AMERICAN ACADEMY OF OPHTHALMOLOGY®

The Role of Artificial Intelligence in Epiretinal Membrane Care: A Scoping Review

David Mikhail, MD(C), MSc(C),[1,2] *Daniel Milad, MD,*[2,3,4] *Fares Antaki, MD, CM,*[2,4] *Karim Hammamji, MD,*[2,4] *Cynthia X. Qian, MD,*[2,3] *Flavio A. Rezende, MD, PhD,*[2,3] *Renaud Duval, MD, FRCSC*[2,3]

Topic: In ophthalmology, artificial intelligence (AI) demonstrates potential in using ophthalmic imaging across diverse diseases, often matching ophthalmologists' performance. However, the range of machine learning models for epiretinal membrane (ERM) management, which differ in methodology, application, and performance, remains largely unsynthesized.

Clinical Relevance: Epiretinal membrane management relies on clinical evaluation and imaging, with surgical intervention considered in cases of significant impairment. AI analysis of ophthalmic images and clinical features could enhance ERM detection, characterization, and prognostication, potentially improving clinical decision-making. This scoping review aims to evaluate the methodologies, applications, and reported performance of AI models in ERM diagnosis, characterization, and prognostication.

Methods: A comprehensive literature search was conducted across 5 electronic databases including Ovid MEDLINE, EMBASE, Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, and Web of Science Core Collection from inception to November 14, 2024. Studies pertaining to AI algorithms in the context of ERM were included. The primary outcomes measured will be the reported design, application in ERM management, and performance of each AI model.

Results: Three hundred ninety articles were retrieved, with 33 studies meeting inclusion criteria. There were 30 studies (91%) reporting their training and validation methods. Altogether, 61 distinct AI models were included. OCT scans and fundus photographs were used in 26 (79%) and 7 (21%) papers, respectively. Supervised learning and both supervised and unsupervised learning were used in 32 (97%) and 1 (3%) studies, respectively. Twenty-seven studies (82%) developed or adapted AI models using images, whereas 5 (15%) had models using both images and clinical features, and 1 (3%) used preoperative and postoperative clinical features without ophthalmic images. Study objectives were categorized into 3 stages of ERM care. Twenty-three studies (70%) implemented AI for diagnosis (stage 1), 1 (3%) identified ERM characteristics (stage 2), and 6 (18%) predicted vision impairment after diagnosis or postoperative vision outcomes (stage 3). No articles studied treatment planning. Three studies (9%) used AI in stages 1 and 2. Of the 16 studies comparing AI performance to human graders (i.e., retinal specialists, general ophthalmologists, and trainees), 10 (63%) reported equivalent or higher performance.

Conclusion: Artificial intelligence–driven assessments of ophthalmic images and clinical features demonstrated high performance in detecting ERM, identifying its morphological properties, and predicting visual outcomes following ERM surgery. Future research might consider the validation of algorithms for clinical applications in personal treatment plan development, ideally to identify patients who might benefit most from surgery.

Financial Disclosure(s): The author(s) have no proprietary or commercial interest in any materials discussed in this article. *Ophthalmology Science 2025;5:100689*

Supplemental material available at www.ophthalmologyscience.org.

In 1865, Iwanoff first described epiretinal membrane (ERM) as a semitranslucent proliferation of cellular tissue on the retina's inner surface overlying the internal limiting membrane (ILM), often at the macula.[1] Epiretinal membranes consist of an outer layer of noncellular extracellular matrix proteins and an inner epiretinal cell layer, which becomes more contractile due to an accumulation of myofibroblast-like cells and extracellular matrix deposits.[2] Epiretinal membranes affect central vision, with symptoms including metamorphopsia and lowered visual acuity.[3] Symptoms often worsen over time due to progressive thickening of the fibrous tissue. Epiretinal membranes are most commonly idiopathic, whereas some can be secondary to trauma, surgery, or another retinal disorder such as diabetic retinopathy or retinal vein occlusion. This condition commonly affects patients >50 years of age. Prevalence is reported to be between 7% and 11%, increasing to approximately 20% in patients >75 years of age.[1,2] Symptomatic ERMs are primarily managed surgically, requiring a pars plana vitrectomy with ERM and ILM peeling.[1]

Artificial intelligence (AI) is a field within computer science dedicated to designing intelligent machines.[4]

https://doi.org/10.1016/j.xops.2024.100689
ISSN 2666-9145/25

Machine learning (ML) and deep learning (DL) models applied to radiology, dermatology, pathology, and ophthalmology have matched or exceeded human performance on specifically defined tasks.[4] In ophthalmology, AI tools are primarily applied to diagnostic imaging, capable of using fundus photographs, OCT scans, and visual fields. The detection of diabetic retinopathy, glaucoma, and age-related macular degeneration are common applications of AI models that process images.[5–7] Models are developed using existing datasets of images, which are either manually labeled by domain experts for ML-based models or are used "end-to-end" without labeling by DL-based models, although models built using unsupervised pretraining still require certain labeled examples as a baseline.[4] These models are then validated with test datasets, and their performance can be compared with a reference standard (e.g., manual grading of images by an ophthalmologist).

Some studies also show AI's capability to make predictions of visual recovery or other postoperative outcomes based on clinical features, such as, for example, visual acuity, age, prior surgeries, and lens status.[8] Thus, the applications of AI in managing any ophthalmic disorder are highly diverse, with each model being optimized to handle a specific task. Given that the diagnosis of ERM largely depends on OCT findings, AI-assisted tools would ideally be applied to predict which patients may benefit most from ERM surgery in the near future.[1]

This scoping review examines the current literature on AI applications at each stage of ERM management, from diagnosis to treatment. This study reports the diverse approaches to AI development and application in ERM management, specifically considering each model's training and validation methodology, role in ERM care, performance, strengths, and limitations.

Methods

This scoping review was reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews guidelines. A protocol was prospectively registered with Open Science Framework (osf.io/g8tqb), since the International Prospective Register of Systematic Reviews is reserved for systematic reviews. Ethics approval from our institutional review board was not required since this study utilized publicly available published data. Given the high variability in AI development methodologies, applications in ERM care, and performance metrics, a scoping review was undertaken. On searching the literature for existing scoping and systematic reviews, there have been no similar studies compiling the literature on this topic to the authors' knowledge. This scoping review will thus summarize existing literature that leverage AI using ophthalmic imagery or clinical characteristics at any stage of ERM care.

Search Strategy and Eligibility Criteria

Five electronic databases, including Medline, Embase, Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, and Web of Science Core Collection, were searched from inception to November 14, 2023. Our search strategy is found in Appendix 1 (available at www.ophthalmologyscience.org). Neither study type nor language restrictions were placed on the search. The reference list of each included study was manually searched to identify articles that were missed in the original search. Key terms relating to or describing ERM were included, such as macular pucker, preretinal macular fibrosis, surface wrinkling retinopathy, epimacular proliferation, epiretinal fibrosis, epiretinal gliosis, and cellophane maculopathy.

Selection Criteria and Data Collection

The selection criteria included (1) original investigations, including peer-reviewed studies and preprints; (2) applied AI usage of ophthalmic images or clinical characteristics at any stage of ERM management; and (3) study population of any age with any type of ERM and any comorbidity. Study exclusion criteria were (1) study populations of nonhuman participants (e.g., animal studies), postmortem samples, or enucleated eyes; (2) conference abstracts, opinion pieces, reviews, systematic reviews, and meta-analyses; and (3) studies only using regression analyses. We included studies utilizing AI for computer vision, those using AI with structured tabular data, or a combination of the two. However, regression analysis, a traditional ML tool, was omitted from our search because of its ubiquity in modern research.

Articles included in the search were exported to Covidence (Veritas Health Innovation) for screening. Two independent reviewers (D. Mikhail and D. Milad) screened the titles and abstracts of each study. Full-text screening of the retained manuscripts was then performed to determine eligibility for inclusion. Any conflicts were resolved by consensus. Screened and included studies are illustrated using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews flowchart (Fig 1). The data were extracted onto a standardized form on an Excel spreadsheet.

Synthesis of Evidence

Descriptive synthesis of the evidence found in the included studies was conducted. Study and patient characteristics were first summarized. Subsequently, a thorough account of each AI model's development and validation methodologies was included. This is followed by a report of the characteristics of each model, such as the type of AI, the input data (i.e., image or clinical data), the model's role in ERM management, and performance measurements. The reported performance of the AI model relative to the reference standard was also included, considering percent accuracy, area under the curve, sensitivity, specificity, precision, dice similarity coefficient, correlation, and F-measure (F1) scores. Given the lack of standardization in AI outcome reports, quantitative statistical analysis could not be performed. A formal risk of bias assessment was not performed.[9]

Results

Study and AI Characteristics

A total of 390 articles were obtained from the literature search, and after the removal of duplicates, 240 studies remained. From these, there were 33 studies that met the inclusion criteria (Fig 1). The studies' characteristics are reported in Table 1. The studies were conducted in China (36%), Spain (15%), Taiwan (12%), South Korea (12%), Japan (9%), and Portugal (6%). There were 23 studies (70%) that did not specify the type of ERM studied, 5 (15%) focused on idiopathic ERM only, 4 (12%) classified ERM based on size, presence of retinal

Table 1. Characteristics of Included Studies

Primary Author, Publication Year	Journal of Publication	Country of Publication	Type of ERM, if Specified	Total Amount of Data	Training Sample Size (Images)	Validation Sample Size (Images)	Testing Sample Size (Images)	Number of Patients	Number of Eyes	Mean Age (±SD)	Sex n (% Male)
Ayhan, 2024	*Nature Scientific Reports*	Germany	Small ERM (100−1000 µm) Large ERM (>1000 µm)	624 OCT volume scans 11 061 OCT images	8341	1146	1574	461	624	69.5 (95% CI: 67.3−71.7)	NR
Baamonde, 2019	*Biomedical Optics Express*	Spain	ERM	285 OCT images	NR	NR	285	NR	NR	NR	NR
Baamonde, 2017	*Advances in Computational Intelligence*	Spain	ERM	129 OCT images, from which random samples are taken	30	30	30	NR	NR	NR	NR
Baamonde, 2017	*Image Analysis and Processing*	Spain	ERM	129 OCT images	NR	NR	120	NR	NR	NR	NR
Bai, 2022	*Frontiers in cell and developmental biology*	China	ERM	1311 OCT images	787	262	262	439	878	53.2 (17.1)	213 (48.5)
Chen, 2023	*International Ophthalmology*	China	iERM	72 CFP	24	24	24	51	NR	59.7 (6.0)	17 (31.3)
Chen, 2023	*Translational Vision Science & Technology*	China	ERM	37 138 OCT images	27 210	6802	3126	775	NR	NR	334 (43.1)
Choi, 2024	*BMC Medical Informatics and Decision Making*	Korea	ERM	1552 CFP	1239	313	RFMiD Database: 695 JSIEC Database: 64	NR	1552	NR	NR
Crincoli, 2023	*Retina*	Italy	iERM	1233 OCT images	828	180	225	411	NR	71 (7.9)	195 (47.4)
Dong, 2022	*Jama Network Open*	China	ERM	CFP	103 970	20 794	209 967 (4031 ERM cases)	110 784	NR	44 (median)	29 370 (46.3)
Gende, 2022	*Computerized Medical Imaging and Graphics*	Spain	ERM	20 OCT volume scans 2427 OCT images	303	152	152	20	20	NR	NR
Gende, 2021	*Institute of Electrical and Electronics Engineers*	Spain	ERM	20 OCT volume scans 2428 OCT images	304	152	152	20	20	NR	NR
Hsia, 2023	*Asia-Pacific Journal of Ophthalmology*	Taiwan	ERM	600 OCT images	540 (training and validation)	540 (training and validation)	60	511	NR	NR	NR
Hung, 2023	*Methods*	Taiwan	ERM	1947 OCT images	1630 (training and validation)	1630 (training and validation)	317	NR	NR	NR	NR
Irie-Ota, 2024	*PLoS One*	Japan	ERM	Model 1: 22 preoperative features Model 2: Postoperative VA at 1, 3, and 6 months	NA	NA	NA	67	67	69.03 (7.39)	36 (53.7)
Jin, 2023	*Journal of Clinical Medicine*	China	ERM	4547 OCT images	Segmentation: 210 Classification: 3220	Segmentation: 30 Classification: 460	Segmentation: 60 Classification: 387	1046	1593	NR	NR

(*Continued*)

Table 1. (Continued.)

Primary Author, Publication Year	Journal of Publication	Country of Publication	Type of ERM, if Specified	Total Amount of Data	Training Sample Size (Images)	Validation Sample Size (Images)	Testing Sample Size (Images)	Number of Patients	Number of Eyes	Mean Age (±SD)	Sex n (% Male)
Kim, 2021	*Journal of Personalized Medicine*	Korea	ERM	628 CFP	502	126 (validation and testing simultaneous using cross-validation)	126 (validation and testing simultaneous)	99	628	63.6 (7.6)	53 (53.5)
Kim, 2022	*Retina*	Korea	iERM	657 OCT images	460	98	99	657	688	65.1 (7.9)	190 (27.6)
Kuwayama, 2019	*Journal of Ophthalmology*	Japan	ERM	1200 OCT images	12	1100 (validation and testing simultaneous)	1100 (validation and testing simultaneous)	300	600	NR	NR
Lee, 2021	*Electronics*	Korea	ERM	43 221 CFP	33 894	9332	11 707	43 227	25 564	53.4 (10.97)	15 594 (61)
Li, 2022	*British Journal of Ophthalmology*	China	ERM	64 914 CFP	45 390	11 348	8176	5950	11 900	53.4	NR
Lo, 2020	*Scientific Reports*	Taiwan	ERM	3618 OCT images	3141	3141 (training and validation)	477 (training and validation)	964	1475	NR	NR
Lu, 2018	*Translational Vision Science & Technology*	China	ERM	25 134 OCT images	22 017	22 017 (training and validation)	3117 (training and validation)	NR	NR	NR	NR
Parra-Mora, 2021	*Institute of Electrical and Electronics Engineers*	Portugal	ERM	2160 OCT images	1404	156	600	608	NR	NR	NR
Parra-Mora, 2022	*Computers in Biology and Medicine*	Portugal	ERM	3101 OCT images	ERM Dataset: 126 AROI Dataset: 947 HCMS Dataset: 833	AROI: 189	ERM: 124 HCMS: 882	NR	NR	NR	NR
Shao, 2021	*Scientific Reports*	China	ERM and secondary ERM	229 239 CFP	207 228	21 867	144	96	192	69.2	48 (50)
Sonobe, 2019	*International Ophthalmology*	Japan	ERM	529 3D-OCT images	423 (training and validation)	423 (training and validation)	106	389	NR	70.1 (8.2)	79 (38.5)
Tang, 2022	*Ophthalmic Research*	China	ERM	469 OCT images	422	46	NR	404	468	62	183 (45.3)
Touma, 2024	*BioMed Central International Journal of Retina and Vitreous*	Canada	ERM	1173 OCT images (207 ERM)	1055 total (70% training 30% validation) (186 ERM)		118 (21 ERM)	NR	NR	NR	NR
Wang, 2020	*Translational Vision Science & Technology*	China	ERM	28 664 OCT images	11 987	2997	EENT Dataset: 7648 TENTH Dataset: 6032	2254	EENT: 956 TENTH: 754	NR	EENT: 290 (49.5) TENTH: 240 (51.6)
Wen, 2023	*BMC Ophthalmology*	China	iERM	5304 OCT images	3180	1068	1056	NR	442	66.2 (8.42)	31 (35.72)

Table 1. (Continued.)

Primary Author, Publication Year	Journal of Publication	Country of Publication	Type of ERM, if Specified	Total Amount of Data	Training Sample Size (Images)	Validation Sample Size (Images)	Testing Sample Size (Images)	Number of Patients	Number of Eyes	Mean Age (±SD)	Sex n (% Male)
Yan, 2023	*Eye*	China	Stage 0: Normal retina Stage 1: Normal FP and retinal layers Stage 2: Loss of FP. Stage 3: No FP; continuous EIFL Stage 4: Traction disrupting retinal layers Stage 5: Traction causing lamellar or pseudoholes	3953 OCT images	3653 (training and validation)	3653 (training and validation)	300	NR	1921	NR	NR
Yeh, 2023	*Retina*	Taiwan	iERM	1058 OCT images	956 (training and validation)	956 (training and validation)	102	529	529	67.74 (10.24)	240 (45.37)

CI = confidence interval; CFP = color fundus photograph; EENT = Eye and ENT Hospital of Fudan University; EIFL = ectopic inner foveal layer; ERM = epiretinal membrane; FP = foveal pit; iERM = idiopathic epiretinal membrane; NR = not reported; *TENTH* = Shanghai Tenth People's Hospital.

disruptions, and membrane staging, and 1 (3%) included secondary causes of ERM. Twenty-six (79%) of the studies employed AI that solely used images, 5 (15%) used images and clinical features, including visual acuity and symptom duration, 1 (3%) used images and videos, and 1 (3%) used only clinical features.

Thirty of the studies (91%) included an in-depth report of their AI training and validation methodology. In total, 61 algorithms were included. Table 2 summarizes the function, strengths, and limitations of common AI algorithms included in this review. Table S3 (available at www.ophthalmologyscience.org) summarizes the methods and performance of each algorithm in the included studies. All models underwent specialized training protocols with training images. Supervised learning was used in 32 studies (97%), and a combination of both supervised and unsupervised learning was used in 1 study (3%).

Supervised and Unsupervised Learning

Of the algorithms trained via supervised learning, artificial neural networks were most common. Of these, there were 41 (70%) convolutional neural networks (CNNs). Other algorithms utilizing supervised learning included 4 (7%) random forests, 3 (5%) Naive Bayes classifiers, 3 (5%) feed-forward neural networks, 2 (3%) k-nearest neighbors, and 2 (3%) support vector machines.

There was 1 (2%) algorithm trained via unsupervised learning, namely principal component analysis. These tools detected the presence of ERM, either through binary or multiclass classification to distinguish ERM cases from normal cases or other retinal pathologies, respectively. Some multiclass classification also distinguished between different presentations of ERM (e.g., attached or detached from the retina).

Epiretinal Membrane Data and Management

Included studies were categorized by their algorithm's role in ERM care, defined in 3 stages: (1) diagnosis; (2) identification of ERM characteristics; and (3) postoperative prognosis of anatomical or visual recovery or progression. Included studies were subcategorized into whether their model used ophthalmic images or clinical characteristics as input data.

The imaging modalities used as input data were OCT scans (26; 79%) and fundus photographs (7; 18%). In 4 (12%) studies, OCT parameters such as the presence of ectopic inner foveal layer, cotton wool sign, foveal detachment, ellipsoid zone interruption, fibrillary changes, central macular thickness (CMT), the position of the ILM, the ERM size, and a scan of the patient's preoperative macular center were assessed manually.[10–13] In 1 (3%) study, the optical density and optical density ratio of fundus photographs were manually extracted and fed to the model.[14] In 5 (15%) studies, OCT parameters were classified automatically.[15–19] Five (15%) of studies utilized models that used images and clinical data.[13,14,20–22] One study used OCT images and videos.[23] Another used clinical data alone.[24] Clinical features included ERM thickness, patient demographics (e.g., age and sex),

Table 2. Tasks, Advantages, and Disadvantages of the Most Common Artificial Intelligence Algorithms in Epiretinal Membrane Management

AI Algorithm	Algorithm Tasks	Type of AI	Advantages	Disadvantages
CNN	Subset of ANNs that automates feature extraction from input data. Contains convolutional, pooling, and FC layers. Layers become increasingly complex and identify hierarchical feature representations from image pixels until the object of interest is recognized.	Supervised	• Highly accurate in image recognition and classification • Automatic extraction of features improves efficiency and training time • Able to leverage transfer learning to cater premade models trained on large datasets for particular tasks with smaller datasets	• Require computational power and large amounts of labeled data for training • Without large datasets, prone to overfitting, and requires data modification (e.g., manually editing OCT images to generate more data) • Interpretability of model's decision-making is made challenging due to the "black box" phenomenon • Vulnerable to adversarial attacks on input data, which can lead to incorrect predictions; raises concerns about privacy and reliability, especially in clinical settings
FCN	Subset of ANNs that augment CNNs by changing the FC layers with convolutional layers. These layers enable pixel-by-pixel classification to identify objects and object boundaries inside images.	Supervised	• Flexibility in using images of any size and of different resolutions • Able to precisely segment and localize objects within images via spatial maps • Require fewer parameters than CNNs, increasing efficiency and speed	• Computationally demanding to train, requiring plenty of diverse datasets • Struggle to distinguish between similarly shaped objects, requiring further information • Challenging to integrate domain expertise (e.g., ophthalmology-specific parameters) to improve performance, requiring custom adjustments
SVM	ML algorithms are used in binary classification and regression. Classifies objects by identifying the clearest boundary or point of difference that separates them into different categories (i.e., hyperplane). Note: SVMs generally process structured data.	Supervised	• Effective in high-dimensional spaces with many features (thus suitable for image recognition) • Manage nonlinear data (along with linear) via kernel functions • Less sensitive to outliers and generalizable to different datasets • Less susceptible to overfitting	• Performance is kernel function dependent (difficult to predict which function the model will choose) • Inefficient multiclass classification • Less effective if the data are "noisy" or contain missing values
RF	Constructs a group of decision trees trained on random parts of the dataset to make predictions. Outputs mean (regression) or mode (classification) of tree predictions. Ensemble learning enables combination of multiple model predictions to improve accuracy and minimize overfitting Note: RFs generally process structured data.	Supervised	• Accurate and versatile • Capable of handling missing or complex data • Ensemble learning boosts robustness and generalizability to different datasets	• Decision-making is based on combined classifications of hundreds to thousands of trees, making interpretability challenging • Struggles with new data that does not fit into a category already learned in the training data • Biased toward highly representative classes in the data

Table 2. (Continued.)

AI Algorithm	Algorithm Tasks	Type of AI	Advantages	Disadvantages
PCA	Reduces data dimensionality while maintaining its variability. Targets variables that maximize data variance (i.e., principal components) and transforms the data into uncorrelated variables categorized by level of importance Note: PCAs generally process structured data.	Unsupervised	• Improves efficiency and reduces noise in the data • Incorporated into ML models to improve performance • Data in lower dimensions enable simpler identification of the underlying data structure and relationships between variables	• Assumes that principal components are linearly related to the original features, which ignores nonlinear relationships between the data • Dimensionality reduction leads to inevitable loss of information, which may negatively influence the performance of the accompanying ML model

AI = artificial intelligence; ANN = artificial neural network; CNN = convolutional neural network; FC = fully connected; FCN = fully convolutional networks; ML = machine learning; PCA = principal component analysis; RF = random forest; SVM = support vector machine.

preoperative best-corrected visual acuity (BCVA), symptom duration, retinal vascular oxygen saturation, inner and outer segment junction (IS/OS line) integrity, foveolar detachment, foveal detachment, and CMT. Of these 5 studies, 2 (40%) applied AI to ERM diagnosis (stage 1), and 3 (60%) designed their models to predict postoperative outcomes. In total, 23 studies (70%) were classified under stage 1 of ERM management, 1 (3%) in stage 2, and 6 (18%) in stage 3. In addition, 3 (9%) studies implemented AI in stages 1 and 2. Thirteen studies (39%) utilized AI models capable of multiclass classification, which distinguished between different retinal disorders or graded the severity of the ERM.[10,14,16–20,23,25–33]

Artificial Intelligence Performance

Each study reported its AI model's performance quantitatively. There were 30 (91%) studies that performed preliminary training and internal validation on their models and reported the specific amount of data used. Three (9%) studies stated that their model was trained but did not specify the data used in their training set.[11,12] Nine (27%) papers externally validated their model, and each reported their model's performance as shown in Table S3.

Twenty-two (67%) studies completed comparative analyses on their model seen in Table S4 (available at www.ophthalmologyscience.org). Six (18%) compared their model either to identical models at different hyperparameters or to different algorithms altogether. Each study reported equivalent or higher performance in their model compared to the AI comparator.[12,14,19,23,32,33] None of the studies confirmed the significance of observed differences between the models' AUCs by performing the DeLong test.[34]

The remaining 12 (33%) papers performed additional comparisons between their model and human graders. Graders were retina specialists, general ophthalmologists, ophthalmology trainees, medical students, or were generally categorized as clinician experts. In these comparisons, the human graders did not establish the ground truth. Ten (83%) of these studies reported that their AI model performed comparably to or outperformed the human graders. The remaining 2 (17%) reported mixed results, with the AI model outperforming only some of the human graders.[14,35] Four additional papers (13%) used human graders to establish the ground truth and noted inferior performance of their model.[10,16,28,29] Four (13%) studies used the Cohen Kappa statistic to measure interrater reliability between the human comparators and the model.[18,25,31,36]

Discussion

The majority of AI models implemented in ERM management were CNNs trained via supervised learning. Supervised learning is an ML approach that uses labeled training data, typically annotated by domain experts, to train models by learning to map inputs to their corresponding ground truth labels.[37,38] These models have had numerous applications in ophthalmology, including detecting geographic atrophy from OCT scans.[39,40] Generally, the

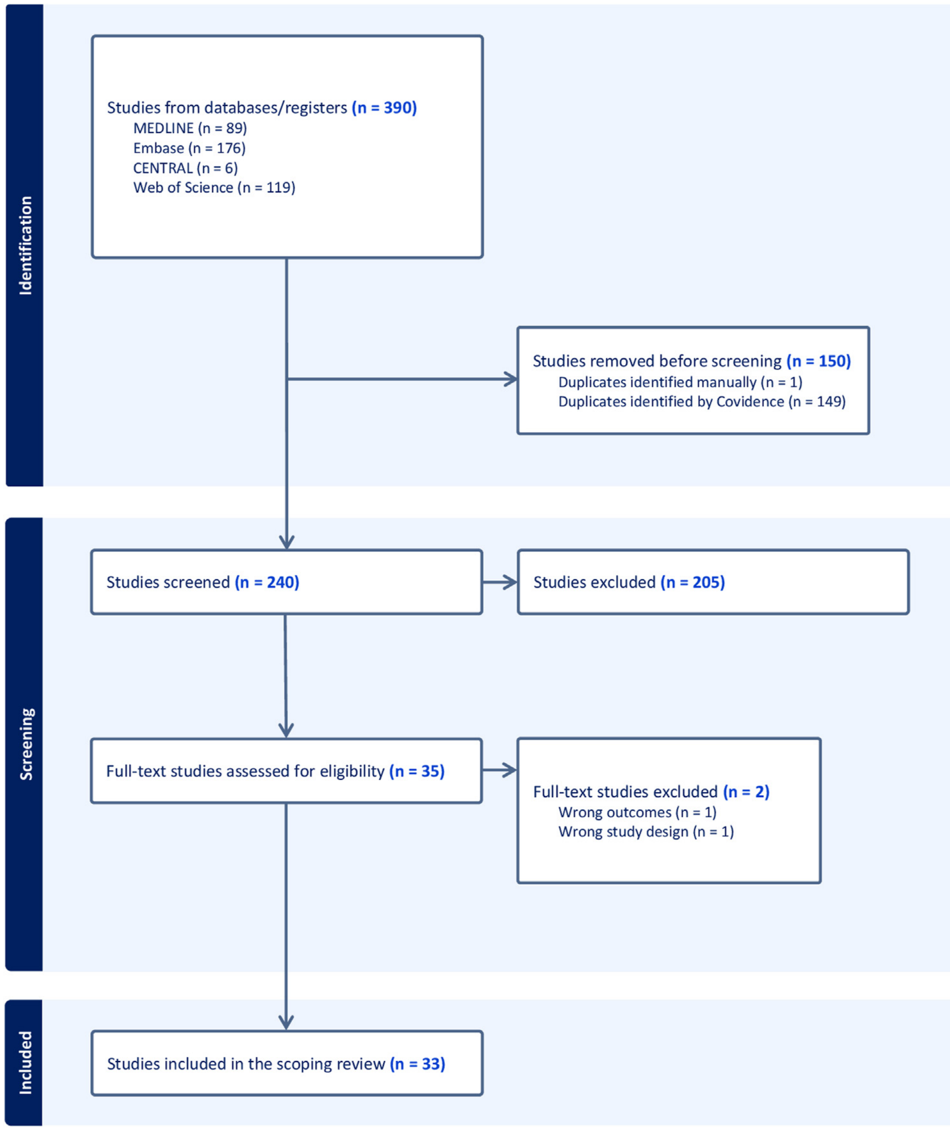

Figure 1. PRISMA-ScR flowchart diagram. PRISMA-ScR = Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews.

models included in this review used OCT scans or fundus photographs for either binary or multiclass ERM diagnosis (stage 1). Clinical features were used solely by 1 (3%) study or in conjunction with ophthalmic imaging in 5 (15%) studies.[13,20–22,24,41] These models were most often designed to predict postoperative anatomical or functional vision outcomes. Since winning the ImageNet Large Scale Visual Recognition Competition in 2012, CNNs have rapidly advanced to become the state-of-the-art DL algorithm for computer vision.[42–44] Convolutional neural networks are distinguished from traditional ML models by their ability to identify essential image features without requiring manual feature extraction by humans.[45] Convolutional neural networks can be catered for specific applications by modifying the depth and organization of their layers, the datasets on which they are trained, and the methods employed to reduce overfitting.[46] These separate variations are also known as "architectures." In this scoping review, the most common CNN architecture was ResNet, which is a deep network capable of complex tasks such as medical image processing.[47] In medical image analysis, CNNs have shown potential in classification, detection, and segmentation of a wide range of images in various specialties.[43,48]

Most (91%) of the studies outlined how they trained, validated, and tested their model. All the studies except 3 reported the specific number of images used for model training. Since some models were designed for multiclass classification, training images consisted of several ocular pathologies, and the distribution of ERM-specific images was not always clarified. Given that AI models for ERM are dependent on large quantities of high-quality images depicting the pathology, it is essential to document and report the number of ERM-specific images on which the model was trained. Moreover, only 9 (27%) papers performed external validation. Of these, 7 (78%) recorded performance metrics on the external dataset. One paper reported external validation but did not report their model's performance on the external dataset.[18] In ML model development, external validation involves testing the model on new data from different sources compared to the data used in training and internal validation.[49] External validity is an essential tool used to confirm the generalizability of a model to new data and its applicability to specific fields of study.[49] Despite the increasing use of external validation in prediction model development, it remains currently underutilized in prediction model studies.[50] In addition, external validity as an indicator for ML model generalizability has been debated, with some researchers maintaining that validity must be continuously assessed over time to ensure robust performance.[51]

Each study recorded quantitative measurements of its model's performance. The most common metrics were accuracy, area under the curve, sensitivity, and specificity. Since this study focused on model performance assessing ERM datasets, some studies only reported their model's overall performance on multiple datasets, which was noted in Table S3. In general, the models with higher performance (achieving accuracies >90%) were CNNs. Deep learning−based models like CNNs were also advantageous in bypassing several time-consuming steps required for classical ML models to segment OCT images of the ERM.[12] These steps included preprocessing and manual feature extraction from the images.

Some studies developed more than 1 model, performing comparative analyses between models or benchmarking their model against state-of-the-art models identified in prior research. Several trends in these comparisons were observed. Firstly, ensemble models tended to outperform individual models. One study combined 6 CNN architectures to create a high-performing ensemble model.[31] Ensemble learning is a ML technique that leverages the power of multiple AI algorithms to make more accurate predictions than any single model.[52] Another article combined their CNN ensemble model with a Swin Transformer, which improved model performance.[35] Unlike typical transformers used in natural language processing, Swin Transformers enhance image processing through a unique self-attention mechanism with shifting windows.[53,54] These processes enable models to efficiently scale to large images and datasets and capture those images both in smaller patches and as a whole.[54]

Secondly, the performance of multimodal models tended to be lower than models using images alone. Multimodal models must seamlessly fuse ophthalmic imaging data with key clinical features into a single feature set.[55] Wen et al[21] used a multimodal deep fusion network to combine spectral domain OCT image data with clinical variables such as age, preoperative BCVA, symptom duration, foveolar detachment, and CMT. Despite noting improved predictive power of their model, reported performance was lower than models used exclusively for diagnostic purposes. When comparing performance between different AI models, it is thus critical to consider the model's task, the data used, and often the type and architecture of the model.

Of the 16 studies comparing AI to human graders, 10 (63%) reported comparable or superior performance by their model. Artificial intelligence models and human comparators were tested on each stage of ERM management. Specific tasks included detection of ERM (stage 1), identification of specific attributes of the ERM (stage 2), and postoperative prognoses (stage 3). Artificial intelligence models were also used by graders as an assistive tool, and grader performance was compared with and without AI assistance.[56,57] In both studies, human grader accuracy in identifying ERM presence and classifying severity improved with AI assistance. It was reported that junior clinicians' improvements in accuracy were greater than senior clinicians' improvements. These results indicate likely applications for AI-assisted management in ophthalmology education, both for trainees and ophthalmologists alike. Notably, these studies reported that AI-assisted human performance remained comparable to or was surpassed by AI models operating independently.[56,57]

The International Vitreomacular Traction Study Group proposed specific classifications for many disorders of the vitreomacular surface; however, an ERM classification system has not yet been proposed.[58,59] Although clinical studies have made independent ERM classifications, not all have been validated, nor have they been shown to be clinically useful.[59−61] The integration of AI detection and classification models may be used to validate existing ERM classification systems, confirming perceived differences between ERM classes. Existing AI models with reliable ERM detection and the capacity of staging disease progression via integrated segmentation and classification networks may prove useful for this task.[10,62] Moreover, multimodal models integrating image and clinical data may identify new associations between ERM morphology and corresponding patient data that could prompt innovative and robust classification systems.

Beyond ERM classification, AI models may help predict the outcomes of surgical treatment for this disease. Currently, ERMs are managed surgically, most often via pars plana vitrectomy, ERM peeling, and ILM peeling.[11,63] However, approximately 10% to 20% of patients with idiopathic ERM will have either unchanged or worse visual acuity postoperatively.[21,64] Thus, there is a need for improved prognostic prediction of ERM surgery preoperatively. In this scoping review, 6 (18%) of the studies leveraged AI to predict visual outcomes following ERM surgery.[11,13,21,22,24] Crincoli et al[11] developed a DL model that used the presence of fibrillary changes along with known risk factors from spectral domain OCT scans to accurately predict 1-year postoperative BCVA. Kim et al[13] built a multimodal DL model capable of integrating patient demographic factors, comorbid diseases, and specific surgical techniques into its prediction of postoperative central foveal thickness and central foveal thickness changes. These predictions were used to guide the choice of combined vs. step-by-step phacovitrectomy for ERM.[65] Predictions were performed with and without clinical data consideration. Models using both clinical information (e.g., age, sex, and presence of hypertension or diabetes mellitus) and surgical information (i.e., whether ILM peeling was performed) outperformed models using OCT scans alone.[13] However, the model was noted to be unreliable in predicting final BCVA and BCVA changes, potentially due to insufficient patient data and clinical variables. Accurate visual outcome predictions may offer patients reasonable expectations regarding potential visual improvements and risks following surgery, facilitating informed patient decision-making.

There is also debate over the timing of ERM removal, with conflicting findings that both earlier and delayed surgery yield a better final BCVA.[64] The potential for accurate postoperative prognosis may enhance our understanding of optimal timing for surgery and may even guide ophthalmologists' choice of surgical technique. Hsia et al[65] developed a DL model capable of discerning the severity of visual impairment from ERM. Their model was able to determine the degree of visual impairment caused by ERM in patients with ERM and visually significant cataracts to guide the choice of single or combined surgery on individual OCT images. Given that ERMs may be complicated by other ocular comorbidities, the choice of performing advanced surgical procedures or combined procedures is difficult to predict preoperatively. Deep learning models are capable of considering numerous risk factors, clinical variables, and data than ordinary prognostic tools.

Foundation models are large AI models based on deep neural networks and self-supervised learning. Seminal work by Zhou et al[66] used RETFound, a foundation model, designed for generalizable disease detection from retinal images. RETFound was trained on 1.6 million unlabeled retinal images using self-supervised learning and subsequently fine-tuned with a much smaller set of labeled data to accurately detect both ocular pathologies and complex systemic disorders from these images. This model outperformed comparator models, which required considerable annotations from experts and were designed with highly specific applications. Foundation models could be adopted to efficiently design models trained on large swaths of data with specific downstream applications, such as the diagnosis and prognosis of ERM.

Although the majority of models used supervised learning, none of the studies were based on reinforcement learning. Reinforcement learning is a learning strategy using feedback in the form of rewards or penalties, allowing the model to refine its performance through iterative learning to achieve a particular goal. Reinforcement learning could enhance all aspects of ophthalmic surgery by optimizing surgical timing and enabling intraoperative automation, which could range from warning systems to autonomous robotic procedures.[67] The management of ERMs is clinically challenging, requiring consideration of patient vision, subjective complaints, imaging findings, and risks to surgery. Models trained using reinforcement learning may improve surgeons' expertise and safety in managing complex ERM cases.

The most common limitations cited in the included studies were related to data quality, amount, and diversity. Artificial intelligence models are highly reliant on the quality of data used to train them. Poor model performance was attributed to inadequate illumination of OCT scans, which affected accuracy of classification.[15] Further, 1 model misclassified ILM for ERM due to instances when the OCT was aimed perpendicularly at the ILM, creating a brighter region.[15] Variation in image quality often depends on the imaging devices and modalities used by different medical centers. Thus, acquiring data from different patient samples and hospitals often require added image quality control steps. Given the inherent variation in these preprocessing techniques by technicians and ophthalmologists, a standardized protocol is needed. Moreover, insufficient ERM-specific datasets and patient sample diversity during model development led to worse performance, particularly when algorithms classified mild ERMs, which were underrepresented in the training data.[12] Publicly available ERM datasets are rare, according to the literature, compared to diabetic eye disease, glaucoma, and age-related macular degeneration.[68] None of the included studies used publicly available datasets for their training data, but 1 study (3%) used public data for external validation.[69] All other studies used proprietary data, most often acquired via ethics approval from their affiliate hospital. As AI models evolve to support increasingly complex applications in ophthalmology and medicine, the growing demand for high-quality data underscores the need for new, publicly accessible datasets.

Artificial intelligence models, particularly DL systems, suffer from the black box phenomenon, which describes the inherent lack of interpretability of AI model decision-making by humans.[70] Gradient-weighted class activation mapping (Grad-CAM) is one approach to combat the black box problem that highlights the regions on an image most relevant to the model's classification decisions.[35] One study found that the heat maps generated by Grad-CAM confirmed the regions on an OCT used by ophthalmologists to detect ERM.[35] Although Grad-CAM is a commonly used technique to make AI more explainable, its validity in localizing objects in an image has been debated.[71] For instance, Grad-CAM was found to highlight image regions not implicated in its decision-making, prompting the creation of new techniques such as HiResCAM, which may address these shortcomings.[72] Additionally, attention mechanisms, such as image attention and region attention modules, can augment model interpretability by identifying and focusing on the most relevant regions of the input data and prioritizing critical data for classification.[73,74] These mechanisms can surpass traditional methods such as Grad-CAM by integrating interpretability into the model's design.[39]

To conclude, this scoping review has explored the current literature surrounding AI implementation in ERM care. We discussed the most commonly used AI models, different architectures, training and validation strategies, and tasks by 30 included articles. The models' performance metrics were reported and summarized, along with trends noted in their strengths and weaknesses. None of the articles in this review integrated their models clinically to manage ERM. As AI models progress to handle more complex, multimodal data, their clinical utility may expand to offer personalized management plans based on a multitude of factors. Future research should test AI models on larger ERM datasets and expand the current image repository to include ERMs at different stages and severity.

Footnotes and Disclosures

Originally received: May 21, 2024.
Final revision: December 2, 2024.
Accepted: December 16, 2024.
Available online: December 20, 2024. Manuscript no. XOPS-D-24-00157.

[1] Temerty Faculty of Medicine, University of Toronto, Toronto, Canada.

[2] Department of Ophthalmology, University of Montreal, Montreal, Canada.

[3] Department of Ophthalmology, Hôpital Maisonneuve-Rosemont, Montreal, Canada.

[4] Department of Ophthalmology, Centre Hospitalier de l'Université de Montréal (CHUM), Montreal, Canada.

Disclosure(s):

All authors have completed and submitted the ICMJE disclosures form.

The authors have no proprietary or commercial interest in any materials discussed in this article.

HUMAN SUBJECTS: No human subjects were included in this study.

No animal subjects were used in this study.

Author Contributions:

Conception and design: Mikhail, Milad, Antaki, Hammamji, Qian, Rezende, Duval

Data collection: Mikhail, Milad

Analysis and interpretation: Mikhail, Milad, Antaki

Obtained funding: N/A

Overall responsibility: Mikhail, Milad, Antaki, Hammamji, Qian, Rezende, Duval

Abbreviations and Acronyms:

AI = artificial intelligence; **BCVA** = best-corrected visual acuity; **CMT** = central macular thickness; **CNN** = convolutional neural network; **DL** = deep learning; **ERM** = epiretinal membrane; **ILM** = internal limiting membrane; **ML** = machine learning.

Keywords:

Artificial intelligence, Deep learning, Machine learning, Epiretinal membrane.

Correspondence:

Renaud Duval, MD, FRCSC, Department of Ophthalmology, Université de Montréal, 2900 Édouard Montpetit Boulevard, Montreal H3T 1J4, Quebec, Canada. E-mail: renaud.duval@gmail.com.

References

1. Kanukollu VM, Agarwal P. Epiretinal membrane. In: StatPearls. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK560703/; 2024. Accessed February 19, 2024.
2. Fung AT, Galvin J, Tran T. Epiretinal membrane: a review. *Clin Exp Ophthalmol*. 2021;49:289–308.
3. Shao E, Liu C, Wang L, et al. Artificial intelligence-based detection of epimacular membrane from color fundus photographs. *Sci Rep*. 2021;11:19291.
4. Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, et al. Artificial intelligence in retina. *Prog Retin Eye Res*. 2018;67:1–29.
5. Dong L, Yang Q, Zhang RH, Wei WB. Artificial intelligence for the detection of age-related macular degeneration in color fundus photographs: a systematic review and meta-analysis. *eClinicalMedicine*. 2021;35:100875.
6. Yousefi S. Clinical applications of artificial intelligence in glaucoma. *J Ophthalmic Vis Res*. 2023;18:97–112.
7. Lim JI, Regillo CD, Sadda SR, et al. Artificial intelligence detection of diabetic retinopathy. *Ophthalmol Sci*. 2022;3:100228.
8. Artificial Intelligence. American academy of ophthalmology. https://www.aao.org/eyenet/article/artificial-intelligence; 2017. Accessed March 9, 2024.
9. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Inf Libr J*. 2009;26:91–108.
10. Ayhan MS, Neubauer J, Uzel MM, et al. Interpretable detection of epiretinal membrane from optical coherence tomography with deep neural networks. *Sci Rep*. 2024;14:8484.
11. Crincoli E, Savastano MC, Savastano A, et al. New artificial intelligence analysis for prediction of long-term visual improvement after epiretinal membrane surgery. *Retina*. 2023;43:173–181.
12. Gende M, Moura J, Novo J, Ortega M. End-to-end multi-task learning approaches for the joint epiretinal membrane segmentation and screening in OCT images. *Comput Med Imag Graph*. 2022;98:102068.
13. Kim SH, Ahn H, Yang S, et al. Deep learning-based prediction of outcomes following noncomplicated epiretinal membrane surgery. *Retina*. 2022;42:1465.
14. Chen X, Xue Y, Wu X, et al. Deep learning-based system for disease screening and pathologic region detection from optical coherence tomography images. *Transl Vis Sci Technol*. 2023;12:29.
15. Baamonde S, de Moura J, Novo J, et al. Automatic identification and characterization of the epiretinal membrane in OCT images. *Biomed Opt Express*. 2019;10:4018–4033.
16. Baamonde S, de Moura J, Novo J, et al. Feature definition and selection for epiretinal membrane characterization in optical coherence tomography images. In: Battiato S, Gallo G, Schettini R, Stanco F, eds. *Image Analysis and Processing - ICIAP 2017. Lecture Notes in Computer Science*. Springer International Publishing; 2017:456–466.
17. Baamonde S, de Moura J, Novo J, Ortega M. Automatic detection of epiretinal membrane in OCT images by means of local luminosity patterns. In: Rojas I, Joya G, Catala A, eds. *Advances in Computational Intelligence. Lecture Notes in Computer Science*. Springer International Publishing; 2017:222–235.

18. Bai J, Wan Z, Li P, et al. Accuracy and feasibility with AI-assisted OCT in retinal disorder community screening. *Front Cell Dev Biol*. 2022;10:1053483.
19. Parra-Mora E, Cazañas-Gordon A, Proença R, da Silva Cruz LA. Epiretinal membrane detection in optical coherence tomography retinal images using deep learning. *IEEE Access*. 2021;9:99201–99219.
20. Wang L, Wang G, Zhang M, et al. An intelligent optical coherence tomography-based system for pathological retinal cases identification and urgent referrals. *Transl Vis Sci Technol*. 2020;9:46.
21. Wen D, Yu Z, Yang Z, et al. Deep learning-based post-operative visual acuity prediction in idiopathic epiretinal membrane. *BMC Ophthalmol*. 2023;23:361.
22. Yeh TC, Chen SJ, Chou YB, et al. Predicting visual outcome after surgery in patients with idiopathic epiretinal membrane using a novel convolutional neural network. *Retina Phila Pa*. 2023;43:767–774.
23. Touma S, Hammou BA, Antaki F, et al. Comparing code-free deep learning models to expert-designed models for detecting retinal diseases from optical coherence tomography. *Int J Retina Vitr*. 2024;10:37.
24. Irie-Ota A, Matsui Y, Imai K, et al. Predicting postoperative visual acuity in epiretinal membrane patients and visualization of the contribution of explanatory variables in a machine learning model. *PLoS One*. 2024;19:e0304281.
25. Dong L, He W, Zhang R, et al. Artificial intelligence for screening of multiple retinal and optic nerve diseases. *JAMA Netw Open*. 2022;5:e229960.
26. JCM | free full-text | iERM: an interpretable deep learning system to classify epiretinal membrane for different optical coherence tomography devices: a multi-center analysis. https://www.mdpi.com/2077-0383/12/2/400. Accessed March 9, 2024.
27. Kim KM, Heo TY, Kim A, et al. Development of a fundus image-based deep learning diagnostic tool for various retinal diseases. *J Pers Med*. 2021;11:321.
28. Kuwayama S, Ayatsuka Y, Yanagisono D, et al. Automated detection of macular diseases by optical coherence tomography and artificial intelligence machine learning of optical coherence tomography images. *J Ophthalmol*. 2019;2019:e6319581.
29. Electronics | free full-text | development of decision support software for deep learning-based automated retinal disease screening using relatively limited fundus photograph data. https://www.mdpi.com/2079-9292/10/2/163. Accessed March 9, 2024.
30. Li Y, Feng W, Zhao X, et al. Development and validation of a deep learning system to screen vision-threatening conditions in high myopia using optical coherence tomography images. *Br J Ophthalmol*. 2022;106:633–639.
31. Lu W, Tong Y, Yu Y, et al. Deep learning-based automated classification of multi-categorical abnormalities from optical coherence tomography images. *Transl Vis Sci Technol*. 2018;7:41.
32. Parra-Mora E, da Silva Cruz LA. LOCTseg: a lightweight fully convolutional network for end-to-end optical coherence tomography segmentation. *Comput Biol Med*. 2022;150:106174.
33. Sonobe T, Tabuchi H, Ohsugi H, et al. Comparison between support vector machine and deep learning, machine-learning technologies for detecting epiretinal membrane using 3D-OCT. *Int Ophthalmol*. 2019;39:1871–1877.
34. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837–845.
35. Hung CL, Lin KH, Lee YK, et al. The classification of stages of epiretinal membrane using convolutional neural network on optical coherence tomography image. *Methods San Diego Calif*. 2023;214:28–34.
36. Lo YC, Lin KH, Bair H, et al. Epiretinal membrane detection at the ophthalmologist level using deep learning of optical coherence tomography. *Sci Rep*. 2020;10:8424.
37. Sarker IH. Machine learning: algorithms, real-world applications and research directions. *SN Comput Sci*. 2021;2:160.
38. Cunningham P, Cord M, Delany SJ. Supervised learning. In: Cord M, Cunningham P, eds. *Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval*. Springer; 2008:21–49.
39. Shi X, Keenan TDL, Chen Q, et al. Improving interpretability in machine diagnosis. *Ophthalmol Sci*. 2021;1:100038.
40. Elsawy A, Kenan TD, Chen Q, et al. Attention-based 3D convolutional networks for detection of geographic atrophy from optical coherence tomography scans. *Medical Imaging 2023: Image Process*. 2023;12464:764–768.
41. Chen K, Mao J, Liu H, et al. Screening of idiopathic epiretinal membrane using fundus images combined with blood oxygen saturation and vascular morphological features. *Int Ophthalmol*. 2023;43:1215–1228.
42. Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis*. 2015;115:211–252.
43. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights Imaging*. 2018;9:611–629.
44. Nichols JA, Herbert Chan HW, Baker MAB. Machine learning: applications of artificial intelligence to imaging and diagnosis. *Biophys Rev*. 2018;11:111–118.
45. Suganyadevi S, Seethalakshmi V, Balasamy K. A review on deep learning in medical image analysis. *Int J Multimed Inf Retr*. 2022;11:19–38.
46. Alzubaidi L, Zhang J, Humaidi AJ, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data*. 2021;8:53.
47. Xu W, Fu YL, Zhu D. ResNet and its application to medical image processing: research progress and challenges. *Comput Methods Programs Biomed*. 2023;240:107660.
48. Cai L, Gao J, Zhao D. A review of the application of deep learning in medical image classification and segmentation. *Ann Transl Med*. 2020;8:713.
49. Ho SY, Phua K, Wong L, Bin Goh WW. Extensions of the external validation for checking learned model interpretability and generalizability. *Patterns*. 2020;1:100129.
50. Ramspek CL, Jager KJ, Dekker FW, et al. External validation of prognostic models: what, why, how, when and where? *Clin Kidney J*. 2021;14:49–58.
51. Youssef A, Pencina M, Thakur A, et al. External validation of AI models in health should be replaced with recurring local validation. *Nat Med*. 2023;29:2686–2687.
52. Mahajan P, Uddin S, Hajati F, Moni MA. Ensemble learning for disease prediction: a review. *Healthc Basel Switz*. 2023;11:1808.
53. Yan S, Wang C, Chen W, Lyu J. Swin transformer-based GAN for multi-modal medical image translation. *Front Oncol*. 2022;12:942511.
54. Liu Z, Lin Y, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows. *arXiv*. 2021. https://doi.org/10.48550/arXiv.2103.14030.

55. Salvi M, Loh HW, Seoni S, et al. Multi-modality approaches for medical support systems: a systematic review of the last decade. *Inf Fusion*. 2024;103:102134.
56. Tang Y, Gao X, Wang W, et al. Automated detection of epiretinal membranes in OCT images using deep learning. *Ophthalmic Res*. 2023;66:238−246.
57. Yan Y, Huang X, Jiang X, et al. Clinical evaluation of deep learning systems for assisting in the diagnosis of the epiretinal membrane grade in general ophthalmologists. *Eye*. 2024;38: 730−736.
58. Duker JS, Kaiser PK, Binder S, et al. The international vitreomacular traction study Group classification of vitreomacular adhesion, traction, and macular hole. *Ophthalmology*. 2013;120:2611−2619.
59. Stevenson W, Prospero Ponce CM, Agarwal DR, et al. Epiretinal membrane: optical coherence tomography-based diagnosis and classification. *Clin Ophthalmol Auckl NZ*. 2016;10:527−534.
60. Szigiato AA, Antaki F, Javidi S, et al. Risk factors for epiretinal membrane formation and peeling following pars plana vitrectomy for primary rhegmatogenous retinal detachment, an OCT guided analysis. *Int J Retina Vitr*. 2022;8:70.
61. Govetto A, Lalane RA, Sarraf D, et al. Insights into epiretinal membranes: presence of ectopic inner foveal layers and a new optical coherence tomography staging scheme. *Am J Ophthalmol*. 2017;175:99−113.
62. Jin K, Yan Y, Wang S, et al. iERM: an interpretable deep learning system to classify epiretinal membrane for different optical coherence tomography devices: a multi-center analysis. *J Clin Med*. 2023;12:400.
63. Dawson SR, Shunmugam M, Williamson TH. Visual acuity outcomes following surgery for idiopathic epiretinal membrane: an analysis of data from 2001 to 2011. *Eye*. 2014;28: 219−224.
64. Matoba R, Morizane Y. Surgical treatment of epiretinal membrane. *Acta Med Okayama*. 2021;75:403−413.
65. Hsia Y, Lin YY, Wang BS, et al. Prediction of visual impairment in epiretinal membrane and feature analysis: a deep learning approach using optical coherence tomography. *Asia-Pac J Ophthalmol Phila Pa*. 2023;12:21−28.
66. Zhou Y, Chia MA, Wagner SK, et al. A foundation model for generalizable disease detection from retinal images. *Nature*. 2023;622:156−163.
67. Nath S, Korot E, Fu DJ, et al. Reinforcement learning in ophthalmology: potential applications and challenges to implementation. *Lancet Digit Health*. 2022;4:e692−e697.
68. Khan SM, Liu X, Nath S, et al. A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability. *Lancet Digit Health*. 2021;3:e51−e66.
69. Choi JY, Ryu IH, Kim JK, et al. Development of a generative deep learning model to improve epiretinal membrane detection in fundus photography. *BMC Med Inform Decis Mak*. 2024;24:25.
70. Hassija V, Chamola V, Mahapatra A, et al. Interpreting black-box models: a review on explainable artificial intelligence. *Cogn Comput*. 2024;16:45−74.
71. Chattopadhay A, Sarkar A, Howlader P, Balasubramanian VN. Grad-CAM++: generalized gradient-based visual explanations for deep convolutional networks. In: *2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA*. Piscataway, NJ: IEEE; 2018:839−847.
72. Draelos RL, Carin L. Use HiResCAM instead of Grad-CAM for faithful explanations of convolutional neural networks. *arXiv*. 2021. https://doi.org/10.48550/arXiv.2011.08891.
73. Shi X, Xing F, Xie Y, et al. Loss-based attention for deep multiple instance learning. *Proc AAAI Conf Artif Intell*. 2020;34:5742−5749.
74. Shi X, Xing F, Xu K, et al. Loss-based attention for interpreting image-level prediction of convolutional neural networks. *IEEE Trans Image Process*. 2021;30:1662−1675.