Research Square

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

Explainable Multimodal Deep Learning Model for Early Prediction of Treatment-Requiring Retinopathy of Prematurity

Fatemeh Sefidbaf
Mashhad University of Medical Sciences
Fatemeh Baharvand Ahmadi
Mashhad University of Medical Sciences
Nasser Shoeibi
Mashhad University of Medical Sciences
Saeid Eslami
Mashhad University of Medical Sciences
Amin Amiri Tehranizadeh
AmiriMA@mums.ac.ir

Mashhad University of Medical Sciences

Research Article

Keywords: Retinopathy of Prematurity, multimodal analysis, machine learning, deep learning, fundus images

Posted Date: March 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8838478/v1

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

Additional Declarations: No competing interests reported.

Abstract

Background: Retinopathy of prematurity (ROP) is a leading preventable cause of childhood blindness. Current screening guidelines, based primarily on gestational age and birth weight, result in numerous unnecessary examinations. We aimed to develop an explainable multimodal deep learning model for early prediction of treatment-requiring ROP.

Methods: In a retrospective cohort of 384 preterm infants (203 treated, 181 untreated) from a tertiary center in Iran (2021-2024), we integrated four directional fundus images, semi-supervised vessel segmentation maps (using a U-Net-based adversarial domain-adaptation approach), and comprehensive clinical/demographic data. A multi-view fusion model with an attention mechanism extracted vessel-aware features, reduced via PCA, and combined with clinical variables. Six multimodal feature sets were evaluated using 14 machine learning classifiers with 5-fold stratified cross-validation.

Results: The best-performing models (Extra Trees and Random Forest) on the full multimodal feature set achieved a test accuracy of 0.987, an AUC-ROC of 0.999, an F1-score of 0.988, and near-perfect specificity (up to 1.000). Interpretability analyses (SHAP and Grad-CAM) confirmed that predictions were primarily driven by vascular morphology features (PCA1) and posterior pole abnormalities consistent with plus disease.

Conclusions: The proposed explainable multimodal model significantly outperforms clinical-only approaches and represents a promising tool for risk stratification in ROP screening. It has the potential to reduce unnecessary examinations, infant stress, and healthcare burden while facilitating timely intervention. External multicenter validation is warranted.

Introduction

Retinopathy of Prematurity (ROP) is a disease characterized by abnormal retinal blood vessel growth in preterm newborns [1]. The condition typically starts when these infants are exposed to high oxygen levels (hyperoxia), causing existing vessels to close and regress. When oxygen levels normalize again, the resulting relative hypoxia drives excessive new vessel formation, mainly through vascular endothelial growth factor (VEGF) [2]. Known risk factors include extremely low gestational age or birth weight, inflammation during pregnancy or after birth, the general immaturity of premature infants, lung-related complications, and anemia [3]. Each year, around 15 million premature infants are born globally, placing approximately 1.4 million at risk for ROP [4]. Roughly 32,000 of these infants progress to severe visual impairment or blindness [4]. In Iran, prevalence rates reported in various studies range widely from 5.6% to 70.3% [5–30]. Available treatments include laser photocoagulation, cryotherapy, and anti-VEGF injections. Still, early diagnosis via timely screening is essential for successful management [31, 32]. Current guidelines rely mostly on gestational age and birth weight, which frequently leads to screening many infants who will not develop treatment-requiring disease [33–35]. Frequent eye examinations can be stressful and uncomfortable for the infants, while also creating emotional and financial strain for families [36, 37]. Factors like delayed referrals, missed appointments, shortage of pediatric retinal specialists, and inadequate parental awareness further raise the chances of preventable blindness [4]. Recent studies have shown accuracies over 90% for image-only models in detecting plus disease and cases that need intervention. These results come from applying convolutional neural networks to fundus photographs (Shoeibi et al., 2025[38]; Wang et al., 2021[39]). Multimodal approaches – the ones that mix retinal images with clinical and demographic data – usually do better than image-only models. This advantage is clearest in predicting disease progression or reactivation after anti-VEGF treatment (Wu et al., 2025[40]). Researchers have also come up with explainable AI methods. These provide clearer interpretations of ROP stages and key pathological features, which helps increase trust among clinicians and makes adoption easier (Wang et al., 2021[39]). On the other hand, systematic reviews make it clear that many of these models still have drawbacks: not enough external validation, datasets that are too small, and limited inclusion of wider clinical factors. All this points to a continuing need for stronger, more widely applicable, and truly interpretable multimodal systems (Jafarizadeh et al., 2025[41]). The present study introduces an explainable multimodal deep learning model. It combines retinal images with a broad range of clinical and demographic variables – things like infant sex, parental education, pregnancy and birth characteristics, neonatal factors (including NICU stay duration), and eye-specific severity measures – to predict treatment-requiring ROP. The model is designed to enhance risk stratification, allow earlier detection of high-risk cases, support timely treatment, minimize unnecessary examinations, reduce healthcare system burden, and ultimately improve care quality and blindness prevention in this vulnerable population.

Methods

1. Study Design and Participant Cohort

This retrospective cohort study was conducted at Khatam-al-Anbia Ophthalmology Hospital, the primary tertiary referral center for neonatal ophthalmic care in northeastern Iran. The study was approved by the Mashhad University of Medical Sciences Ethics Committee (IR.MUMS.MEDICAL.REC.1403.338) and adhered to the tenets of the Declaration of Helsinki. Infants were systematically referred for retinopathy of prematurity screening according to national guidelines, which specify eligibility as gestational age < 34 weeks and/or birth weight ≤ 2000 g. Consecutive infants presenting to the dedicated ROP clinic between January 2021 and December 2024 were included. All infants underwent standardized comprehensive ROP evaluation using the RetCam imaging system, documenting ROP stage, zone, and presence of plus disease. Infants with complete evaluation data were eligible; those with unresolved missing data on key variables were excluded. Treatment eligibility (laser photocoagulation or intravitreal anti-VEGF injection) was independently assessed by two senior pediatric ophthalmologists with > 10 years of experience in ROP management. Discrepancies were adjudicated by a third senior ophthalmologist, yielding a kappa coefficient of 0.89 (95% CI 0.85–0.93). From the screened population, 400 treatment-requiring infants were identified. A comparison group of 400 infants without treatment indication was randomly selected. Following data quality assessment, the final analytical cohort comprised 384 infants (203 treated, 181 untreated) with complete multimodal datasets. Written informed consent was waived by the ethics committee due to the retrospective design and use of fully anonymized data.

2. Data Acquisition and Preprocessing

2.1. Clinical and Demographic Variables

Clinical and demographic data were extracted from electronic health records using Python 3.12 with pandas. Variables included infant sex, twin status, ROP features at first visit (stage, zone, and plus disease), gestational age and birth weight, age and weight at first ROP visit, NICU stay duration, parental ages at birth, parental education levels, gravidity, pregnancy initiation mode, pre-pregnancy maternal status, and city of residence. After data cleaning, no missing values remained in the analyzed variables. Numerical variables underwent outlier removal (interquartile range method) and standardization (using StandardScaler fitted on the training set). Categorical variables were label-encoded (using LabelEncoder fitted on the training set). The target variable was binary treatment requirement (1 = treatment required, 0 = no treatment).

2.2. ROP image selection and preprocessing

2.2.1. Image Quality Control

All retinal images underwent a rigorous quality control (QC) procedure prior to inclusion in the analytical pipeline to ensure reliability of downstream segmentation, feature extraction, and classification tasks. The heterogeneous nature of neonatal fundus imaging—characterized by variable illumination, motion artifacts, media opacity, and differences in field-of-view—necessitates stringent QC to prevent poor-quality images from degrading model performance or introducing bias. Therefore, a standardized multi-stage QC framework was implemented to systematically evaluate the suitability of each image. First, all fundus photographs were screened by trained graders for essential diagnostic visibility, including adequate depiction of the posterior pole, optic disc, and major vascular arcades. Images with severe blur, defocus, or motion streaks that obscured vascular structures were excluded. Particular attention was given to vessel clarity and contrast, as vessel segmentation accuracy and subsequent vascular feature extraction depend heavily on the visibility of fine vascular structures. Images with excessive glare, overexposure, or shadowing caused by eyelids or specular reflections were removed if these artifacts interfered with vessel morphology assessment. Cases with incomplete field-of-view that did not capture the posterior pole or displayed only peripheral retina were also excluded, as these images do not provide sufficient information for evaluating plus disease or predicting treatment-requiring ROP. Following the manual grading step, automated quality metrics were applied to ensure consistency across the dataset. These included evaluations of illumination uniformity, contrast distribution, edge sharpness, and signal-to-noise ratio. Images failing predefined

thresholds were flagged for secondary review and excluded when quality was deemed insufficient for reliable vessel segmentation. After completing all QC stages, the four highest-quality fundus images for each eye—captured from four standard directional views—were selected as the final candidates for subsequent processing phases, ensuring that downstream algorithms operated on the most informative and diagnostically reliable representations of the retina. This hybrid QC approach—combining expert review with automated assessment—ensured that only images meeting minimum diagnostic and computational quality requirements were used for model development.

2.2.2 Semi-supervised Vessel Segmentation

Given the scarcity of expertly annotated retinal vessel masks and the abundance of unlabeled ROP fundus images typically encountered in clinical settings, we adopted a semi-supervised segmentation strategy inspired by the Deep Adversarial Network framework of Zhang et al. [42]. As illustrated in Fig. 1, the model employs a U-Net architecture with a ResNet-34 encoder pretrained on ImageNet to enhance feature generalization and training stability. In the supervised warm-up phase, the network was trained using high-quality labeled retinal vessel datasets (DRIVE, STARE, CHASEDB1), with a combined Dice and binary cross-entropy loss.

In the second stage, a dual-branch encoder–decoder model was introduced to jointly process both labeled target-domain images and unlabeled source-domain images. The public datasets were defined as the target domain, representing high-quality labeled vascular structures, whereas the locally acquired ROP images constituted the source domain, reflecting real-world clinical variability. During this phase, the segmentation branch of the network was frozen to preserve the vascular knowledge learned from the target domain, while the reconstruction branch remained trainable to encourage the extraction of domain-invariant retinal representations. Adversarial and reconstruction-based consistency constraints were applied to ensure that the feature distributions of source-domain images gradually aligned with those of the target domain. Importantly, the classifier associated with this stage was trained to output True for target-domain images and False for source-domain images, enabling the model to explicitly learn domain discrimination. This domain-classification signal plays a key role in guiding the encoder toward representations that minimize domain shift, thereby improving robustness in subsequent clinical prediction tasks. In the third stage of the framework, the dual-branch encoder–decoder architecture operates under an adversarial domain-alignment objective. At this point, the reconstruction branch of the network is frozen, preserving the domain-invariant structural representations learned in previous phases, while the segmentation branch remains fully trainable to allow targeted refinement of vessel probability maps for source-domain ROP images. Simultaneously, the domain-classification module of the framework is also frozen, ensuring that its decision boundary remains fixed during this phase. The segmentation branch is then optimized such that its outputs for source-domain images are increasingly likely to be interpreted by the frozen classifier as target-domain segmentation maps. This setup forces the segmentation network to adapt its output distribution toward that of the labeled target-domain datasets, thereby reducing residual domain discrepancies. As a result, the segmentation model learns to generate vessel maps for heterogeneous, real-world ROP images that closely resemble the structured vascular patterns seen in high-quality public datasets, achieving robust domain-invariant segmentation performance.

3. DL model development

As illustrated in Fig. 2, The development of the proposed deep learning framework proceeded in two main stages, beginning with a single-image classification model and subsequently extending to a multi-view architecture capable of leveraging four directional fundus images from each eye. The goal of this staged development was to establish a robust baseline classifier and then enhance it through the incorporation of semi-supervised segmentation priors and multi-task feature extraction.

In the first stage, a standard image-based classifier was trained to predict treatment-requiring ROP from a single posterior-pole fundus image. A convolutional neural network pretrained on ImageNet served as the encoder backbone, enabling the model to benefit from strong visual feature representations despite limited domain-specific data. The classifier was optimized end-to-end using cross-entropy loss, and this single-image model provided an essential baseline for understanding the discriminative retinal features associated with clinically significant ROP.

In the second stage, the framework was expanded to incorporate all four available fundus images captured from different viewpoints around the optic disc. For each view, vessel segmentation probability maps generated from the preceding semi-

supervised segmentation model were used as additional structural inputs. Each fundus image and its corresponding vessel map were passed through one of four encoder streams, all sharing the same architecture but initialized with pretrained weights from the earlier training phase. To accommodate the single-channel vessel segmentation maps while preserving the pretrained structure of the encoder, the first convolutional layer was adapted by computing the channel-wise mean of its pretrained ImageNet filters and applying this averaged kernel to the segmentation input; after this modification of the initial layer, all subsequent convolutional blocks remained identical to the pretrained backbone. Thus, the fundus images were processed using the original three-channel pretrained weights, whereas the segmentation masks were integrated through a minimally adjusted first layer followed by the same deeper encoder architecture.

4. Multimodal Fusion and Machine Learning Evaluation

The multimodal fusion framework integrates four directional fundus images and their corresponding vessel segmentation maps with structured demographic and clinical information to generate comprehensive representations for predicting treatment-requiring ROP. Each fundus image is paired with its vessel probability map and processed jointly through one of four parallel encoder streams with shared architecture. These encoders are initialized with weights pretrained in earlier model development stages, allowing extraction of stable, vessel-aware feature embeddings across retinal viewpoints. Independent processing of each directional view captures spatially distributed vascular cues—including tortuosity, dilation, branching complexity, and disc-centered geometry—that vary with camera angle and illumination conditions. The four feature vectors are projected into a shared latent space and fused using a multi-head attention module that adaptively weights the contribution of each view, prioritizing informative perspectives while suppressing noisier inputs. This produces a robust fused imaging representation. PCA dimensionality reduction is applied to this representation to generate low-dimensional embeddings (PCA1–3).

All clinical and demographic variables included in the study, as well as the composition of the multimodal feature sets, were selected based on clinical expertise from experienced pediatric ophthalmologists. Selection prioritized variables with established or potential associations with ROP severity and treatment requirement.

The fused imaging representation (including PCA1–3) is concatenated with clinical/demographic variables to form five predefined multimodal feature sets for comparative evaluation:

- Set 1: PCA1–3, ROP stage, zone, plus disease at first visit.
- Set 2: Clinical/demographic variables (birth age/weight, age/weight at first visit, NICU duration, parental age/education, sex, twin status, pregnancy initiation, pre-pregnancy issues, gravidity, city of residence).
- Set 3: Set 2 + PCA1–3.
- Set 4: Set 2 + ROP stage, zone, plus disease at first visit.
- Set 5: Set 2 + PCA1–3 + ROP stage, zone, plus disease at first visit.
- Set 6: All features (Set 1 + Set 2)

Data were split 80/20 (stratified by target, random_state = 42). Preprocessing was fitted on the training set only.

Given the mild class imbalance in the dataset (Class Ratio 0:1 = 0.89:1, corresponding to approximately 89 non-treatment cases per treatment case), class weighting was applied during model training to mitigate bias toward the majority class without altering the original data distribution. Specifically, the class_weight=balanced parameter was used in tree-based ensembles (Random Forest, Extra Trees, XGBoost, Gradient Boosting, AdaBoost, Decision Tree) and equivalent weighting strategies were implemented in other classifiers (e.g., class_weight=balanced in Logistic Regression and SVM). This approach was preferred over resampling techniques (e.g., oversampling or undersampling) due to its computational efficiency, preservation of the original data distribution, and avoidance of potential artifacts from synthetic sample generation in a clinically sensitive and moderately sized dataset. The following machine learning models were evaluated on each feature set: Random Forest (n_estimators = 200, max_depth = 10, class_weight='balanced'), XGBoost (n_estimators = 150, max_depth = 6, learning_rate = 0.1), Gradient Boosting, Logistic Regression (C = 0.1, penalty='l2'), SVM (kernel='rbf', C = 1.0, gamma='scale'), MLP (hidden_layers=(50,25), alpha = 0.0001), AdaBoost, Extra Trees, Decision Tree, KNN (n_neighbors = 5), Gaussian Naive Bayes, LDA, and QDA. By combining multi-view vascular morphology

with clinical context across different feature sets, the framework enables comparative assessment of multimodal contributions to treatment prediction.

5. Statistical Analysis of Feature Importance

Univariate statistical analyses were performed to assess the importance and association of individual features with treatment requirement. Descriptive statistics (mean, SD, median, min/max, skewness, kurtosis for numerical; frequency/percentage for categorical) were computed. Normality was tested using Shapiro-Wilk/Kolmogorov-Smirnov. Group comparisons used automated test selection: Student's t-test/Welch's t-test for normal data, Mann-Whitney U for non-normal. Categorical associations used χ^2 or Fisher's exact test with Cramér's V. Pearson correlation was calculated for numerical variables with treatment. These statistical tests provided evidence of feature importance without formal automated selection.

Results

1. Semi-supervised Vessel Segmentation

Figures 3 and 4 present qualitative comparisons of retinal vessel segmentation outputs generated by the proposed semi-supervised framework across multiple ROP cases, each containing four directional fundus views. For each sample, the original ROP images are shown in the top row, followed by the vessel maps produced using an unsupervised baseline method and the refined vessel masks obtained through the proposed semi-supervised segmentation model. These visual results demonstrate the substantial improvements achieved when unlabeled ROP images are leveraged through adversarial and reconstruction-based consistency training. Across all samples, the unsupervised segmentation outputs exhibit common failure patterns, including fragmented vessel structures, loss of fine-caliber peripheral vessels, and incomplete delineation near the optic disc region. These limitations are particularly evident in fundus views with uneven illumination, motion blur, or glare—conditions frequently encountered in neonatal imaging. Additionally, unsupervised masks often show excessive noise or spurious edges in low-contrast regions, reducing their suitability for downstream vascular feature extraction. By contrast, the semi-supervised segmentation maps display markedly enhanced anatomical coherence and vessel continuity across all four views. The proposed method successfully recovers subtle vascular branches, improves delineation of tortuous vessels, and reduces background noise, even in images with challenging acquisition conditions. The integration of labeled public fundus datasets during supervised warm-up, followed by domain adaptation on unlabeled ROP images, enables the model to generalize effectively to neonatal retinal morphology, which differs substantially from adult fundus patterns. This is particularly notable in peripheral zones where vessel caliber is small and prone to segmentation errors in conventional unsupervised methods.Moreover, the semi-supervised approach yields consistent vessel structures across all views within each sample, demonstrating the model’s robustness to variations in angle, illumination, and retinal curvature. This cross-view consistency is essential because the segmentation outputs are subsequently used as structural priors in the multi-view classification framework. Accurate preservation of vascular geometry, especially dilation, branching density, and tortuosity is critical for downstream prediction of treatment-requiring ROP. The improved vessel fidelity achieved via semi-supervised segmentation substantially enhances the reliability of extracted vascular biomarkers.

2. Model Interpretability Analysis

To evaluate the clinical validity and transparency of the proposed predictive framework, we performed an extensive interpretability analysis using gradient-based class activation mapping (Grad-CAM) applied to the final multimodal fusion model. The goal of this analysis was to determine whether the model relies on physiologically meaningful retinal features—particularly vascular abnormalities, when predicting treatment-requiring ROP.

Figure 5 presents representative examples of the original fundus images, their corresponding vessel segmentation maps, and the model-generated attention heatmaps. Across all samples, the visual explanations consistently demonstrated a strong alignment between the model’s decision-making process and established clinical hallmarks of severe ROP. In nearly all cases, the Grad-CAM activation maps concentrated on the posterior pole, including the optic disc and major vascular arcades. This corresponds precisely to the anatomical region where plus disease—the strongest predictor of treatment requirement—is assessed by clinicians. Increased tortuosity and dilation of central vessels are central diagnostic indicators in the ICROP classification system,

and the model's selective attention to these regions suggests that it has successfully learned vascular patterns associated with advanced disease. Notably, the heatmaps rarely emphasized peripheral retinal regions, which aligns with clinical practice, as treatment decisions are largely driven by posterior vascular abnormalities rather than the peripheral stage alone. This focal attention on the posterior pole provides strong evidence that the model's predictions are grounded in pathophysiologic cues rather than spurious correlations. Overlaying the heatmaps onto the segmentation maps further confirmed that the model's attention aligns with specific vascular structures rather than non-informative image regions such as illumination artifacts, imaging borders, or peripheral blur. In multiple examples, activation clusters followed the trajectory of dilated or tortuous vessels, indicating that the model leverages vessel morphology—rather than global image color or brightness—to guide its predictions. In cases with highly engorged vessels or asymmetric vascular branching, the heatmaps highlighted exactly those aberrant areas, demonstrating sensitivity to clinically meaningful nuances. For images with milder vascular changes, the attention maps were more diffuse yet still centered around the arcades, reflecting the model's uncertainty in borderline cases—behavior that mimics real-world variability among human graders. Importantly, the interpretability results reveal consistency between the vessel-aware representations learned during the semi-supervised segmentation stage and the downstream classification behavior. Because the fusion model integrates both raw fundus features and segmentation-derived vascular context, the resulting attention maps capture a blend of structural and morphological cues. This demonstrates that the segmentation-guided feature extraction process not only improves predictive accuracy but also enhances interpretability by explicitly anchoring predictions to vascular pathology. The multimodal design, therefore, avoids common pitfalls in deep learning models—such as reliance on non-biological artifacts—and instead replicates clinically grounded reasoning patterns used by expert ophthalmologists.

3. Quantitative Results

3.1 Descriptive Statistics and Univariate Analysis

The study included 384 preterm infants with complete multimodal data (203 requiring treatment and 181 without treatment; 52.9% vs 47.1%). The mean gestational age was 30.94 ± 3.25 weeks (range: 24–40), the mean birth weight was 1555.41 ± 576.58 g, and the mean NICU stay was 24.30 ± 19.18 days. The majority of infants were male (56.2%), singleton (63.8%), and resided in urban areas (59.9%). ROP stage at the first visit ranged from 0 (38.3%) to 3 (18.0%), with 69.8% of infants showing no plus disease. Univariate analyses revealed significant differences between the treatment and non-treatment groups. Infants requiring treatment had significantly lower gestational age (28.75 ± 2.12 vs 33.40 ± 2.43 weeks; $p < 0.001$, Cohen's $r = 0.853$), lower birth weight ($p < 0.001$, large effect size), longer NICU stays ($p < 0.001$), and higher ROP stage and plus disease prevalence ($p < 0.001$, Cramér's $V > 0.80$). Among the imaging features, PCA1 demonstrated the strongest discriminatory power ($r = 0.964$, $p < 0.001$). Demographic variables showed no significant association with treatment requirement.

3.2 Multimodal Model Performance

Fourteen machine learning models were evaluated across five predefined multimodal feature sets incorporating PCA embeddings and/or clinical variables. Five-fold stratified cross-validation was performed on the training set (80% of data) to ensure robust performance estimation. Tree-based models (Extra Trees, Random Forest, XGBoost, LightGBM) consistently outperformed other algorithms across feature sets containing PCA embeddings (Table 1). For instance, Extra Trees and Random Forest on Set 6 (full multimodal) and Set 4 (clinical + PCA1–3) achieved test accuracy of 0.987, AUC-ROC up to 0.999, and F1-score of 0.988, with near-perfect specificity. Cross-validation confirmed the stability of these results, with mean CV accuracy ranging from 0.984 to 0.986 ± 0.007–0.010 in these sets. In contrast, clinical-only sets (Set 3 and Set 5) exhibited substantially lower performance (maximum test accuracy of 0.896 and 0.948, respectively; CV accuracy 0.888–0.942 ± 0.018–0.026), underscoring the critical contribution of vascular morphology features derived from retinal imaging.

Table 1
Performance of top-performing models on multimodal feature sets (test set & 5-fold CV metrics)

Feature Set	Description	Model	Accuracy (mean ± std)	AUC-ROC (mean ± std)	Test Accuracy	Test AUC-ROC	Test F_1-score	Test Sensitivity	Test Specificity
Set 6	Full multimodal (Clinical/demographic + ROP + PCA1–3)	1. Extra Trees	0.986 ± 0.007	0.995 ± 0.004	0.987	0.999	0.988	0.976	1.000
		2. Random Forest	0.985 ± 0.008	0.994 ± 0.005	0.987	0.998	0.988	0.976	1.000
Set 4	Clinical/demographic + PCA1–3	1. Extra Trees	0.984 ± 0.009	0.994 ± 0.005	0.987	0.998	0.988	1.000	0.972
		2. Random Forest	0.983 ± 0.010	0.993 ± 0.006	0.987	0.997	0.988	0.976	1.000
Set 2	PCA1–3 + ROP stage/zone/plus	1. Random Forest	0.985 ± 0.008	0.992 ± 0.005	0.987	0.990	0.988	0.976	1.000
		2. Extra Trees	0.986 ± 0.007	0.993 ± 0.004	0.987	0.992	0.988	0.976	1.000
Set 5	Clinical/demographic + ROP stage/zone/plus	1. XGBoost	0.942 ± 0.018	0.970 ± 0.012	0.948	0.972	0.952	0.976	0.917
		2. Random Forest	0.938 ± 0.020	0.968 ± 0.013	0.935	0.968	0.942	1.000	0.889
Set 3	Clinical/demographic only	1. Random Forest	0.890 ± 0.025	0.935 ± 0.018	0.896	0.938	0.905	0.927	0.861
		2. Extra Trees	0.888 ± 0.026	0.933 ± 0.019	0.896	0.935	0.905	0.927	0.861

3.3 Model Interpretability (SHAP Analysis, Feature importance)

SHAP analysis was performed on the top-performing models (Extra Trees, Random Forest, XGBoost, and LightGBM) across feature Sets 2, 4, 5, and 6. The results consistently ranked PCA1 as the most influential feature (mean absolute SHAP value > 10–12), followed by ROP stage at the first visit and gestational age. Higher values of PCA1 were strongly associated with an increased likelihood of requiring treatment, reflecting more severe vascular abnormalities in the retina. Synergistic interactions were observed between PCA1 and key clinical risk factors, such as low gestational age, indicating that the models effectively captured the combined effects of imaging-derived vascular morphology and established ROP risk factors. These findings confirm that the models' predictions are grounded in clinically meaningful features rather than spurious correlations.

Discussion

The results of this study demonstrate that our explainable multimodal deep learning model achieves high predictive performance for treatment-requiring retinopathy of prematurity, with Extra Trees and Random Forest classifiers yielding a test accuracy of 0.987, AUC-ROC of 0.999, and F1-score of 0.988 on the full multimodal feature set. These outcomes substantially outperform clinical-only models (maximum test accuracy 0.948) and align with or exceed recent AI advancements in ROP prediction, underscoring the value of integrating retinal imaging features—particularly PCA1, which emerged as the most influential predictor in SHAP and Grad-CAM analyses (r = 0.881 with treatment requirement).

Our findings are consistent with prior image-only convolutional neural network (CNN) models, which have reported accuracies exceeding 90% for detecting plus disease or treatment-requiring cases (Tan et al. [43]; Huang et al. [44]; Yenice et al. [45]). For instance, Tan et al. achieved 96.6% sensitivity in plus disease detection using DL on fundal images, while Huang et al. reported up to 96.14% sensitivity for detecting ROP (no ROP vs. stages 1–2). However, multimodal approaches combining imaging with clinical data have consistently shown superior performance, particularly in predicting disease progression or reactivation post-treatment (Wu et al. [40]; He et al. [46]; Engin et al. [47]). Wu et al. developed a fusion model with an AUC of 0.822 for ROP reactivation after anti-VEGF, integrating clinical and imaging data, while He et al. achieved an AUC of 0.853 using DNN in the e-ROP study. Our model extends this paradigm by incorporating semi-supervised vessel segmentation via a U-Net-based adversarial domain-adaptation approach, addressing challenges in heterogeneous neonatal fundus images and achieving robust domain-invariant vascular feature extraction—advantages not fully realized in prior studies like Redd et al. [48], which reported an AUC of 0.960 for detecting type 1 ROP but relied on basic U-Net vessel segmentation without domain adaptation.

The attention mechanism in our multi-view fusion further enhances interpretability by adaptively weighting informative retinal perspectives, a critical factor for clinical adoption. The dominance of PCA1 in SHAP analyses highlights the pivotal role of vascular morphology (tortuosity, dilation, and abnormal branching) in identifying severe ROP, consistent with findings from Krishnan et al. [49], who identified gestational age and device type as key predictors in ML models, and Karkhaneh et al. [50], who reported 85% sensitivity in digital imaging for referral-warranted ROP in an Iranian cohort. Grad-CAM heatmaps confirmed model attention on the posterior pole, aligning with the International Classification of ROP criteria for plus disease—the strongest predictor of treatment need. This biological alignment addresses a key limitation of many DL models: reliance on non-biological artifacts, as noted in systematic reviews (Jafarizadeh et al.[41]).

Compared to prior multimodal ROP models, our approach offers several advantages. While Takeda et al. [51] achieved AUCs of 0.93–0.94 using non-imaging ML for ROP occurrence prediction, and Shoeibi et al.[52] reported 96% accuracy in XGBoost for treatment-needed ROP, these studies often lack external validation, comprehensive clinical integration, or focus on Iranian-specific cohorts (e.g., Karkhaneh et al.). By contrast, our model incorporates a broad range of demographic and neonatal variables (e.g., gestational age, NICU duration) alongside vascular features, achieving near-perfect specificity (up to 1.000) and minimizing false positives—an essential consideration for reducing unnecessary examinations in vulnerable neonates, as emphasized by Engin et al. The use of 5-fold stratified cross-validation on the training set further ensures robustness and generalizability within our cohort, surpassing the internal validation in Wu et al.

Despite these strengths, several limitations must be acknowledged. First, the study was conducted at a single tertiary center in northeastern Iran, potentially limiting generalizability to other populations or settings with different ROP prevalence and screening practices, as highlighted in reviews by Jafarizadeh et al. External validation on multicenter or international cohorts—such as those in the e-ROP study (He et al.) or global datasets (Krishnan et al.)—is essential to confirm performance across diverse ethnicities and healthcare systems. Second, while the dataset was rigorously quality-controlled, neonatal fundus imaging remains technically challenging due to motion artifacts and variable illumination; future work could explore additional preprocessing techniques or larger datasets to further enhance robustness. Third, although SHAP and Grad-CAM provided strong interpretability, prospective clinical studies are needed to evaluate the model's real-world impact on screening efficiency and blindness prevention. The clinical implications of this work are significant. Current ROP screening guidelines rely primarily on gestational age and birth weight, resulting in frequent examinations of infants who ultimately do not require treatment. By achieving high accuracy and near-perfect specificity, our model could enable risk stratification, allowing clinicians to prioritize high-risk cases for timely intervention while reducing unnecessary retinal examinations—echoing benefits seen in non-imaging models by Takeda et al. This approach has the potential to decrease infant stress, alleviate parental burden, and optimize resource allocation in neonatal intensive care units, particularly in resource-limited settings like Iran.In **conclusion**, this explainable multimodal deep learning model represents a promising tool for early prediction of treatment-requiring ROP. By integrating vascular morphology from retinal images with clinical data, the framework not only achieves superior predictive performance but also provides interpretable insights aligned with clinical reasoning. Future multicenter validation and prospective implementation studies, building on global AI advancements (Jafarizadeh et al.), are warranted to translate these findings into routine clinical practice, ultimately contributing to improved outcomes for preterm infants at risk of vision loss.

Abbreviations

- AI: Artificial Intelligence
- AUC-ROC: Area Under the Receiver Operating Characteristic Curve
- CNN: Convolutional Neural Network
- DL: Deep Learning
- Grad-CAM: Gradient-weighted Class Activation Mapping
- ICROP: International Classification of Retinopathy of Prematurity
- ML: Machine Learning
- NICU: Neonatal Intensive Care Unit
- PCA: Principal Component Analysis
- QC: Quality Control
- ResNet: Residual Network
- RetCam: RetCam Imaging System
- RF: Random Forest
- ROP: Retinopathy of Prematurity
- SHAP: SHapley Additive exPlanations
- SVM: Support Vector Machine
- U-Net: U-shaped Network
- VEGF: Vascular Endothelial Growth Factor
- XGBoost: eXtreme Gradient Boosting

Declarations

Ethics approval and consent to participate

This study was approved by the Mashhad University of Medical Sciences Ethics Committee (approval number: IR.MUMS.MEDICAL.REC.1403.338) and adhered to the tenets of the Declaration of Helsinki. Written informed consent was waived due to the retrospective nature of the study and the use of anonymized data.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This study received no specific funding from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contribution

All authors contributed equally to the study, including conceptualization, data collection, analysis, interpretation, and manuscript preparation. FS, FBA, NS, SE, and AAT each participated in all aspects of the research and approved the final manuscript.

Acknowledgements

Not applicable.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request. The complete analysis pipeline code is also available upon request from the corresponding author.

References

1. Kong L, Fry M, Al-Samarraie M, Gilbert C, Steinkuller PG. An update on progress and the changing epidemiology of causes of childhood blindness worldwide. J aapos. 2012;16(6):501–7.
2. Du Y, Dang Y. Recent Advances in Understanding the Role of PEST Sequence-Containing Proteins in Retinal Neovascularization. Curr Drug Targets. 2025.
3. Kim SJ, Port AD, Swan R, Campbell JP, Chan RVP, Chiang MF. Retinopathy of prematurity: a review of risk factors and their clinical significance. Surv Ophthalmol. 2018;63(5):618–37.
4. Kaur K, Mikes BA. Retinopathy of Prematurity. StatPearls. Treasure Island (FL): StatPearls Publishing Copyright © 2025. StatPearls Publishing LLC.; 2025.
5. Karkhaneh R, Shokravi N. Assessment of retinopathy of prematurity among 150 premature neonates in Farabi Eye Hospital. Acta Medica Iranica. 2001:35–8.
6. Nakshab M, Bayani GH, Ëshaghi M. Prevalence of retinopathy in premature neonates in neonatal intensive care unit of Boali sina hospital in 2001. J Mazandaran Univ Med Sci. 2003;13(39):63–70.
7. Karkhaneh R, Riazi Esfahani M, Ghojehzadeh L, Kadivar M, Nayeri F, Chams H. Nili Ahmadabadi M. Incidence and risk factors of retinopathy of prematurity. Bina J Ophthalmol. 2005;11(1):81–90.
8. Mansouri MR, Kadivar M, Karkhaneh R, Riazi Esfahani M, Nili Ahmadabadi M, Faghihi H, Mirshahi A, Sadat-Nayeri F, Farahvash MS, Tabatabaei A, Adelpour A. Prevalence and risk factors of retinopathy of prematurity in very low birth weight or low gestational age infants. Bina J Ophthalmol. 2007;12(4):428–34.
9. Naderian G, Iranpour R, Mohammadizadeh M, Fazel Najafabadi F, Badiei Z, Naseri F, et al. The frequency of retinopathy of prematurity in premature infants referred to an ophthalmology clinic in Isfahan. J Isfahan Med Sch. 2011;29(128):126–30.
10. Fayyazi A, Heidarzadeh M, Fayzalahzadeh H, Golzar A, Sadegi K. Prevalence of retinopathy of prematurity in preterm infant hospitalized in Tabriz Alzahra Hospital's NICU. Med J tabriz Univ Med Sci. 2009;30(4):63–6.
11. Sadeghi K, Heidary E, Hashemi F, Heydarzadeh M. Incidence and risk factors of retinopathy of prematurity. Med J Tabriz Univ Med Sci. 2009;30(2):73–7.
12. Riazi-Esfahani M, Alizadeh Y, Karkhaneh R, Mansouri MR, Kadivar M, Ahmadabadi MN, Nayeri F. Retinopathy of prematurity: single versus multiple-birth pregnancies. J ophthalmic Vis Res. 2008;3(1):47.
13. Karkhaneh R, Mousavi SZ, Riazi-Esfahani M, Ebrahimzadeh SA, Roohipoor R, Kadivar M, Ghalichi L, Mohammadi SF, Mansouri MR. Incidence and risk factors of retinopathy of prematurity in a tertiary eye hospital in Tehran. Br J Ophthalmol. 2008;92(11):1446–9.
14. Ghalichi L. Incidence, severity and risk factors for retinopathy of prematurity in premature infants with late retinal examination. Bina J Ophthalmol. 2008 Jul 10.
15. Khatami SF, YOUSEFI AE, FATEHI BG, MAMOURI GA. Retinopathy of prematurity among 1000–2000 gram birth weight newborn infants.
16. Naderian G, Parvaresh M, Rismanchiyan A, Sajadi V. Refractive errors after laser therapy for retinopathy of prematurity. Bina J Ophthalmol. 2009;15(1):13–8.
17. Fouladinejad M, Motahari MM, Gharib MH, Sheishari F, Soltani MO. The prevalence, intensity and some risk factors of retinopathy of premature newborns in Taleghani Hospital, Gorgan, Iran.
18. Saeidi R, Hashemzadeh A, Ahmadi S. RAHMANI S. Prevalence and predisposing factors of retinopathy of prematurity in very low-birth-weight infants discharged from NICU.

19. Mousavi SZ, Karkhaneh R, Riazi-Esfahani M, Mansouri MR, Roohipoor R, Ghalichi L, Kadivar M, Nili-Ahmadabadi M, Naieri F. Retinopathy of prematurity in infants with late retinal examination. J ophthalmic Vis Res. 2009;4(1):24.
20. NADERIAN GA, MOULAVI VH, Hadipour M, Sajadi V. Prevalence and risk factors for retinopathy of prematurity in Isfahan.
21. Bayat-Mokhtari M, Pishva N, Attarzadeh A, Hosseini H, Pourarian S. Incidence and risk factors of retinopathy of prematurity among preterm infants in Shiraz/Iran. Iran J Pediatr. 2010;20(3):303.
22. Ebrahim M, Ahmad RS, Mohammad M. Incidence and risk factors of retinopathy of prematurity in Babol, North of Iran. Ophthalmic Epidemiol. 2010;17(3):166–70.
23. MOUSAVI SZ, RIAZI EM, Rouhipour R, JABARVAND M, Ghalichi L, NILI AM, GHASEMI F, AALAMI HZ, Karkhaneh R. Characteristics of advanced stages of retinopathy of prematurity.
24. Ghaseminejad A, Niknafs P. Distribution of retinopathy of prematurity and its risk factors. Iran J Pediatr. 2011;21(2):209.
25. Afarid M, Hosseini H, Abtahi B. Screening for retinopathy of prematurity in South of Iran. Middle East Afr J Ophthalmol. 2012;19(3):277–81.
26. Feghhi M, Altayeb SM, Haghi F, Kasiri A, Farahi F, Dehdashtyan M, Movasaghi M, Rahim F. Incidence of retinopathy of prematurity and risk factors in the south-western region of Iran. Middle East Afr J Ophthalmol. 2012;19(1):101–6.
27. Abrishami M, Maemori GA, Boskabadi H, Yaeghobi Z, Mafı-Nejad S, Abrishami M. Incidence and risk factors of retinopathy of prematurity in mashhad, northeast iran. Iran Red Crescent Med J. 2013;15(3):229.
28. Kazem Sabzehei M, Afje h SA, Farahani AD, Shamshiri AR, Esmaili F. Retinopathy of Prematurity: Incidence, Risk Factors, and Outcome. Archives Iran Med (AIM). 2013;16(9).
29. Ahmadpour-kacho M, Pasha YZ, Rasoulinejad SA, Hajiahmadi M, Pourdad P. Correlation between retinopathy of prematurity and clinical risk index for babies score. Tehran Univ Med J. 2014;72(6).
30. Rasoulinejad SA, Montazeri M. Retinopathy of prematurity in neonates and its risk factors: a seven year study in northern Iran. Open Ophthalmol J. 2016;10:17.
31. Broxterman EC, Hug DA. Retinopathy of Prematurity: A Review of Current Screening Guidelines and Treatment Options. Mo Med. 2016;113(3):187–90.
32. Kaur K, Gurnani B, Kannusamy V, Yadalla D. A tale of orbital cellulitis and retinopathy of prematurity in an infant: First case report. Eur J Ophthalmol. 2022;32(6):Np20–3.
33. Hutchinson AK, Melia M, Yang MB, VanderVeen DK, Wilson LB, Lambert SR. Clinical Models and Algorithms for the Prediction of Retinopathy of Prematurity: A Report by the American Academy of Ophthalmology. Ophthalmology. 2016;123(4):804–16.
34. Palmer EA, Flynn JT, Hardy RJ, Phelps DL, Phillips CL, Schaffer DB, et al. Incidence and early course of retinopathy of prematurity. The Cryotherapy for Retinopathy of Prematurity Cooperative Group. Ophthalmology. 1991;98(11):1628–40.
35. Revised indications for the. treatment of retinopathy of prematurity: results of the early treatment for retinopathy of prematurity randomized trial. Arch Ophthalmol. 2003;121(12):1684–94.
36. Ophthalmology So P, AAo O, AAo. Ophthalmology AAfP, Strabismus. Screening Examination of Premature Infants for Retinopathy of Prematurity. Pediatrics. 2006;117(2):572–6.
37. Desai S, Athikarisamy SE, Lundgren P, Simmer K, Lam GC. Validation of WINROP (online prediction model) to identify severe retinopathy of prematurity (ROP) in an Australian preterm population: a retrospective study. Eye (Lond). 2021;35(5):1334–9.
38. Shoeibi N, Ameri N, Hoseinkhani MR, Ansari Astaneh MR, Motamed Shariati M, Hosseini SM, Abrishami M, Abrishami M, Zamani G, Heidarzadeh HR. Deep learning algorithms for timely diagnosis of retinopathy of prematurity requiring treatment. Eye 2025 Nov 3:1–8.
39. Wang J, Ji J, Zhang M, Lin JW, Zhang G, Gong W, Cen LP, Lu Y, Huang X, Huang D, Li T. Automated explainable multidimensional deep learning platform of retinal images for retinopathy of prematurity screening. JAMA Netw open. 2021;4(5):e218758.
40. Wu R, Zhang Y, Huang P, Xie Y, Wang J, Wang S, Lin Q, Bai Y, Feng S, Cai N, Lu X. Prediction of Reactivation After Antivascular Endothelial Growth Factor Monotherapy for Retinopathy of Prematurity: Multimodal Machine Learning Model Study. J Med Internet Res. 2025;27:e60367.

41. Jafarizadeh A, Maleki SF, Pouya P, Sobhi N, Abdollahi M, Pedrammehr S, Lim CP, Asadi H, Alizadehsani R, Tan RS, Islam SM. Current and future roles of artificial intelligence in retinopathy of prematurity. Artif Intell Rev. 2025;58(6):1–55.
42. Zhang Y, Yang L, Chen J, Fredericksen M, Hughes DP, Chen DZ. Deep adversarial networks for biomedical image segmentation utilizing unannotated images. InInternational conference on medical image computing and computer-assisted intervention. 2017 Sep 4 (pp. 408–416). Cham: Springer International Publishing.
43. Ταν Z, Siμkiv S, Lαi C, Dαi S. Deeπ leαrvivg αlgoριτhμ φορ αυτοματεδ διαγνοσισ οφ ρετινοπατhy οφ πρεματυριτy πlυσ δισεασε. Τρανσlατιοναl visιov σcievce τechvology. 2019;8(6):23.
44. Huang YP, Basanta H, Kang EY, Chen KJ, Hwang YS, Lai CC, Campbell JP, Chiang MF, Chan RV, Kusaka S, Fukushima Y. Automated detection of early-stage ROP using a deep convolutional neural network. Br J Ophthalmol. 2021;105(8):1099–103.
45. Kıran Yenice E, Kara C, Erdaş ÇB. Automated detection of type 1 ROP, type 2 ROP and A-ROP based on deep learning. Eye. 2024;38(13):2644–8.
46. He D, Luo X, Ying B, Quinn GE, Baumritter A, Chen Y, Ying GS, He L. Machine learning models for predicting treatment-requiring retinopathy of prematurity in the e-ROP study. Translational Vis Sci Technol. 2025;14(8):14.
47. Durmaz Engin C, Ozturk T, Ozkan O, Oztas A, Selver MA, Tuzun F. Prediction of retinopathy of prematurity development and treatment need with machine learning models. BMC Ophthalmol. 2025;25(1):1–2.
48. Redd TK, Campbell JP, Brown JM, Kim SJ, Ostmo S, Chan RVP, et al. Evaluation of a deep learning image assessment system for detecting severe retinopathy of prematurity. Br J Ophthalmol. 2019;103(5):580–4.
49. Κριsηναν Α, Sαlμαν Sαρκαρ Μ, Βαδανατη L. Μαchive Leαρvivg Αππlicατιονσ iv Ρετινοπατhy οφ Ρρεματυριτy Diαgνοσισ Uσivg τhe ROP Retiναl Iμage Δατασετ. J Neοναταl Supg [Ιντερνετ]. 2025 Φεβ. 7 [citeδ 2025 Dec. 28];14(1S):820-7. Αναilαβle φρομ: hττπσ://jveοναταlσupg.coμ/ivδex.πhπ/jvσ/αρτicle/vieω/1607
50. Καρkhαveh R, Ahμαδραji A, Εσφαhανi MR, Roohiπουρ R, Δαστjανi ΑΦ, Iμανi M, Khοδαβανδε A, Εβραhiμiαδiβ N, Ahμαδαβαδi MN. The αccupαcy οφ δigiταl iμagivg iv διαgνοσισ οφ ρετινοπατhy οφ πρεματυριτy iv Ιραν: α πilοτ στυδy. Jουρναl οφ Οπhτhαlμic & Viσiov Reσεαρch. 2019 Jαν;14(1):38.
51. Takeda Y, Kaneko Y, Sugimoto M, Yamashita H, Sasaki A, Mitsui T. Prediction Models for Retinopathy of Prematurity Using Nonimaging Machine Learning Approaches: A Regional Multicenter Study. Ophthalmol Sci. 2025;5(4):100715.
52. Shoeibi N, Abrishami M, Hosseini SM, Ansari-Astaneh MR, Farrahi R, Gharib B, et al. Development and validation of machine learning classifiers for predicting treatment-needed retinopathy of prematurity. BMC Med Inf Decis Mak. 2025;25(1):221.

Figures

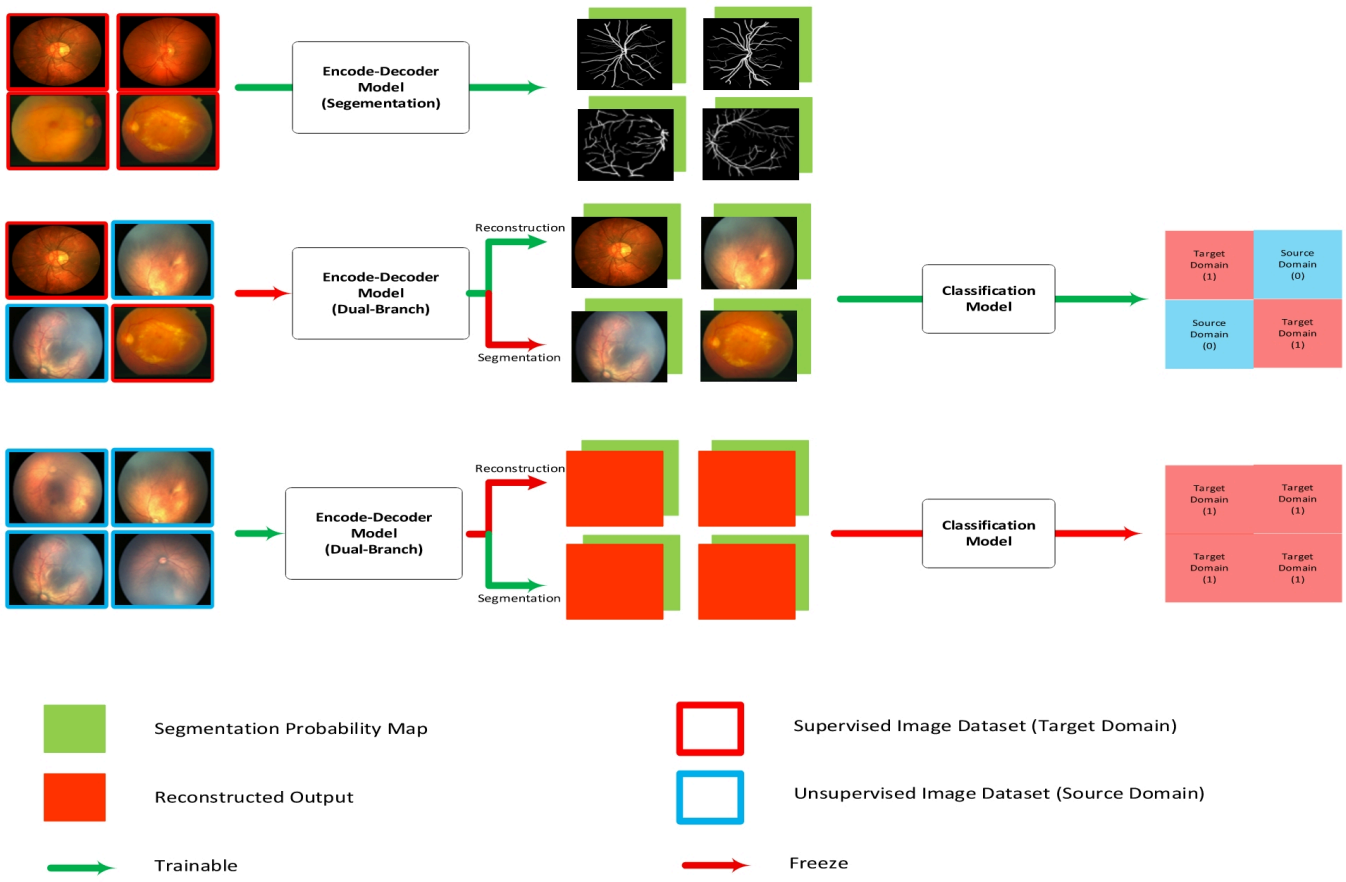


Figure 1

Overview of the proposed semi-supervised learning framework for ROP analysis. This pipeline leverages segmentation, reconstruction, and adversarial consistency to learn vessel-aware features for predicting treatment-requiring ROP. Training proceeds in three stages: supervised segmentation, semi-supervised representation learning with frozen segmentation weights, and classifier training using the learned features while the reconstruction branch remains fixed. Green arrows indicate trainable components, red arrows denote frozen modules, and colored frames represent target-domain labeled (red) and source-domain unlabeled (blue) images.

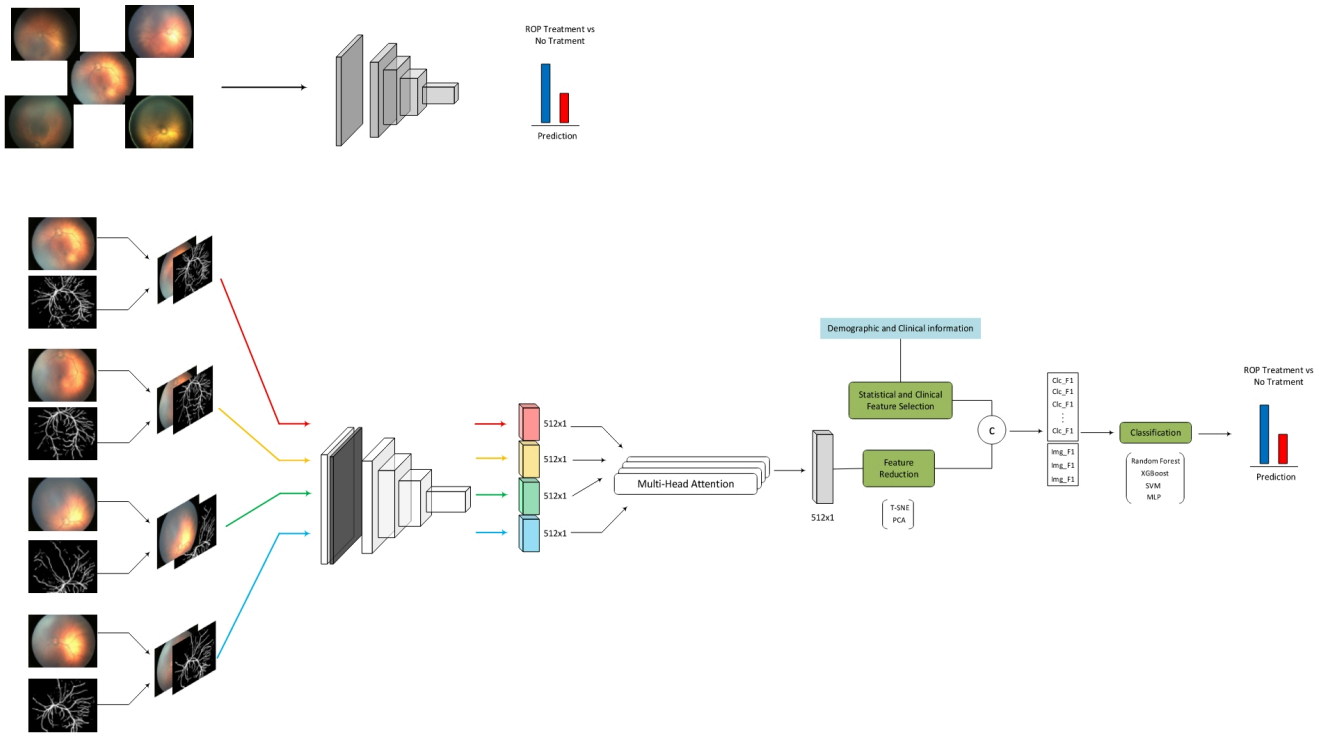


Figure 2

Deep learning model development pipeline: The framework is built in two stages: (1) training a baseline single-image classifier using an ImageNet-pretrained encoder, and (2) extending the model to a multi-view architecture that processes four directional fundus images and their corresponding vessel segmentation maps through parallel pretrained encoders. Single-channel segmentation inputs are integrated by adapting the first convolutional layer via channel-wise kernel averaging, enabling consistent vessel-aware feature extraction across views.

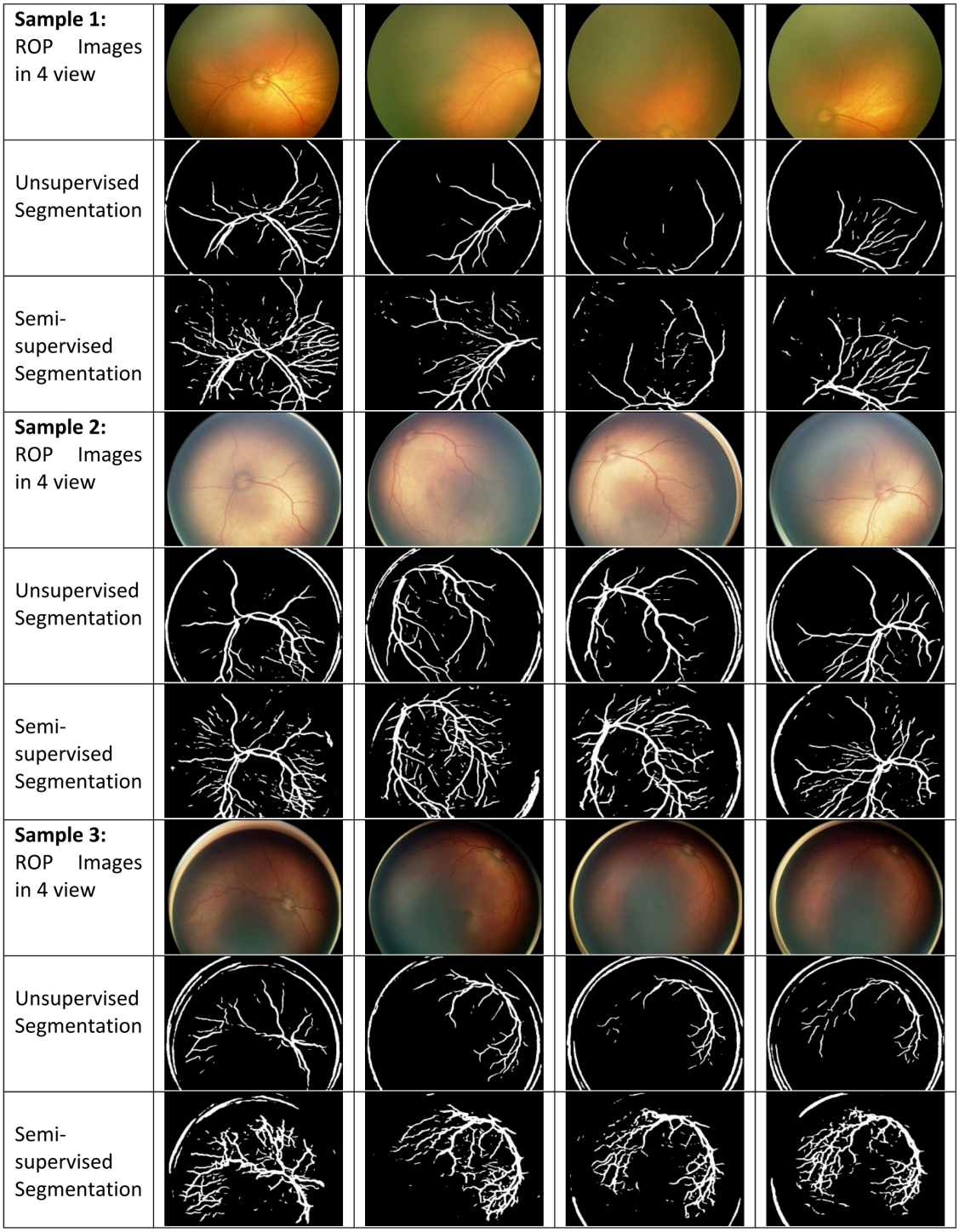


Figure 3

Examples of treatment-requiring ROP cases shown across four fundus views. Each column represents one of the four directional fundus views. The rows, from top to bottom, correspond to the original fundus images, the unsupervised segmentation outputs, and the semi-supervised segmentation results.

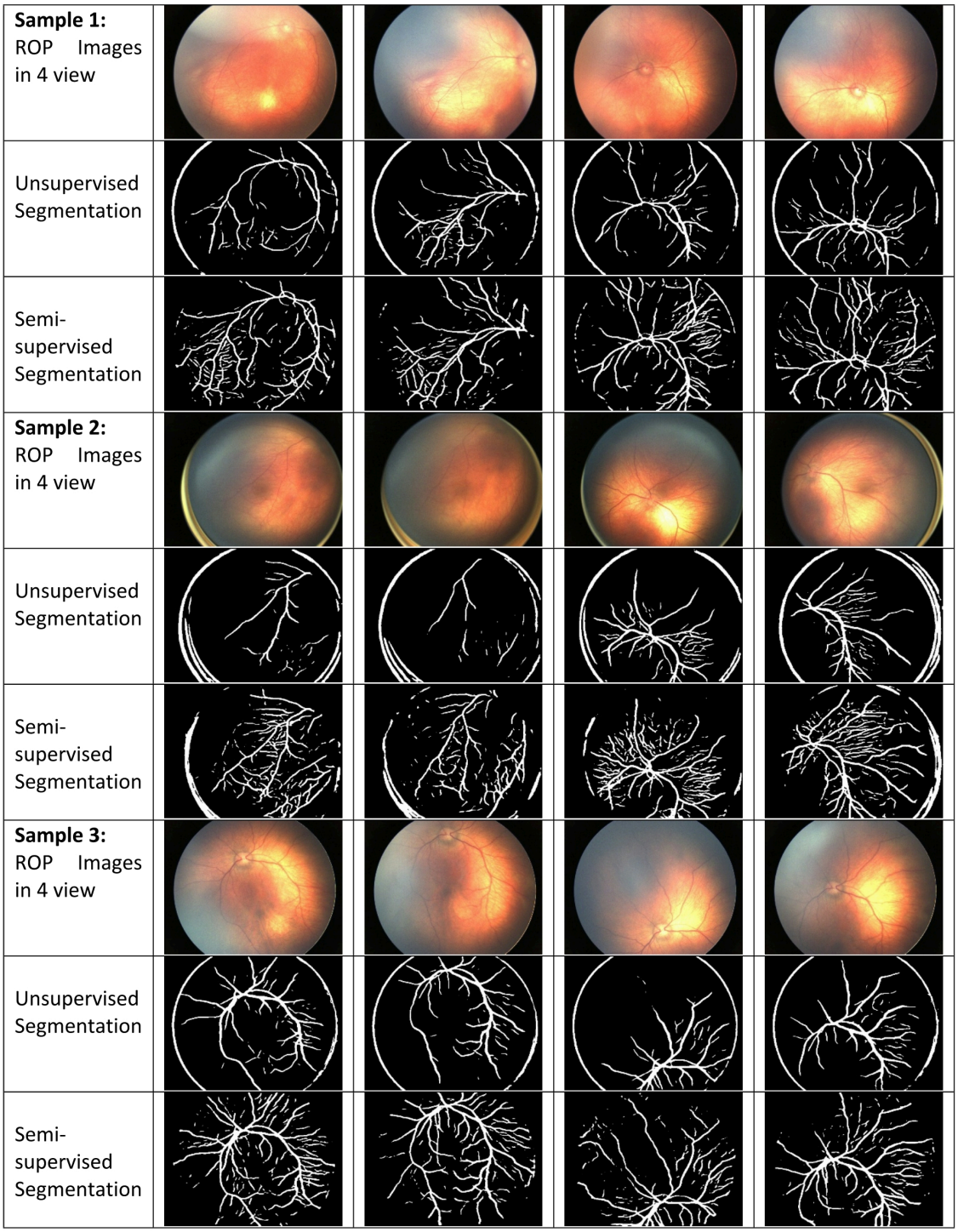


Figure 4

Examples of non–treatment-requiring ROP cases in four views. Each column represents one of the four directional fundus views. The rows, from top to bottom, correspond to the original fundus images, the unsupervised segmentation outputs, and the semi-supervised segmentation results.

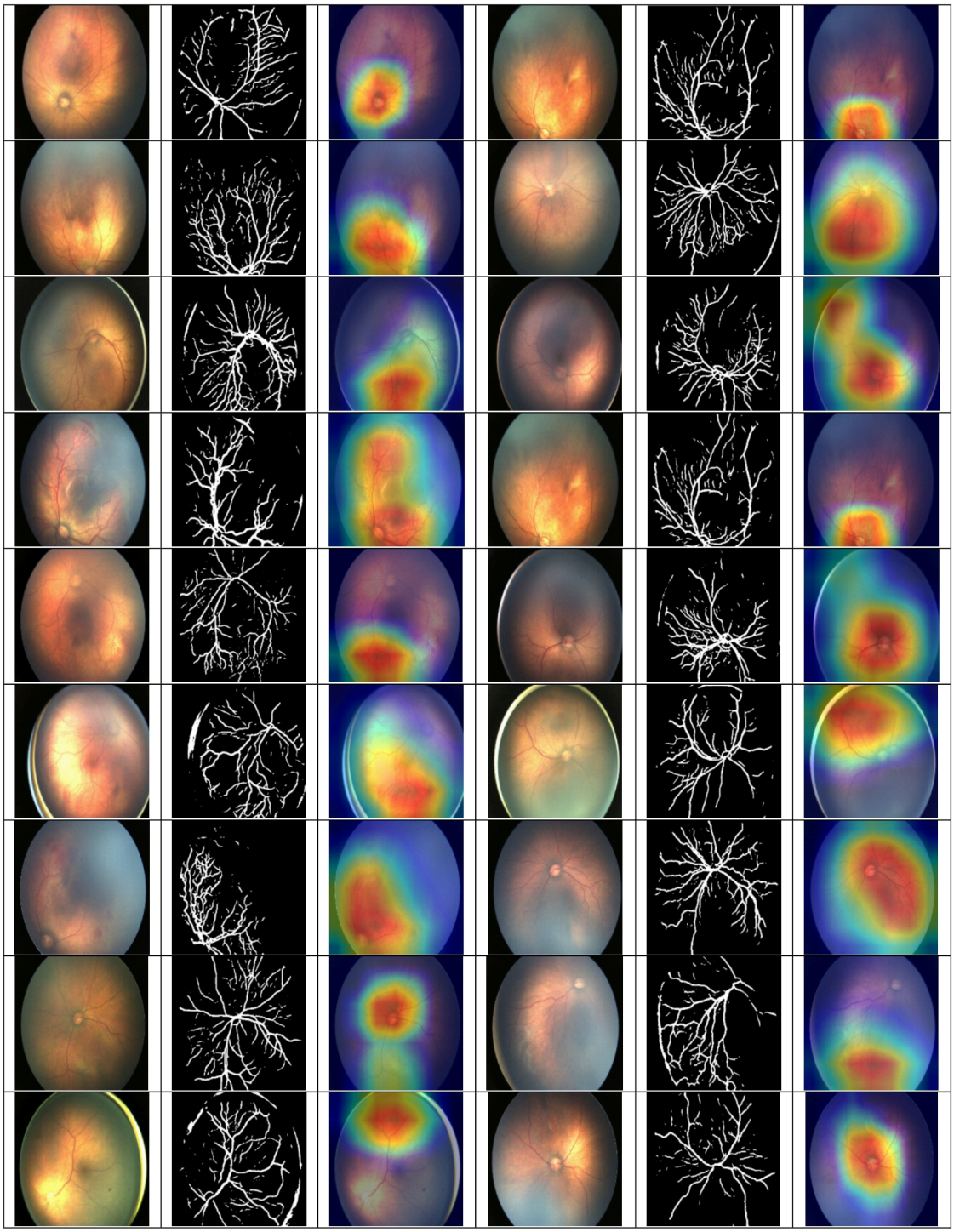


Figure 5

Model interpretability analysis. Interpretability heatmaps highlight the retinal regions most influential to the model’s predictions.