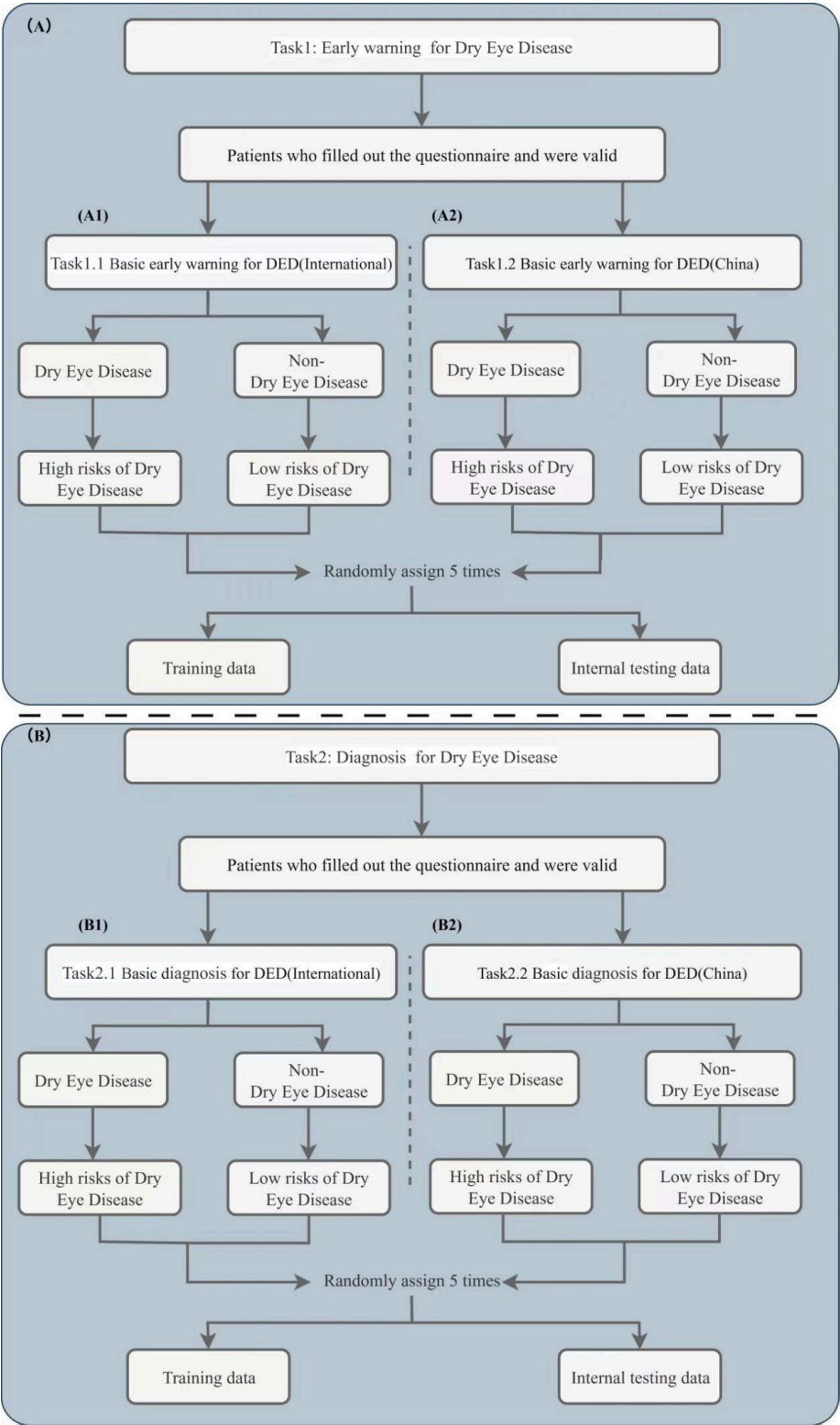


Figure 2

The flowchart illustrates the development and testing processes for the warning task and diagnosis task. All experiments were independently repeated 5 times. For each repetition, data splitting and model training were performed anew, and the final results are reported as the mean values from the 5 repeated experiments, followed by testing and internal validation testing.

AUC Curves of DED-MLLM Across Four Tasks

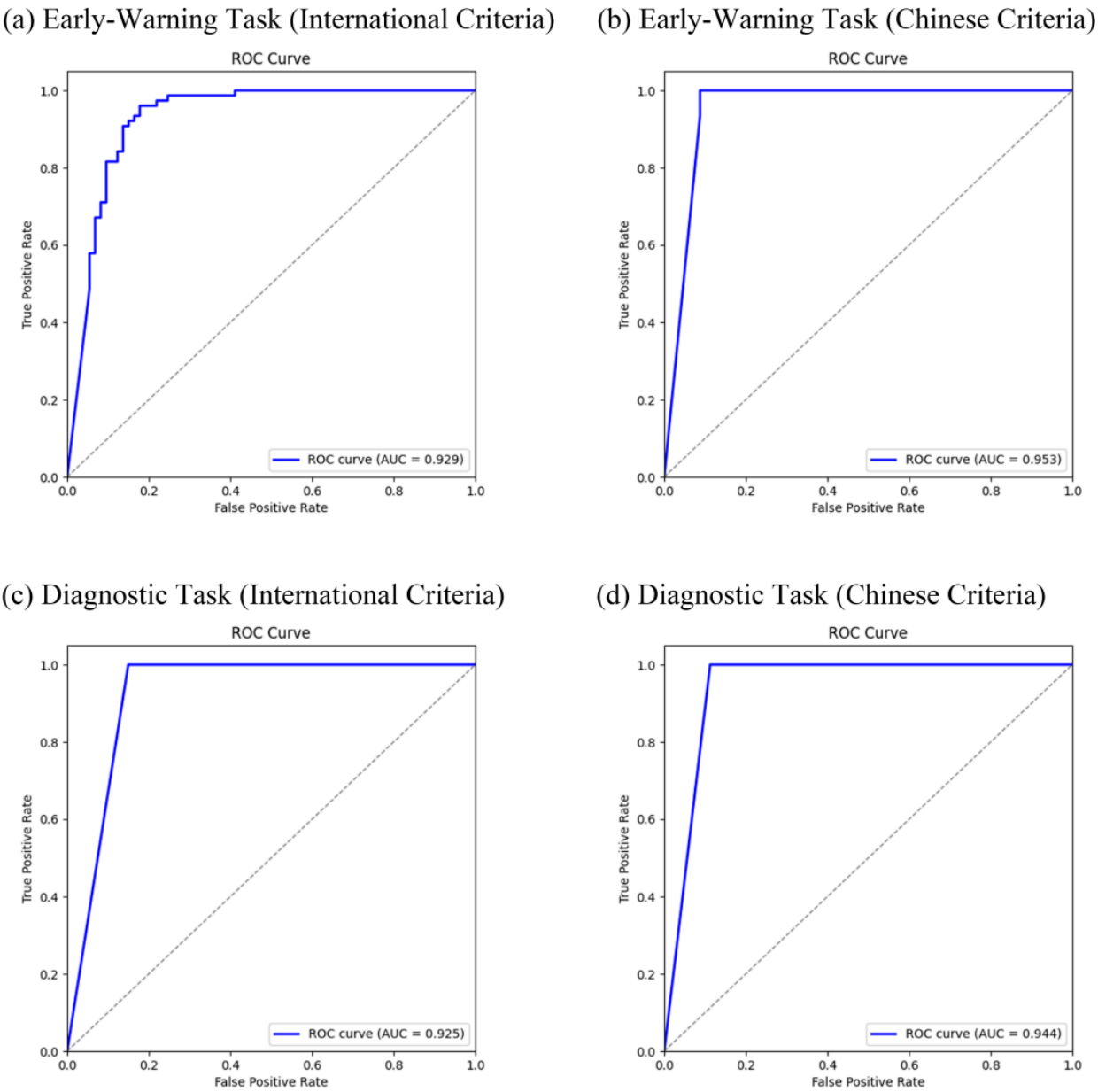


Figure 3

These figures present the AUC curves of the model across four tasks. Panels (a) and (b) display the results of early-warning tasks based on international diagnostic criteria and Chinese diagnostic criteria, with AUC values of 0.929 and 0.953, respectively. Panels (c) and (d) illustrate the results of diagnostic tasks based on international diagnostic criteria and Chinese diagnostic criteria, with AUC values of 0.925 and 0.944, respectively. The results indicate that the model exhibits excellent diagnostic and early-warning performance, with higher predictive accuracy when using the Chinese diagnostic criteria.

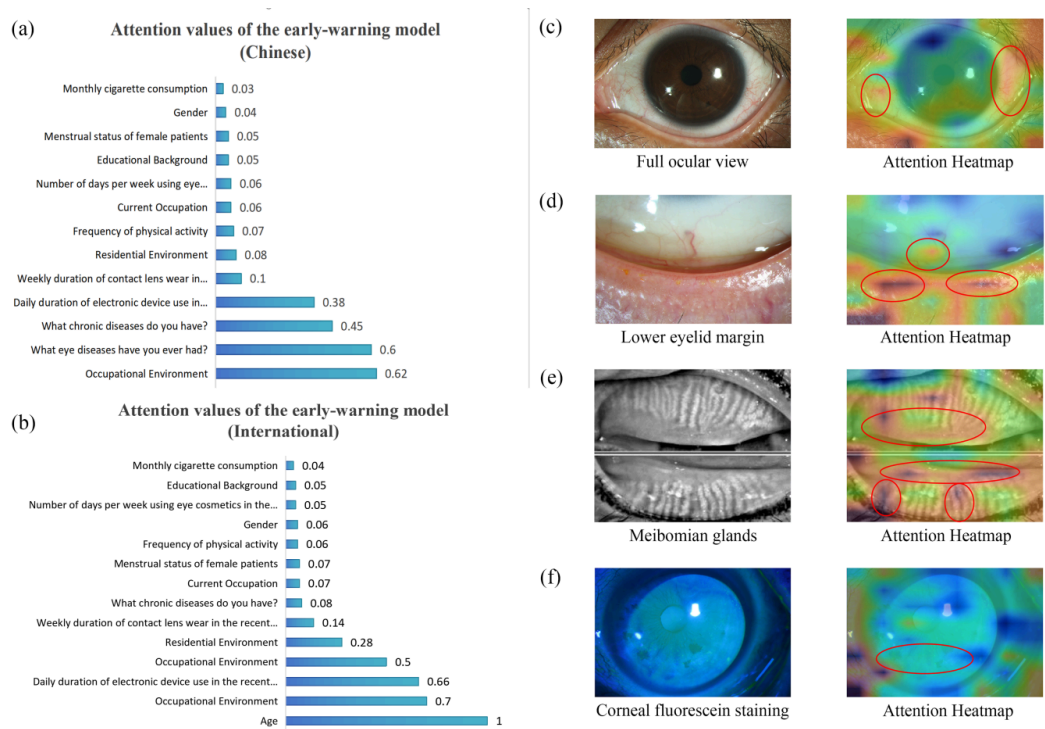
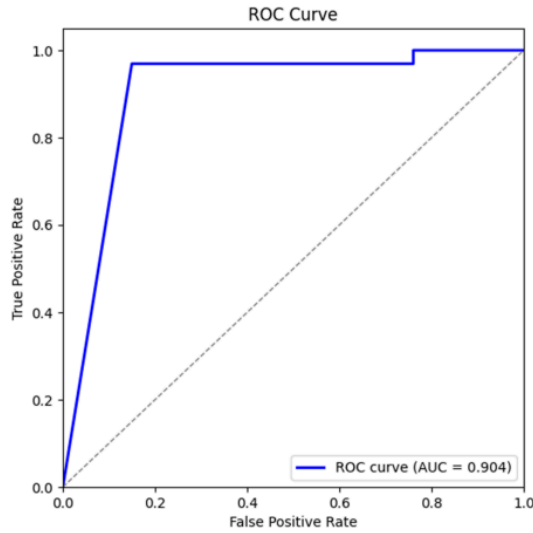


Figure 4

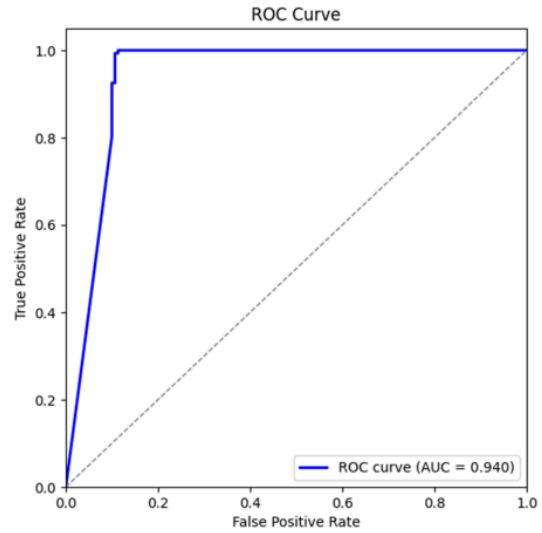
Interpretability analysis of the DED-MLLM. (a-b) Attention weights of patient characteristics in the early-warning task under Chinese (a) and International (b) criteria. "Patient age" was the most prominently attended feature in both scenarios. Differences in other weighted features (e.g., electronic device usage behavior vs. comorbid chronic disorders) reflect the model's adaptation to population heterogeneity across different contexts. (c-f) Visualization of model attention on ophthalmic images using heatmaps. (c) Full ocular view; (d) Lower eyelid margin; (e) Meibomian glands; (f) Corneal fluorescein staining. Warm colors (red/orange) indicate regions with high attention weights. The model focused on pathological signs such as conjunctival vascular dilation, meibomian gland obstruction/atrophy, and corneal punctate staining.

AUC Curves for Different Ablation Settings

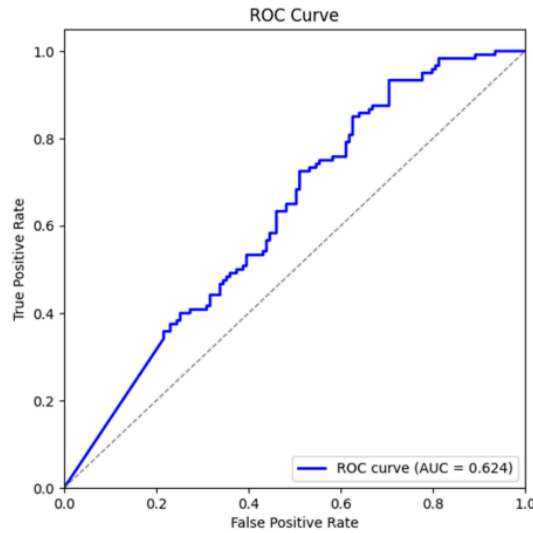
(a) International Criteria (w/o Imaging)



(b) Chinese Criteria (w/o Imaging)



(c) International Criteria (w/o Tabular)



(d) Chinese Criteria (w/o Tabular)

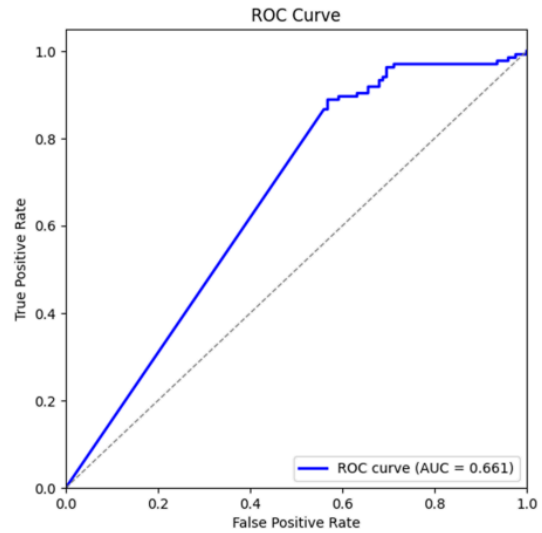
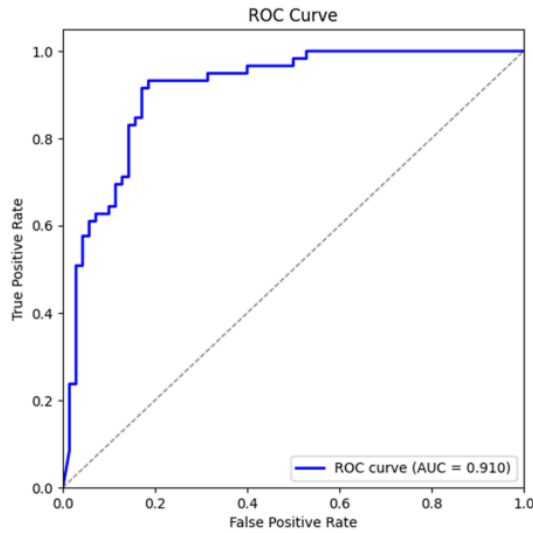


Figure 5

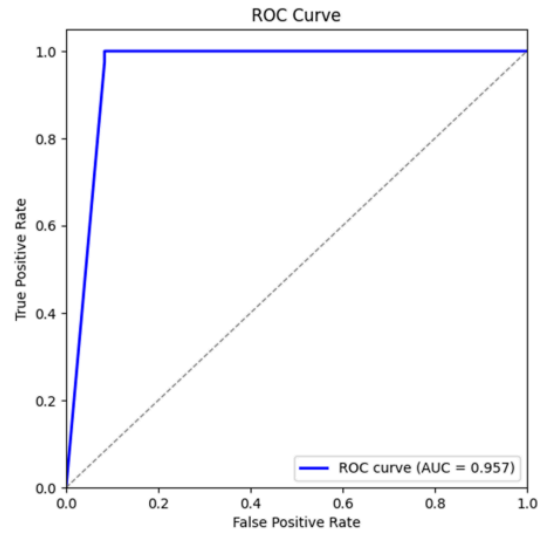
The figures above present the ROC curves of the model in diagnostic tasks after removing a specific module. Panels (a) and (b) correspond to the diagnostic task (International) and diagnostic task (China), respectively, after removal of the image module. The AUC values are 0.904 and 0.940, respectively. Compared with the multimodal model, the decreases in AUC are minor, at 0.021 and 0.004. Panels (c) and (d) represent the diagnostic task (International) and diagnostic task (China), respectively, after removal of the tabular module. The corresponding AUC values are 0.624 and 0.661. In comparison with the multimodal model, the AUC values are reduced by 0.301 and 0.283, respectively.

AUC Curves for External Validation

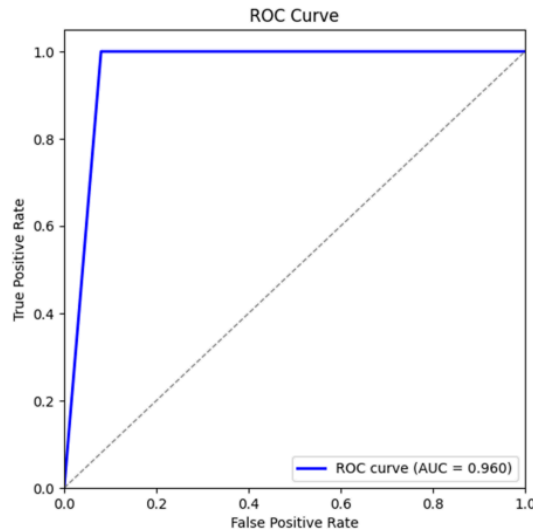
(a) Early-Warning Task (International Criteria)



(b) Early-Warning Task (Chinese Criteria)



(c) Diagnostic Task (International Criteria)



(d) Diagnostic Task (Chinese Criteria)

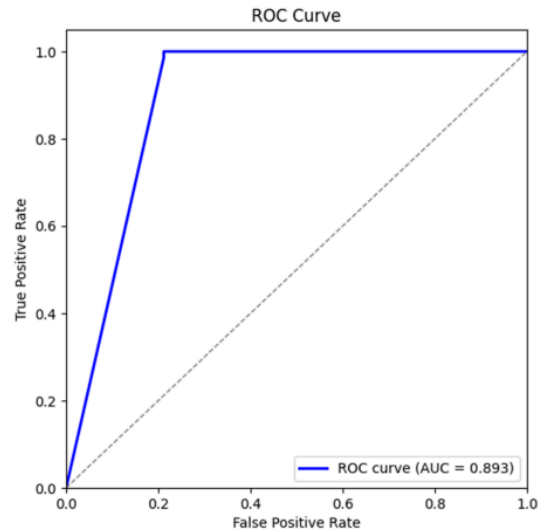


Figure 6

The figure above presents the AUC curves of the model across four tasks in external validation. Panels (a) and (b) display the results of the early-warning tasks according to international standards and Chinese standards, with AUC values of 0.910 and 0.957, respectively. Panels (c) and (d) illustrate the results of the diagnostic tasks according to international standards and Chinese standards, with corresponding AUC values of 0.960 and 0.893. These findings indicate that the model exhibits strong robustness.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTables.docx](#)