Received: 26 January 2026 | Accepted: 1 July 2026

DOI: 10.1111/aos.70205

Acta Ophthalmologica

ORIGINAL ARTICLE

Explainable deep learning self-supervised vision transformer for fundus-based myopia classification

Eirini Maliagkani[1] | **Christos Tsoutsas**[2] | **Nikolaos Papageorgiou**[2] | **Ioannis D. Apostolopoulos**[2] | **Nikolaos Papandrianos**[2] | **Ilias Georgalas**[1] | **Elpiniki Papageorgiou**[2]

[1]1st Department of Ophthalmology, General Hospital of Athens "G. Gennimatas", National and Kapodistrian University of Athens, Athens, Greece

[2]ACTA Lab, Department of Energy Systems, University of Thessaly, Larisa, Greece

Correspondence
Eirini Maliagkani, 1st Department of Ophthalmology, General Hospital of Athens "G. Gennimatas", National and Kapodistrian University of Athens, Leoforos Mesogeion 154, 11527 Athens, Greece.
Email: emaliagk@med.uoa.gr; eirini.maliagani@gmail.com

Abstract

Purpose: To develop and validate a self-supervised vision transformer (ViT) for automated three-class classification of Normal, High Myopia (HM), and Pathologic Myopia (PM) from colour fundus photographs, with anatomically faithful interpretability.

Methods: A DINOv2 self-supervised ViT was fine-tuned using a two-stage transfer-learning protocol with class-balanced sampling. Model robustness was assessed using 10-fold stratified cross-validation, and an ensemble of the best fold-specific checkpoints was evaluated once on a strictly held-out independent test set. Comparative performance was assessed against ResNet-50, VGG-16, and EfficientNet-B3 using paired, two-sided Student's *t*-tests on fold-level macro–F1 scores. Model explainability was examined using a Dual-Rollout Attention approach that fuses early-layer spatial detail with late-layer semantic information.

Results: On the independent test set, the ensemble achieved an overall accuracy of 97.03% (95% CI 96.75–97.32%) and a macro F1-score of 97.08% (95% CI 96.80–97.36%). The DINOv2-based framework outperformed ResNet-50, VGG-16, and EfficientNet-B3 architectures (all $p<0.001$). Dual-Rollout Attention highlighted clinically relevant regions including optic disc morphology, peripapillary atrophy, and areas consistent with chorioretinal thinning.

Conclusion: A self-supervised ViT combined with transformer-specific interpretability enables accurate, robust, and transparent automated classification of high and pathologic myopia from fundus photographs. This approach may support standardized assessment and scalable screening, pending external and prospective validation.

KEYWORDS
deep learning, explainable artificial intelligence, fundus photography, high myopia, pathologic myopia, vision transformer

1 | INTRODUCTION

Myopia represents a major and growing public-health burden and is increasingly recognized as a chronic, progressive disease with cumulative structural damage and long-term ocular morbidity (Pärssinen, 2025; Sankaridurg et al., 2021; Flitcroft et al., 2019). Epidemiologic projections suggest that by 2050 approximately half of the world's population will be myopic and nearly 10% will have high myopia (HM; typically spherical equivalent < −6.00 D), a phenotype that substantially increases the risk of pathologic myopia (PM) and irreversible vision loss (Flitcroft et al., 2019). Clinically, HM is driven primarily by axial elongation and early posterior-pole remodelling, whereas PM is defined by degenerative lesions, including posterior staphyloma, chorioretinal atrophy, and myopic choroidal neovascularization, that signify progression to

a vision-threatening disorder (Benahmed et al., 2025; Flores-Moreno et al., 2025; Ohno-Matsui et al., 2021).

Colour fundus photography remains the most scalable and cost-efficient imaging modality for surveillance and screening. However, distinguishing advanced HM from early PM often relies on subtle, globally distributed cues and is sensitive to image quality and reader expertise, leading to clinically meaningful inter-observer variability in borderline presentations (Du et al., 2021; Ohno-Matsui et al., 2021; Ting et al., 2017).

Deep learning (DL), in particular convolutional neural networks (CNNs), have achieved clinician-level performance for several ophthalmic tasks, including diabetic retinopathy detection from fundus photographs and referral triage using optical coherence tomography (OCT), and have been extended to myopia to support PM detection and myopic maculopathy grading within screening workflows (Du et al., 2021; Ting et al., 2019; Gulshan et al., 2016; Abràmoff et al., 2018; De Fauw et al., 2018). Nonetheless, although deep CNNs can aggregate information over increasingly large image regions as network depth increases, their inherent local inductive bias may limit their ability to capture retina-wide contextual and structural relationships efficiently. Such global information can be important for PM assessment, particularly for distinguishing HM from early PM (Catania et al., 2024; Dosovitskiy et al., 2020; Qi et al., 2025).

Vision Transformers (ViTs) can address this limitation through patch tokenization and multi-head self-attention, enabling joint modelling of fine morphologic detail and long-range spatial dependencies (Dosovitskiy et al., 2020; Shamshad et al., 2023). Because ViTs are typically more data-intensive in the absence of convolutional inductive bias, self-supervised pre-training strategies provide an alternative to creating and training such networks from scratch and improve robustness to domain shift when fine-tuned on curated retinal datasets (Oquab et al., 2024; Touvron et al., 2021; Zhou et al., 2023).

This study developed, validated, and interpreted a DL framework for three-class classification of Normal, HM, and PM from colour fundus photographs using a self-supervised ViT. We benchmarked performance against established CNN architectures (ResNet-50, VGG-16, EfficientNet-B3) to isolate architectural effects under controlled training and sampling conditions (He et al., 2016; Raghu et al., 2021; Simonyan & Zisserman, 2014; Tan & Le, 2019).

In addition, to support clinical interpretability, we introduce Dual-Rollout Attention (Chefer et al., 2021), a method to visualize important regions of the input image by considering the entire ViT response to the input image, rather than focusing only on some layers.

2 | MATERIALS AND METHODS

2.1 | Patient cohort and dataset

This study analysed a combined image-level cohort assembled from two publicly available repositories of colour fundus photographs: the High and Pathologic Myopia Images (HPMI) dataset and the Ocular Disease Intelligent Recognition (OIA-ODIR) dataset (Huang et al., 2023; Li et al., 2021). HPMI served as the source of High Myopia (HM) and Pathologic Myopia (PM) images, whereas OIA-ODIR provided Normal images.

The public HPMI release includes 4011 fundus images with annotations for HM and PM confirmed by multiple ophthalmic examinations, including visual acuity and axial length measurements. OIA-ODIR is a binocular fundus dataset comprising 10 000 images from the left and right eyes of 5000 patients and includes age and sex metadata. It was collected from 487 clinical centres across 26 provinces in China, and images were acquired using multiple camera systems (e.g., Canon, Zeiss, Kowa), resulting in heterogeneous native image resolutions. Accordingly, the two datasets were not assumed to share identical acquisition conditions.

For HPMI, the publicly available release provides image-level labels and dataset-level descriptions but does not include harmonized demographic information (e.g., age, sex, race/ethnicity) or detailed acquisition-device metadata comparable to OIA-ODIR. These limitations reflect the constraints of the publicly available source datasets.

Ground-truth labels were used as provided in the original public datasets. For the HPMI dataset, annotations for HM and PM were defined by the dataset curators. For the OIA-ODIR dataset, diagnostic labels were based on clinical assessments provided in the original dataset and were used to identify Normal eyes. The clinical definitions of HM and PM reflected in the included datasets are broadly consistent with established descriptions in the ophthalmic literature (Kapetanaki et al., 2025; Ohno-Matsui et al., 2015).

We adopted a class-balanced subsampling strategy that used the full complement of available HPMI HM images ($n=773$) as the reference class size. For the PM class, 700 images were randomly sampled from the HPMI dataset and for the Normal class, 700 images were randomly sampled from the OIA-ODIR dataset. The final integrated dataset included 2173 images across three diagnostic categories: HM ($n=773$), PM ($n=700$), and Normal ($n=700$), as summarized in Table 1.

In this study, the unit of analysis was the image rather than the patient; therefore, patient-level

TABLE 1 Distribution of retinal fundus images across diagnostic classes and datasets.

Class	Source dataset(s)	Number of images (total)	Number of images (hold-out test set)
Normal	OIA-ODIR	700	140
High Myopia (HM)	HPMI	773	155
Pathologic Myopia (PM)	HPMI	700	140
Total	-	2173	435

uniqueness was not enforced, and multiple images from the same patient may have been included, particularly in OIA-ODIR, which contains paired left- and right-eye images.

All images were de-identified prior to analysis. Image quality criteria were defined a priori, including a clear view of both the optic disc and macula, sufficient focus to permit identification of fine lesions, and minimal artefact burden (e.g., lashes, dust, arc shadows). Images with co-existing ocular disease likely to confound myopia grading, such as advanced diabetic retinopathy, glaucomatous optic neuropathy, or unrelated macular disease, were considered ineligible. A second, independent quality-control pass was performed to verify adherence to these criteria across the combined dataset. All procedures conformed to the principles of the Declaration of Helsinki (World Medical Association, 2013). As the study used publicly available anonymized data, ethics approval was waived.

2.2 | Data curation and pre-processing

Figure 1 summarizes the workflow from data curation and pre-processing through cross-validated training to held-out evaluation. The original dataset was divided into training, validation, and test subsets through class-stratified sampling. A total of 80% of the images were allocated to the development phase (training and validation), which consisted of model training and cross-validation, while the remaining 20% were held out as a test set.

All images were resized to 518 × 518 pixels to match the native resolution for which the ViT was pre-trained (Oquab et al., 2024). Further, pixel intensity values for each colour channel were normalized to the ImageNet mean and standard deviation (Deng et al., 2009). This pre-processing strategy aligns the statistical properties of the input images with those expected by the pre-trained backbone and improves training stability (Deng et al., 2009; Oquab et al., 2024).

Data augmentation was applied exclusively to the training set and performed online during training. Augmentation included random horizontal flips ($p = 0.5$), small affine perturbations (translation and scaling limits of ±5% and a rotation limit of 15°, with reflection padding at the borders), and moderate brightness and contrast variations (20%), each applied with a probability of 50%. The transformations were constrained to produce realistic augmented images by retaining clinically relevant structures and remaining anatomically plausible for both left and right eyes (Xu et al., 2023). No images were excluded during the quality-control process, as all images met the predefined quality criteria.

2.3 | Model architecture

The classification framework was built on a pre-trained DINOv2 ViT (ViT-S/14), implemented using the timm (PyTorch Image Models) library (Oquab et al., 2024; Wightman, 2019). DINOv2 was selected because recent literature supports self-supervised transformer backbones for retinal imaging tasks in which diagnostically relevant cues are spatially distributed across the fundus rather than confined to a single local lesion (Azad et al., 2024; Hou et al., 2026; Shamshad et al., 2023; Zhou et al., 2023). It was also selected for its ability to capture both local and global contextual relationships, which is important for recognizing the diffuse and spatially distributed anatomical changes associated with myopia (Dosovitskiy et al., 2020; Oquab et al., 2024).

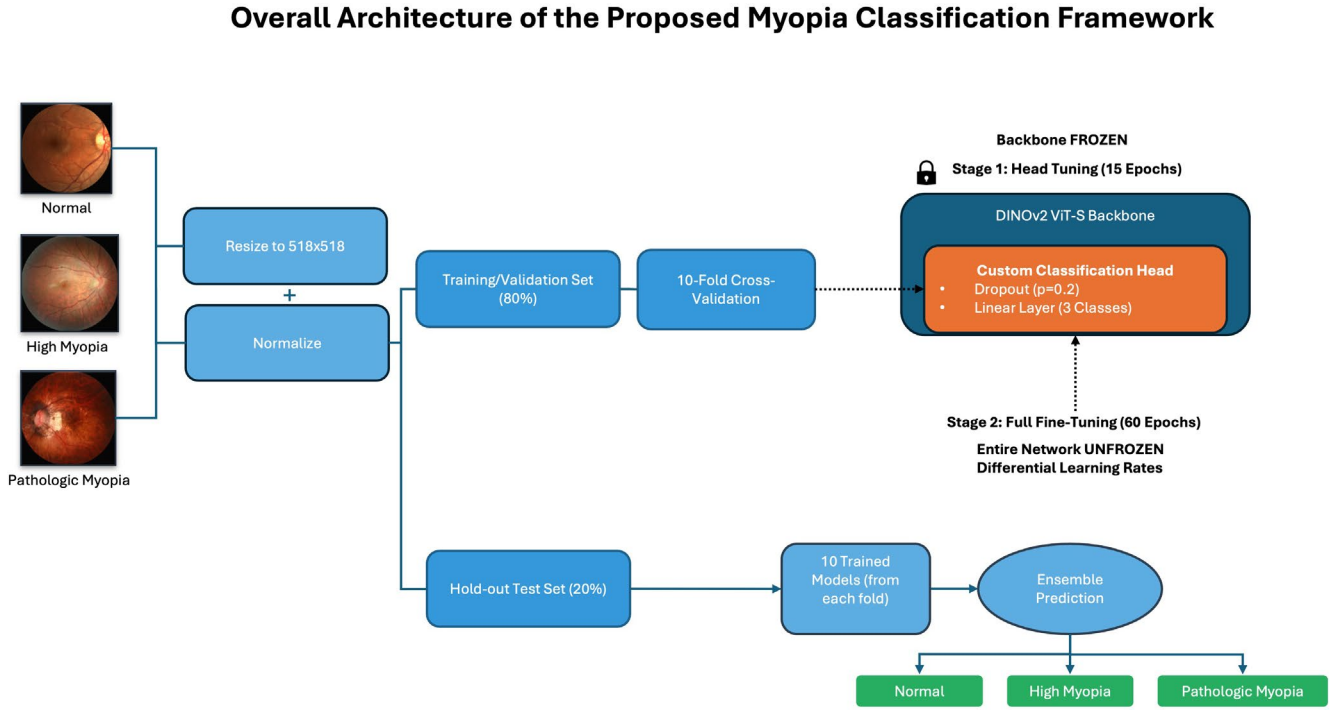

FIGURE 1 Overall methodology flowchart.

To adapt the pre-trained backbone to the three-class classification task, the original classification head was replaced by a custom one consisting of a dropout layer ($p=0.2$) and a fully connected linear layer (Srivastava et al., 2014). This head mapped the 384-dimensional feature vector output produced by the ViT-S/14 backbone to three output logits corresponding to Normal, HM, and PM (Oquab et al., 2024).

2.4 | Design and model training

DINOv2 is a general-purpose self-supervised ViT, pre-trained on large-scale natural image corpora rather than medical ophthalmic imagery. Therefore, domain adaptation was necessary. A two-stage transfer-learning protocol was adapted during the ten-fold cross validation (applied to each fold independently) (Oquab et al., 2024).

In Stage 1, the backbone was frozen and only a randomly initialized classification head was trained using AdamW (decoupled weight decay) under a cosine learning-rate schedule with a short warm-up phase (Loshchilov & Hutter, 2019). In Stage 2, we retained the learned weights of the classification head from Stage 1, but unfroze the layers of the ViT backbone. During this stage, the layers' initial weight values were those obtained from their prior training on non-medical images. These weights were now learnable again, allowing for domain-specific knowledge to be obtained through training. We controlled the learning freedom of this step ahead by introducing a layer-wise learning-rate decay (LLRD), assigning smaller learning rates to earlier transformer blocks to preserve pre-trained low-level representations. During training, focal loss was employed to further mitigate residual class imbalance by down-weighting easy examples, while label smoothing was applied to reduce overconfident predictions and improve generalization.

Training was conducted for a maximum of 60 epochs, as preliminary exploratory experiments indicated that extending the training schedule beyond this threshold did not yield systematic improvements in validation performance. Both validation accuracy and macro-F1 score plateaued around 60 epochs, with no consistent gains observed when training was extended beyond this point. To reduce overfitting and avoid unnecessary computation, training was controlled using early stopping with a patience of 12 epochs, monitored on validation macro-F1 score (Ferro et al., 2023; Tharwat, 2021). During 10-fold cross-validation, the best-performing model checkpoint from each fold was saved. The resulting ten fold-specific models were subsequently ensembled and used to generate predictions on the independent held-out test set. For this procedure, each image of the test set was processed by the inner ten submodels, and their prediction probabilities were averaged to produce the final model's class prediction. Detailed model architecture choices, optimization settings, training hyperparameters, regularization strategies, and data-processing configurations are summarized in Table 2.

2.5 | Benchmark models

The proposed ViT framework was benchmarked against CNN architectures that represent the current state-of-the-art in medical image analysis (Litjens et al., 2017; Ting et al., 2019). The selected models ResNet-50, VGG-16, and EfficientNet-B3 were chosen based on their established performance across several medical imaging tasks and their frequent use as baseline comparators in recent ophthalmic artificial intelligence (AI) studies (Litjens et al., 2017; Muchuchuti & Viriri, 2023; Tan & Le, 2019; Ting et al., 2019). To ensure that differences in performance could be attributed solely to architectural design rather than confounding experimental factors, all benchmark models were trained and evaluated under identical conditions to the proposed DINOv2 ViT framework, including data partitions, pre-processing, augmentation, optimization strategy, and evaluation protocol.

Statistical comparisons between the proposed DINOv2-based framework and CNN baselines were performed using paired, two-sided Student's *t*-tests on fold-level macro–F1 scores across the 10 cross-validation folds. No formal adjustment for multiple comparisons was applied, as the number of comparisons was limited and defined a priori.

2.6 | Explainable AI: Dual-rollout attention

To generate transformer-specific attribution maps, we applied a Dual-Rollout Attention procedure, building on the attention rollout method (Abnar & Zuidema, 2020), to the fine-tuned DINOv2 ViT-S/14 model (Figure 2). For each transformer block, self-attention maps were extracted and averaged across attention heads to obtain a single attention map per layer. To account for residual connections, an identity component was added prior to normalization.

The transformer layers were then divided into two groups: early layers and late layers. Separate attention rollouts were computed for each group. From each rollout, the attention between the classification token and the image patches was extracted, reshaped to the patch grid, and independently normalized to generate an early-layer map and a late-layer map. The early-layer map preserves finer spatial detail, whereas the late-layer map captures higher-level semantic information. This design mitigates the tendency of full attention propagation to produce overly diffuse attribution maps (Chefer et al., 2021) by assigning different contributions to early and late transformer layers.

The final Dual-Rollout heatmap was obtained by combining the two normalized maps using fixed weights of 0.4 for the early-layer map and 0.6 for the late-layer map. These coefficients were empirically selected to balance spatial precision with semantic relevance, placing slightly greater emphasis on late-layer representations that encode more abstract and clinically meaningful features (Raghu et al., 2021).

For visualization, the fused heatmap was resized to the original image resolution, smoothed with a Gaussian

TABLE 2 Summary of key model hyperparameters and training configuration.

Parameter	Value	Justification
Model architecture		
Base Model	vit_small_patch14_dinov2	Self-supervised ViT-Small with strong pretrained features for transfer learning.
Input Image Size	518 × 518 pixels	Matches DINOv2 native resolution.
Batch Size	32	Balances GPU memory constraints with stable batch statistics for high-resolution inputs.
Optimization configuration		
Optimizer	AdamW	Decoupled weight decay implementation, standard for Vision Transformers.
LR Scheduler	Cosine with Warmup	Smooth learning-rate decay; warmup used in both stages (Stage-1: 2 epochs, Stage-2: 5 epochs).
Weight Decay (Backbone)	0.05	Moderate regularization applied to backbone transformer parameters.
Weight Decay (Head)	0.01	Lighter regularization for the newly initialized classification head.
Excluded from Decay	Bias, norms, tokens, pos/rel-pos	Bias terms, layer norms, class/dist tokens, positional embeddings and relative position terms are excluded from decay.
Stage 1: Head-only training		
Learning Rate	2×10^{-4}	Appropriate rate for training a randomly initialized linear head.
Epochs	15	Allows the classification head to stabilize before unfreezing backbone layers.
Stage 2: Full model fine-tuning		
Base LR (Top Blocks)	1×10^{-5}	Conservative learning rate for gentle adaptation of pretrained features.
Layer-wise Decay	0.9 per layer	Earlier transformer blocks receive progressively lower learning rates.
Effective LR Range	3.14×10^{-6} to 1×10^{-5}	Covers from earliest to latest transformer blocks (12 blocks total).
Head Learning Rate	10^{-4}	Classification head uses 10X higher LR than backbone top blocks.
Epochs	60	Sufficient duration for full-model convergence with cosine LR decay.
Loss function & regularization		
Loss Function	Focal Loss + Label Smoothing	Addresses class imbalance while reducing model overconfidence.
Focal Loss γ	2.0	Standard setting to down-weight easily classified examples.
Label Smoothing	0.05	Prevents overconfident predictions and improves generalization.
Dropout (Head)	0.2	Moderate dropout applied before the final classification layer.
Data preprocessing & augmentation		
Training Sampler	WeightedRandomSampler	Ensures balanced sampling across an imbalanced class distribution.
Train Transforms	Resize → HFlip → Shift/Scale/Rotate → Bright/Contrast → Normalize → ToTensorV2	Adds controlled geometric/photometric augmentation to improve robustness.
Val/Test Transforms	Resize → Normalize → ToTensorV2	Minimal preprocessing pipeline for consistent evaluation
Experiment configuration		
Mixed Precision	AMP (autocast + GradScaler)	Accelerated training reduces memory consumption without accuracy loss.
Random Seed	42 + Run_ID	Each run uses a distinct but reproducible seed; splits and training dynamics vary per run.
Cross-Validation Folds	10	Stratified 10-fold CV for robust evaluation and tuning.
Test Split	20%	Hold-out test set reserved for final evaluation and performance reporting.
Early Stopping	Patience = 12 (macro-F1)	Stops training when macro-F1 does not improve for 12 epochs to limit overfitting.

filter, restricted to the circular retinal field, and overlaid on the fundus photograph using a green-to-red colour scale.

2.7 | Experimental environment

All experiments were performed on a local workstation equipped with an NVIDIA GeForce RTX 4080 SUPER GPU, providing sufficient computational resources for accelerated model training and inference.

3 | RESULTS

3.1 | Classification performance of the DINOv2 framework

Following 10-fold stratified cross-validation on the development cohort, the DINOv2-based framework achieved a mean validation macro-F1 score of 0.98, with narrow 95% confidence intervals across folds.

Training and validation performance across all 10 cross-validation folds over 60 fine-tuning epochs are

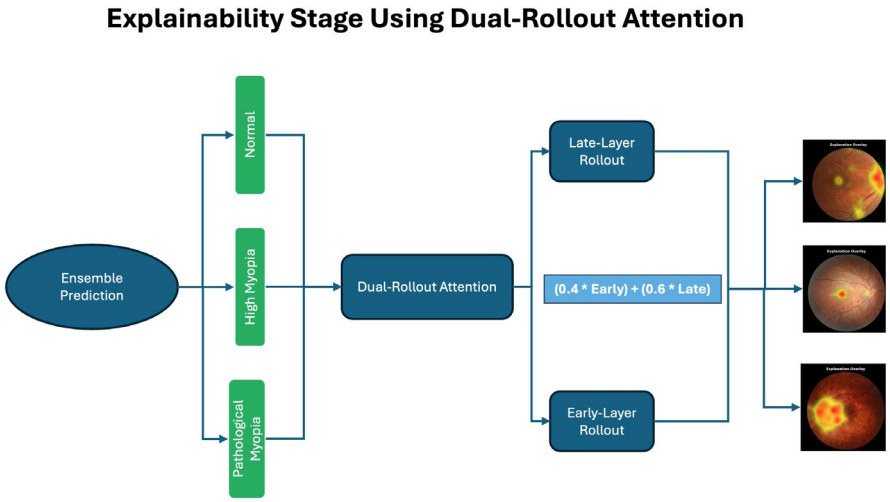

FIGURE 2 Schematic of the Dual-Rollout Attention mechanism.

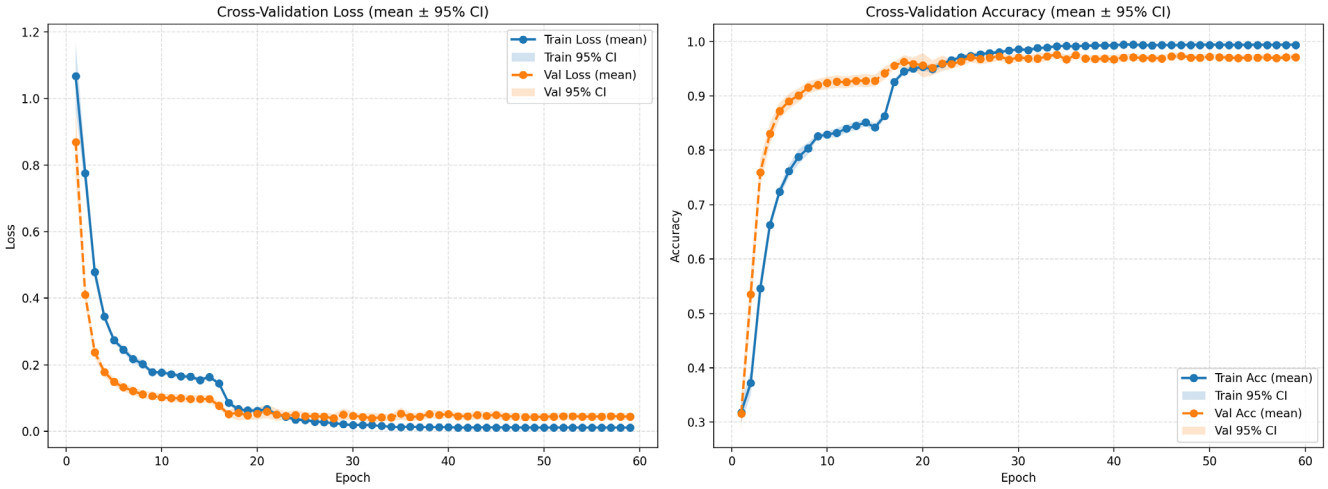

FIGURE 3 Cross-validation learning curves (mean ± 95% CI) for loss and accuracy.

shown in Figure 3. Validation accuracy and loss curves remained consistent across folds, indicating stable optimization and a well-balanced, unbiased stratified split. Validation accuracy increased rapidly after the initial head-only training stage and subsequently plateaued, with minimal gains observed after approximately epochs 20–25. The early-stopping criterion (patience = 12, monitored on validation macro-F1 score) identified this plateau region as the stopping point in most folds.

Across the ten validation folds, DINOv2 achieved an average macro-F1 score of 0.98 (95% CI, 0.9709–0.9888). Per-class accuracy was 99.94% (95% CI, 99.81–100.00%) for Normal, 97.99% (95% CI, 97.09–98.88%) for HM, and 98.04% (95% CI, 97.13–98.96%) for PM. Corresponding sensitivity and specificity values are summarized in Table 3.

On the independent held-out test set, the final model achieved an overall accuracy of 97.03% (95% CI 96.75–97.32%) and a macro-F1 score of 97.08% (95% CI 96.80–97.36%). A confusion matrix summarising the distribution of correct and incorrect predictions across the three diagnostic categories is shown in Figure 4.

3.2 | Class-specific performance

Per-class performance metrics, including accuracy, sensitivity (recall), specificity, and F1 score are summarized in Tables 3 and 4. On the independent test set, accuracy was 99.66% (95% CI, 99.54–99.77%) for Normal, 97.15% (95% CI, 96.86–97.44%) for HM, and 97.26% (95% CI, 97.01–97.51%) for PM. Corresponding sensitivities were 98.93% (95% CI, 98.57–99.29%) for Normal, 96.06% (95% CI, 95.16–96.97%) for HM, and 96.21% (95% CI, 95.31–97.12%) for PM.

3.3 | Comparative performance against CNN baselines

Performance comparisons between the DINOv2-based framework and the selected benchmark CNN architectures (ResNet-50, VGG-16, and EfficientNet-B3) are summarized in Table 4. Receiver operating characteristic (ROC) curves for all models are presented in Figure 5, enabling visual comparison of model performance across classification thresholds. The corresponding area under

TABLE 3 Validation and test performance (mean with 95% CI) by class and metric.

Metric	Validation mean (95% CI)	Test mean (95% CI)
Normal		
Accuracy	99.94% (99.81–100.00%)	99.66% (99.54–99.77%)
Sensitivity (Recall)	99.82% (99.42–100.00%)	98.93% (98.57–99.29%)
Specificity	100.00% (100.00–100.00%)	100.00% (100.00–100.00%)
F1-Score	99.91% (99.71–100.00%)	99.46% (99.28–99.64%)
High myopia		
Accuracy	97.99% (97.09–98.88%)	97.15% (96.86–97.44%)
Sensitivity (Recall)	97.57% (96.59–98.55%)	96.06% (95.16–96.97%)
Specificity	98.21% (97.17–99.26%)	97.75% (97.37–98.13%)
F1-Score	97.19% (95.96–98.42%)	96.00% (95.58–96.42%)
Pathologic myopia		
Accuracy	98.04% (97.13–98.96%)	97.26% (97.01–97.51%)
Sensitivity (Recall)	96.61% (94.48–98.73%)	96.21% (95.31–97.12%)
Specificity	98.73% (98.21–99.24%)	97.76% (97.26–98.26%)
F1-Score	96.94% (95.49–98.39%)	95.77% (95.39–96.15%)

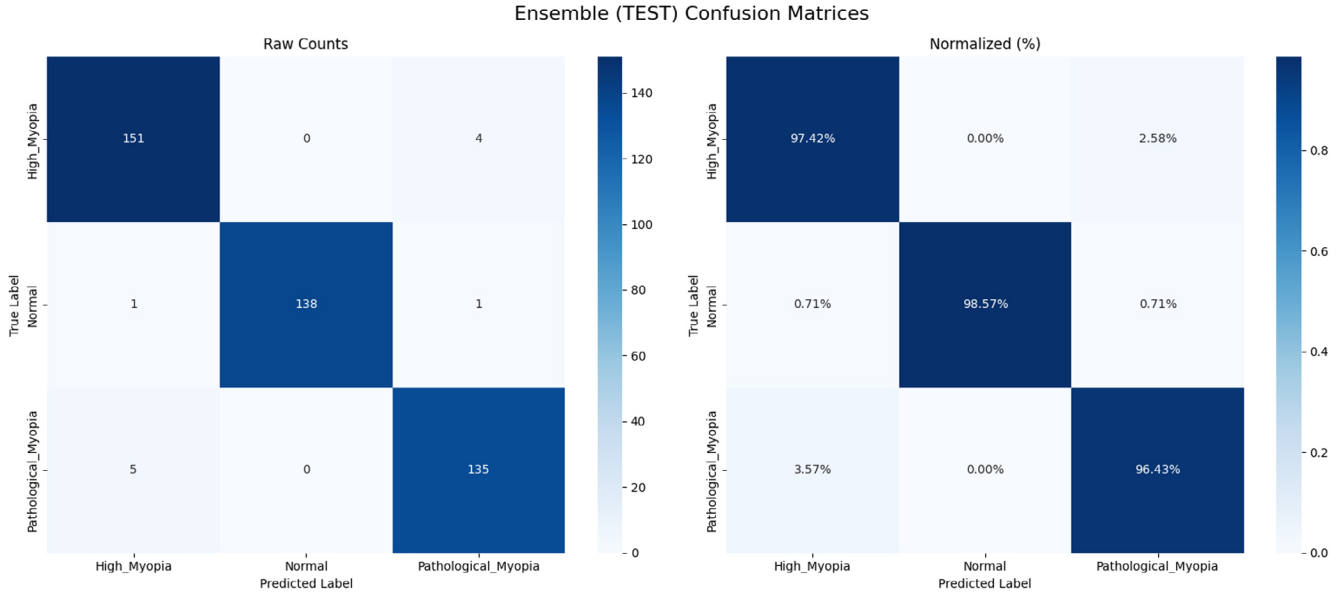

FIGURE 4 Confusion matrix for the DINOv2-based model on the independent held-out test set.

the curve (AUC) values are reported in the figure legend. The ViT-based model consistently outperformed all examined CNN baselines in terms of macro-averaged evaluation metrics, including ResNet-50 ($p<0.001$), VGG-16 ($p<0.001$), and EfficientNet-B3 ($p<0.001$).

3.4 | Qualitative analysis via dual-rollout attention maps

Representative attention heatmaps generated using the Dual-Rollout Attention method for images from the independent test set are shown in Figures 6 and 7. For images classified as Normal, the attention maps generally displayed diffuse, low-intensity activation distributed across the fundus. For HM, attention maps demonstrated localized activation patterns involving peripapillary regions and optic disc morphology. For PM, attention maps highlighted regions corresponding to diffuse or patch chorioretinal atrophy and other degenerative changes.

To further investigate model behaviour in erroneous predictions, all misclassified cases from the independent test set ($n=11$) are presented in Figures 8 and 9. These examples include both false-positive and false-negative classifications, along with their corresponding prediction probabilities and Dual-Rollout attention maps.

In several cases, misclassifications occurred between HM and PM, often with relatively close probability scores, indicating model uncertainty near the decision boundary. Attention maps in these cases frequently highlighted clinically relevant regions, such as the optic disc and peripapillary areas, but with variable spatial focus or incomplete coverage of pathological features. Not paying enough attention to such features or completely missing them explains the model's inability to discriminate between the myopic categories and also explains the close probability

TABLE 4 Performance metrics on the hold-out test set by DINOv2 and selected CNNs.

Class	Accuracy	Sensitivity	Specificity	F1
EfficientNet-B3				
Overall	88.46% (87.56–89.36%)	88.83% (87.95–89.70%)	94.25% (93.80–94.69%)	88.55% (87.64–89.46%)
Normal	97.52% (96.73–98.31%)	96.93% (95.37–98.49%)	97.80% (96.75–98.84%)	96.18% (94.98–97.38%)
High Myopia	89.13% (88.32–89.94%)	78.13% (75.34–80.92%)	95.21% (93.82–96.61%)	83.63% (82.26–85.00%)
Pathologic Myopia	90.28% (89.47–91.08%)	91.43% (90.04–92.81%)	89.73% (88.36–91.10%)	85.84% (84.82–86.85%)
ResNet-50				
Overall	94.02% (93.51–94.53%)	94.15% (93.67–94.62%)	97.00% (96.76–97.25%)	94.12% (93.62–94.61%)
Normal	99.56% (99.34–99.79%)	98.71% (98.04–99.39%)	99.97% (99.89–100.00%)	99.32% (98.96–99.67%)
High Myopia	94.39% (93.98–94.80%)	90.58% (88.89–92.27%)	96.50% (96.04–96.96%)	91.99% (91.33–92.65%)
Pathologic Myopia	94.09% (93.58–94.61%)	93.14% (92.34–93.95%)	94.54% (93.53–95.55%)	91.04% (90.36–91.73%)
VGG-16				
Overall	94.23% (93.31–95.15%)	94.32% (93.41–95.24%)	97.10% (96.64–97.55%)	94.31% (93.39–95.22%)
Normal	98.97% (98.20–99.73%)	98.00% (95.92–100.00%)	99.42% (99.22–99.62%)	98.37% (97.13–99.61%)
High Myopia	94.51% (93.91–95.10%)	91.61% (90.34–92.88%)	96.11% (95.68–96.53%)	92.23% (91.37–93.09%)
Pathologic Myopia	94.99% (94.37–95.60%)	93.36% (92.67–94.04%)	95.76% (94.78–96.75%)	92.32% (91.46–93.18%)
DINOv2				
Overall	97.03%★ (96.75–97.32%)	97.07% ★ (96.80–97.34%)	98.50% ★ (98.36–98.65%)	97.08% ★ (96.80–97.36%)
Normal	99.66% ★ (99.54–99.77%)	98.93% ★ (98.57–99.29%)	100.00% ★ (100.00–100.00%)	99.46% ★ (99.28–99.64%)
High Myopia	97.15% ★ (96.86–97.44%)	96.06% ★ (95.16–96.97%)	97.75% ★ (97.37–98.13%)	96.00% ★ (95.58–96.42%)
Pathologic Myopia	97.26% ★ (97.01–97.51%)	96.21% ★ (95.31–97.12%)	97.76% ★ (97.26–98.26%)	95.77% ★ (95.39–96.15%)

Note: Best value per metric (within each class/overall row) is marked with ★.

scores. Conversely, some PM cases misclassified as HM demonstrated attention focused on localized regions, potentially underrepresenting the full extent of diffuse degenerative changes.

Misclassification of Normal images as myopic categories in the only two False Negative cases of the test dataset was associated with attention concentrated on peripheral retinal regions or image artefacts, suggesting that non-pathological structures may have influenced the model's predictions.

Overall, these findings suggest that model errors are more likely to occur in borderline or visually ambiguous cases, where pathological features are subtle, spatially heterogeneous, or partially represented in the image.

4 | DISCUSSION

4.1 | Principal findings

In this work, we developed, validated, and interpreted a DL framework for three-class classification of Normal, HM, and PM from colour fundus photographs, using a self-supervised DINOv2 ViT with a two-stage transfer-learning protocol and a ten-fold cross-validated ensemble evaluated on an independent held-out test set. Under identical training, sampling, and evaluation conditions, the proposed ViT model consistently outperformed established CNN baselines.

CNNs have traditionally served as a strong baseline for fundus image classification; however, our results demonstrate that a self-supervised ViT can achieve superior performance, particularly in sensitivity for clinically relevant classes, without sacrificing specificity. Although VGG-16 and ResNet-50 demonstrated competitive performance, and it is possible that further architecture-specific optimization could improve their results, all models were intentionally trained under identical conditions to ensure a fair and controlled comparison. These findings support the view that large pre-trained transformer models are not excessive for moderately sized medical imaging datasets, but instead can provide meaningful performance advantages when appropriately adapted. The consistent outperformance over established CNN baselines under identical experimental conditions suggests that transformer-based architectures may better capture the global retinal patterns relevant to myopic disease severity.

4.2 | Comparison with prior work

Prior studies applying DL to myopia classification have largely relied on CNN-based architectures and have focused predominantly on binary classification tasks, most commonly distinguishing pathologic from non-pathologic myopia (Du et al., 2021; Ting et al., 2019). While these approaches established the feasibility of automated detection, they do not fully

reflect real-world clinical decision-making, where clinicians must distinguish among Normal eyes, HM, and PM, with the HM–PM boundary being particularly consequential.

By explicitly addressing this three-class problem and adopting a ViT architecture, the present study extends earlier work and aligns more closely with clinical practice. ViTs are inherently well-suited to modelling the global, spatially distributed retinal changes characteristic of progressive myopia, including diffuse chorioretinal atrophy and posterior pole remodelling, which may be incompletely captured by architectures with predominantly local receptive fields.

4.3 | Generalization and dataset considerations

Although the proposed framework achieved strong performance on an independent held-out test set, a reduction from near-ceiling cross-validation metrics to slightly lower test performance was observed. This pattern is expected and likely reflects partial "in-distribution familiarity" introduced by dataset composition and controlled sampling strategies, rather than purely universal disease-feature learning.

In ophthalmic imaging, truly external validation often introduces additional sources of distribution shift, including differences in camera vendors, acquisition protocols, field of view, image quality, population characteristics, and labeling conventions. These factors typically lead to further performance degradation beyond that observed with internal held-out testing. Therefore, the present findings underscore the importance of external, multi-center validation by fully characterizing generalizability.

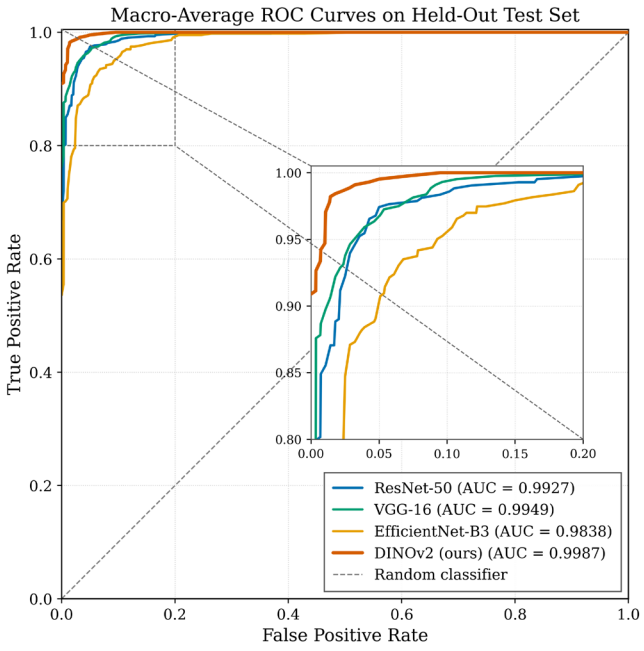

FIGURE 5 Receiver operating characteristic (ROC) curves for DINOv2 and CNN baseline models on the independent test set (AUC = area under the curve).

4.4 | Interpretability and clinical transparency

A persistent barrier to the clinical adoption of DL systems is their perceived "black-box" nature (Borys et al., 2023; Rudin, 2019). Clinicians must be confident that a model's predictions are driven by clinically meaningful structures rather than spurious correlations or imaging artefacts (Borys et al., 2023; Geirhos et al., 2020). This requirement is especially important in conditions such as PM, where diagnostic decisions may carry long-term visual consequences.

To address this challenge, we incorporated Dual-Rollout Attention as an interpretability mechanism tailored to ViTs. By combining attention information from early and late transformer layers, this approach produces attribution maps that preserve spatial precision while retaining semantic relevance. In qualitative analyses, the resulting heatmaps showed alignment with clinically recognised features of myopic degeneration.

Although Dual-Rollout Attention produced clinically plausible attribution maps in many cases, not all visualizations consistently highlighted every anatomically relevant feature, reflecting a known limitation of attention-based explainability methods, which emphasize regions most influential for a given prediction rather than providing exhaustive lesion-level delineation.

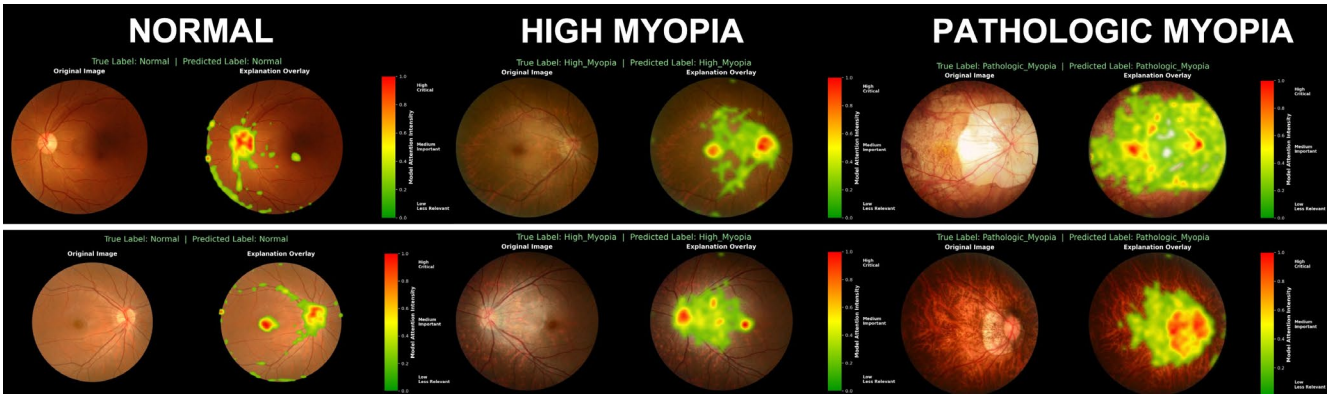

FIGURE 6 Representative examples of model explainability using the Dual-Rollout Attention method. For each diagnostic category (Normal, High Myopia, and Pathologic Myopia), a fundus photograph from the held-out test set is shown alongside its corresponding attention heatmap, illustrating regions contributing to the model's prediction.

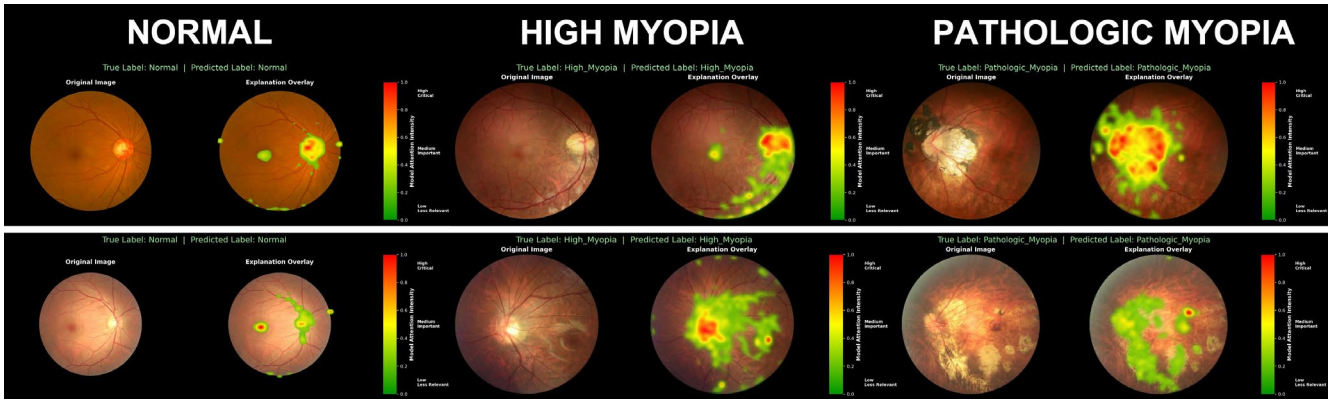

FIGURE 7 Additional representative examples of Dual-Rollout Attention visualizations from the held-out test set, demonstrating the consistency of model focus across different images within each diagnostic category (Normal, High Myopia, and Pathologic Myopia).

Beyond transparency, such visualizations may facilitate systematic error analysis and support hypothesis generation regarding under-recognized retinal patterns.

Extending this qualitative analysis, the examination of misclassified cases further provides insight into the model's decision-making process and limitations. Errors were most commonly observed at the interface between HM and PM, reflecting the inherent clinical difficulty in distinguishing these categories, particularly in early or borderline presentations. In such cases, prediction probabilities were often distributed across competing classes, indicating model uncertainty rather than confident misclassification.

Inspection of the corresponding attention maps revealed that, even in erroneous predictions, the model frequently focused on clinically relevant regions, such as the optic disc, peripapillary atrophy, and areas suggestive of chorioretinal degeneration. However, attention was occasionally restricted to localized regions or influenced by peripheral structures and imaging artefacts, which may have contributed to incomplete or biased feature representation.

These findings highlight that misclassifications often arise from subtle phenotypic overlap, heterogeneous spatial presentation of pathology, or partial feature representation within a single image. Importantly, this analysis is based on all misclassified cases from the independent test set rather than selected examples, supporting the robustness of these observations.

4.5 | Clinical implications

An automated system capable of reliably classifying eyes as Normal, HM, or PM has several potential clinical applications. Such a tool could support scalable screening in regions with rising myopia prevalence, standardise grading and reduce inter-observer variability, particularly at the HM–PM interface. Also, it enables earlier referral and closer monitoring for eyes at risk of pathologic progression.

Foundation-model approaches, including self-supervised ViTs, may further enhance robustness across imaging devices and populations when coupled with appropriate calibration and post-deployment monitoring. In routine practice, the integration of such systems as background decision-support tools or "second readers" could reduce grading burden, streamline referral pathways, and improve triage efficiency.

4.6 | Limitations and future directions

This study has several limitations. Although images were sourced from multiple centers, the demographic and ethnic diversity of the dataset remains limited, and broader external validation across different populations, imaging devices, and care settings is required to identify potential performance disparities. Additionally, because the Normal and myopic cohorts were derived from different public datasets and comprehensive demographic and acquisition metadata were not available for all images, we cannot exclude the possibility that dataset-specific factors, including camera characteristics, population differences, retinal pigmentation, or other unmeasured confounders, contributed to model performance. The retrospective nature of the datasets also limits representation of real-world variability, including technician-dependent factors, media opacities, and suboptimal image acquisition.

Furthermore, because reliable subject identifiers were not available across all source datasets, subject-level partitioning could not be enforced. As a result, paired-eye images from the same individual may have been included in different data subsets, potentially introducing sample correlation and leading to optimistic estimates of model performance.

While the DINOv2 model provided strong initialization through natural-image pre-training, retina-specific self-supervised pre-training strategies or future medical foundation models may further improve data efficiency and robustness, particularly for under-represented phenotypes and borderline presentations. Finally, although Dual-Rollout Attention enhances the plausibility of attribution maps, its practical utility for clinicians requires prospective, user-centered evaluation in clinical screening and follow-up settings, particularly to determine whether such explanations objectively improve grading consistency, confidence in referral decisions, and detection of clinically relevant errors.

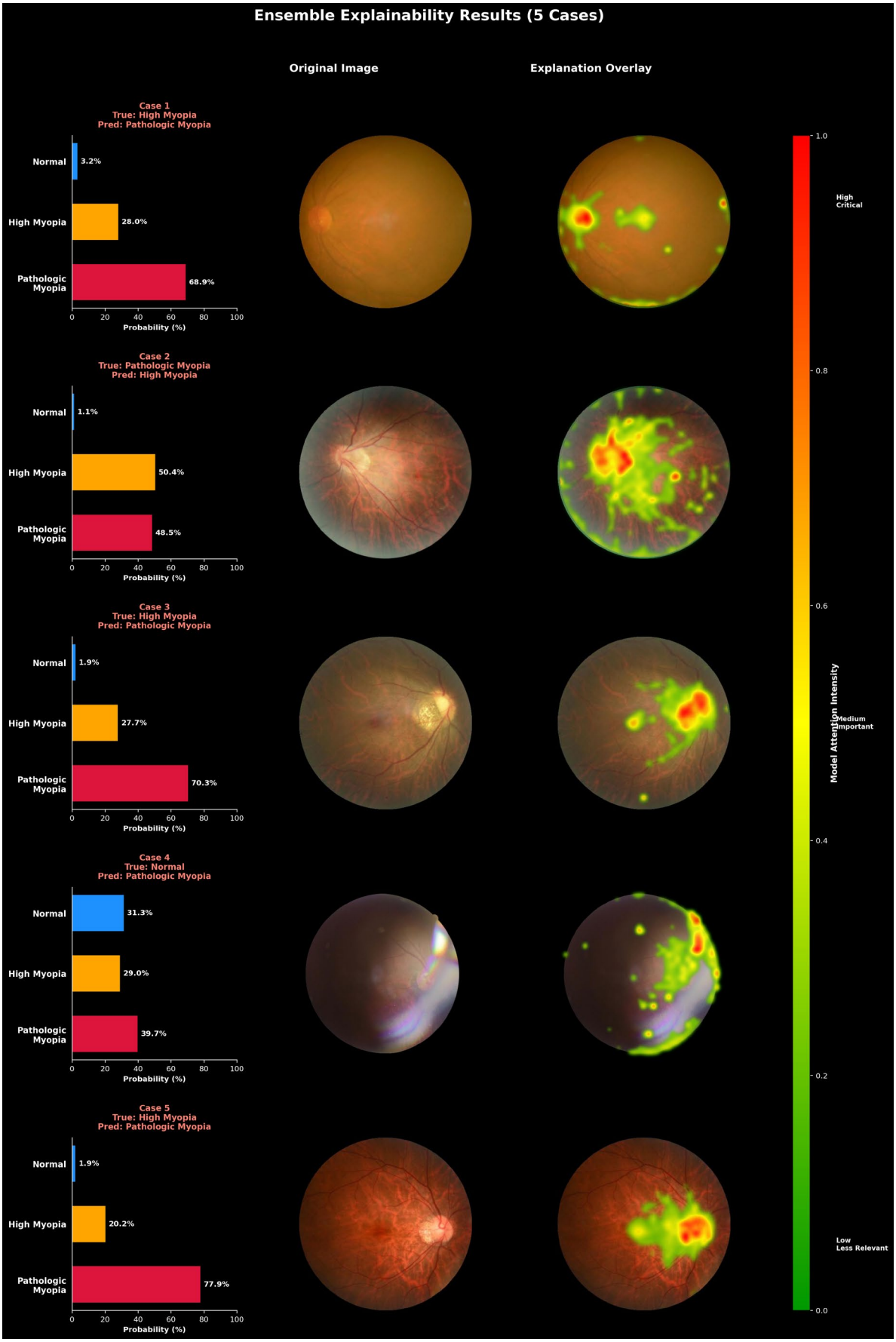

FIGURE 8 All misclassified cases from the independent test set (panel 1). Each panel displays a fundus photograph alongside the corresponding Dual-Rollout Attention heatmap and model-predicted class probabilities. Cases include both false-positive and false-negative predictions, primarily involving confusion between High Myopia and Pathologic Myopia, as well as Normal images incorrectly classified as myopic categories. Probability scores are shown for all three classes, illustrating both confident errors and borderline predictions.

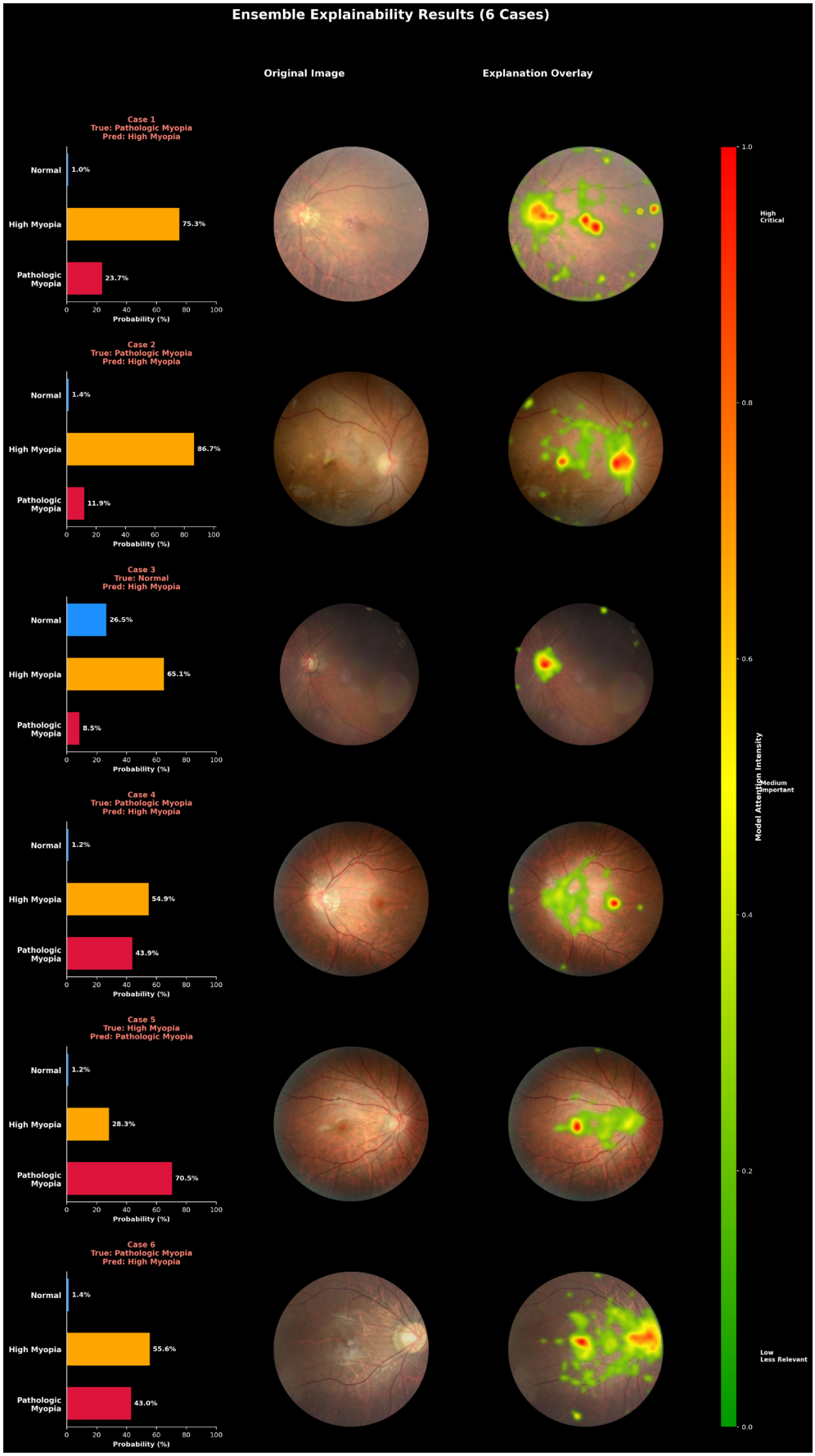

FIGURE 9 All misclassified cases from the independent test set (panel 2). Each panel displays a fundus photograph alongside the corresponding Dual-Rollout Attention heatmap and model-predicted class probabilities. Additional examples of false-positive and false-negative predictions are presented, including confusion between High Myopia and Pathologic Myopia and misclassification of Normal images. Probability scores reflect varying levels of model confidence, ranging from high-certainty errors to borderline cases.

5 | CONCLUSIONS

In this study, we developed and validated a self-supervised ViT framework for automated three-class classification of Normal, HM, and PM from colour fundus photographs. By combining DINOv2 pre-training with a two-stage transfer-learning strategy and transformer-specific interpretability, the proposed approach achieved strong diagnostic performance while providing clinically meaningful visual explanations. These findings support the potential of self-supervised transformer models as scalable, transparent tools for myopia assessment, with external and prospective validation needed to confirm robustness in real-world clinical settings.

ACKNOWLEDGEMENTS

The publication of this article in OA mode was financially supported by HEAL-Link.

FUNDING INFORMATION

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ORCID

Eirini Maliagkani https://orcid.org/0009-0008-3679-8807
Ioannis D. Apostolopoulos https://orcid.org/0000-0001-6439-9282

REFERENCES

Abnar, S. & Zuidema, W. (2020) Quantifying attention flow in transformers. In: *Proceedings of the 58th annual meeting of the Association for Computational Linguistics*. Stroudsburg, PA: Association for Computational Linguistics, pp. 4190–4197. Available from: https://doi.org/10.18653/v1/2020.acl-main.385

Abràmoff, M.D., Lavin, P.T., Birch, M., Shah, N. & Folk, J.C. (2018) Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *npj Digital Medicine*, 1, 39. Available from: https://doi.org/10.1038/s41746-018-0040-6

Azad, R., Kazerouni, A., Heidari, M., Aghdam, E.K., Molaei, A., Jia, Y. et al. (2024) Advances in medical image analysis with vision transformers: a comprehensive review. *Medical Image Analysis*, 91, 103000. Available from: https://doi.org/10.1016/j.media.2023.103000

Benahmed, R., Dormegny, L., Gaudric, A., Philippakis, E., Sauer, A., Bourcier, T. et al. (2025) Perivascular Chorioretinal atrophy: an unusual feature in pathologic myopia eyes. *American Journal of Ophthalmology*, 271, 498–506. Available from: https://doi.org/10.1016/j.ajo.2024.12.022

Borys, K., Schmitt, Y.A., Nauta, M., Seifert, C., Krämer, N., Friedrich, C.M. et al. (2023) Explainable AI in medical imaging: an overview for clinical practitioners—beyond saliency-based XAI approaches. *European Journal of Radiology*, 162, 110786. Available from: https://doi.org/10.1016/j.ejrad.2023.110786

Catania, F., Chapron, T., Crincoli, E., Miere, A., Abdelmassih, Y., Beaumont, W. et al. (2024) Deep learning for prediction of late recurrence of retinal detachment using preoperative and postoperative ultra-wide field imaging. *Acta Ophthalmologica*, 102(7), e984–e993. Available from: https://doi.org/10.1111/aos.16693

Chefer, H., Gur, S. & Wolf, L. (2021) Transformer interpretability beyond attention visualization. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*. New York, NY: IEEE, pp. 782–791. Available from: https://doi.org/10.1109/CVPR46437.2021.00084

De Fauw, J., Ledsam, J.R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S. et al. (2018) Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine*, 24(9), 1342–1350. Available from: https://doi.org/10.1038/s41591-018-0107-6

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. & Fei-Fei, L. (2009) ImageNet: a large-scale hierarchical image database. In: *2009 IEEE conference on computer vision and pattern recognition (CVPR)*. New York, NY: IEEE, pp. 248–255. Available from: https://doi.org/10.1109/CVPR.2009.5206848

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T. et al. (2020) An image is worth 16×16 words: transformers for image recognition at scale. *arXiv* Preprint. https://arxiv.org/abs/2010.11929

Du, R., Xie, S., Fang, Y., Igarashi-Yokoi, T., Moriyama, M., Ogata, S. et al. (2021) Deep learning approach for automated detection of myopic maculopathy and pathologic myopia in fundus images. *Ophthalmology. Retina*, 5(12), 1235–1244. Available from: https://doi.org/10.1016/j.oret.2021.02.006

Ferro, M.V., Doval Mosquera, Y., Ribadas Pena, F.J. & Darriba Bilbao, V.M. (2023) Early stopping by correlating online indicators in neural networks. *Neural Networks*, 159, 109–124. Available from: https://doi.org/10.1016/j.neunet.2022.11.035

Flitcroft, D.I., He, M., Jonas, J.B., Jong, M., Naidoo, K., Ohno-Matsui, K. et al. (2019) IMI—defining and classifying myopia: a proposed set of standards for clinical and epidemiologic studies. *Investigative Ophthalmology & Visual Science*, 60(3), M20–M30. Available from: https://doi.org/10.1167/iovs.18-25957

Flores-Moreno, I., Puertas, M., Fernández-Jiménez, M., Franco, L.C., Terrón-Vilalta, M., Eslava, B. et al. (2025) Myopic maculopathy progression: insights into posterior staphyloma and macular involvement. *American Journal of Ophthalmology*, 270, 164–171. Available from: https://doi.org/10.1016/j.ajo.2024.09.035

Geirhos, R., Jacobsen, J.H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M. et al. (2020) Shortcut learning in deep neural networks. *Nature Machine Intelligence*, 2, 665–673. Available from: https://doi.org/10.1038/s42256-020-00257-z

Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A. et al. (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22), 2402–2410. Available from: https://doi.org/10.1001/jama.2016.17216

He, K., Zhang, X., Ren, S. & Sun, J. (2016) Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*. New York, NY: IEEE, pp. 770–778.

Hou, Q., Zhou, Y., Lin Goh, J.H., Zou, K., Yew, S.M.E., Srinivasan, S. et al. (2026) Can a natural image-based foundation model outperform a retina-specific model in detecting ocular and systemic diseases? *Ophthalmology Science*, 6(1), 100923. Available from: https://doi.org/10.1016/j.xops.2025.100923

Huang, S., Li, Z., Lin, B., Zhang, S., Yi, Q. & Wang, L. (2023) HPMI: a retinal fundus image dataset for identification of high and pathological myopia based on deep learning. *Figshare*. https://doi.org/10.6084/m9.figshare.24800232.v1

Kapetanaki, M.V., Maliagkani, E., Tyrlis, K. & Georgalas, I. (2025) Artificial intelligence in myopic maculopathy: a comprehensive review of identification, classification, and monitoring using diverse imaging modalities. *Cureus*, 17(2), e78685. Available from: https://doi.org/10.7759/cureus.78685

Li, N., Li, T., Hu, C., Wang, K. & Kang, H. (2021) A benchmark of ocular disease intelligent recognition: one shot for multi-disease detection. *arXiv*. https://doi.org/10.48550/arXiv.2102.07978

Litjens, G., Kooi, T., Ehteshami Bejnordi, B., Setio, A.A.A., Ciompi, F., Ghafoorian, M. et al. (2017) A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60–88. Available from: https://doi.org/10.1016/j.media.2017.07.005

Loshchilov, I. & Hutter, F. (2019) Decoupled weight decay regularization. In International conference on learning representations (ICLR). arXiv. https://arxiv.org/abs/1711.05101

Muchuchuti, S. & Viriri, S. (2023) Retinal disease detection using deep learning techniques: a comprehensive review. *Journal of Imaging*, 9(4), 84. Available from: https://doi.org/10.3390/jimaging9040084

Ohno-Matsui, K., Kawasaki, R., Jonas, J.B., Cheung, C.M., Saw, S.M., Verhoeven, V.J. et al. (2015) International photographic classification and grading system for myopic maculopathy. *American Journal of Ophthalmology*, 159(5), 877–883.e7. Available from: https://doi.org/10.1016/j.ajo.2015.01.022

Ohno-Matsui, K., Wu, P.C., Yamashiro, K., Vutipongsatorn, K., Fang, Y., Cheung, C.M.G. et al. (2021) IMI pathologic myopia. *Investigative Ophthalmology & Visual Science*, 62(5), 5. Available from: https://doi.org/10.1167/iovs.62.5.5

Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V. et al. (2024) DINOv2: learning robust visual features without supervision. *arXiv*. https://arxiv.org/abs/2304.07193

Pärssinen, O. (2025) Factors associated with the high prevalence of myopia and its decrease-a historical review. *Acta Ophthalmologica*, 103(8), 879–890. Available from: https://doi.org/10.1111/aos.70001

Qi, X., Gapar Md Johar, M., Khatibi, A. & Tham, J. (2025) Exploiting gaussian-based effective receptive fields for object detection. *Scientific Reports*, 15, 25008. Available from: https://doi.org/10.1038/s41598-025-10548-3

Raghu, M., Unterthiner, T., Kornblith, S., Zhang, C. & Dosovitskiy, A. (2021) Do vision transformers see like convolutional neural networks? *Advances in Neural Information Processing Systems*, 34, 12116–12128.

Rudin, C. (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1, 206–215. Available from: https://doi.org/10.1038/s42256-019-0048-x

Sankaridurg, P., Tahhan, N., Kandel, H., Naduvilath, T., Zou, H., Frick, K.D. et al. (2021) IMI impact of myopia. *Investigative Ophthalmology & Visual Science*, 62(5), 2. Available from: https://doi.org/10.1167/iovs.62.5.2

Shamshad, F., Khan, S., Zamir, S.W., Khan, M.H., Hayat, M., Khan, F.S. et al. (2023) Transformers in medical imaging: a survey. *Medical Image Analysis*, 88, 102802. Available from: https://doi.org/10.1016/j.media.2023.102802

Simonyan, K. & Zisserman, A. (2014) Very deep convolutional networks for large-scale image recognition. *arXiv*. https://arxiv.org/abs/1409.1556

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. (2014) Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15, 1929–1958. Available from: https://doi.org/10.5555/2627435.2670313

Tan, M. & Le, Q.V. (2019) EfficientNet: rethinking model scaling for convolutional neural networks. In: *Proceedings of the 36th international conference on machine learning (ICML)*. Paris, France: PMLR, pp. 6105–6114.

Tharwat, A. (2021) Classification assessment methods. *Applied Computing and Informatics*, 17(1), 168–192. Available from: https://doi.org/10.1016/j.aci.2018.08.003

Ting, D.S.W., Cheung, C.Y., Lim, G., Tan, G.S.W., Quang, N.D., Gan, A. et al. (2017) Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA*, 318(22), 2211–2223. Available from: https://doi.org/10.1001/jama.2017.18152

Ting, D.S.W., Pasquale, L.R., Peng, L., Campbell, J.P., Lee, A.Y., Raman, R. et al. (2019) Artificial intelligence and deep learning in ophthalmology. *The British Journal of Ophthalmology*, 103(2), 167–175. Available from: https://doi.org/10.1136/bjophthalmol-2018-313173

Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A. & Jégou, H. (2021) Training data-efficient image transformers and distillation through attention. In: Meila, M. & Zhang, T. (Eds.) *Proceedings of the 38th international conference on machine learning (ICML)*. Paris, France: PMLR, pp. 10347–10357.

Wightman, R. (2019) *PyTorch image models (timm) [software]*. San Francisco, CA: GitHub. Available from: https://doi.org/10.5281/zenodo.4414861

World Medical Association. (2013) World medical association declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*, 310(20), 2191–2194. Available from: https://doi.org/10.1001/jama.2013.281053

Xu, M., Yoon, S., Fuentes, A. & Park, D.S. (2023) A comprehensive survey of image augmentation techniques for deep learning. *Pattern Recognition*, 137, 109347. Available from: https://doi.org/10.1016/j.patcog.2023.109347

Zhou, Y., Chia, M.A., Wagner, S.K., Ayhan, M.S., Williamson, D.J., Struyven, R.R. et al. (2023) A foundation model for generalizable disease detection from retinal images. *Nature*, 622(7981), 156–163. Available from: https://doi.org/10.1038/s41586-023-06555-x

How to cite this article: Maliagkani, E., Tsoutsas, C., Papageorgiou, N., Apostolopoulos, I.D., Papandrianos, N., Georgalas, I. et al. (2026) Explainable deep learning self-supervised vision transformer for fundus-based myopia classification. *Acta Ophthalmologica*, 00, 1–14. Available from: https://doi.org/10.1111/aos.70205