scientific reports

Check for updates

OPEN

Automated detection and segmentation of Weiss ring in fundus photography images using deep learning

Heesuk Kim[1], Sun Young Ryu[2], Tae Keun Yoo[3,4]✉ & Daniel Duck-Jin Hwang[3,5]

We developed and evaluated a deep learning system integrating segmentation and classification for automated detection of Weiss rings on fundus photographs (FPs). A U-Net segmenter and two EfficientNet-B0 classifiers were trained, and their probability outputs were fused by a concatenation meta-classifier. Performance was assessed on an independent test set. Segmentation was evaluated with Dice similarity coefficient and intersection-over-union (IoU); classification with area under the receiver-operating-characteristic curve (AUC), accuracy, sensitivity, and specificity. We used Grad-CAM to provide interpretability and reported inter-observer agreement to contextualize segmentation performance. U-Net achieved Dice 0.578 and IoU 0.421 for Weiss-ring localization, consistent with variability observed in low-contrast/peripapillary cases. The integrated meta-classifier outperformed individual CNNs, yielding AUC 0.903, accuracy 0.812, sensitivity 0.692, and specificity 0.872. Attention maps highlighted peripapillary regions of Weiss-ring appearance, supporting model interpretability. Integrating segmentation with classification improved discrimination relative to classification alone. This FP-based tool is not intended to replace clinical examination for posterior vitreous detachment; rather, it may support archival review, education/quality assurance, and research phenotyping, and serve as an adjunct flag when wider-field imaging or OCT is unavailable. Given the limited field of view and variable visibility of Weiss rings on FPs, prospective validation against ultrawide-field and/or OCT reference standards is warranted.

Keywords Weiss ring, Posterior vitreous detachment, Deep learning, Fundus photography, Segmentation, Classification, Artificial intelligence

Abbreviations

AI	Artificial intelligence
AUC	Area under the curve
CI	Confidence interval
CNN	Convolutional neural network
FP	Fundus photograph
IoU	Intersection over union
OCT	Optical coherence tomography
PVD	Posterior vitreous detachment
PPV	Positive predictive value
NPV	Negative predictive value
ROC	Receiver operating characteristic

The posterior vitreous detachment (PVD) process is a common age-related change in the vitreoretinal interface, occurring when the posterior hyaloid membrane separates from the retina[1]. Although PVD is a normal physiological event in most individuals over the age of 60, its onset, particularly when acute, can be associated with retinal tears and, in some cases, rhegmatogenous retinal detachment[2]. Clinical assessment of vitreous status

[1]Institute of Vision Research, Department of Ophthalmology, Yonsei University College of Medicine, Seoul, South Korea. [2]Department of Refractive Surgery, B&VIIT Eye Center, Seoul, South Korea. [3]Department of Ophthalmology, Hangil Eye Hospital, 35 Bupyeong-Daero, Bupyeong-Gu, Incheon 21388, South Korea. [4]Bright St. Mary's Eye Center, Seoul, South Korea. [5]Lux Mind, Incheon, South Korea. ✉email: eyetaekeunyoo@gmail.com; fawoo2@yonsei.ac.kr

is therefore a routine component of ophthalmic examination, especially in patients presenting with symptoms such as new-onset floaters or photopsia. A Weiss ring is a distinctive, ring-shaped vitreous opacity representing peripapillary glial tissue that has detached from the optic nerve head during complete PVD[3]. Its presence is pathognomonic for completed vitreous separation from the optic disc. While frequently encountered as a benign, incidental finding in older adults, a newly observed Weiss ring in a symptomatic patient can indicate recent vitreous detachment, a period of increased susceptibility to retinal breaks[4]. Identifying this feature can therefore have clinical relevance when interpreted in the appropriate context. However, Weiss rings are not observed in all cases of clinically confirmed PVD, and their appearance on color fundus photographs (FPs) can be subtle, variably shaped, or outside the captured field-of-view.

Despite its pathognomonic significance for complete PVD[5], the Weiss ring is often routinely ignored or misclassified in both clinical image review and artificial intelligence (AI)-based retinal image analysis[6]. In current AI development pipelines, the focus has largely been on posterior segment pathologies such as diabetic retinopathy, age-related macular degeneration, or glaucoma[7], with vitreous opacities either excluded from datasets or mislabeled as image artifacts, peripapillary atrophy, or myelinated nerve fibers. This lack of attention to vitreous interface findings means that automated systems are generally unable to recognize Weiss rings or assess vitreous status—leaving a gap in documentation, research, and potential risk stratification for acute PVD-related pathology. This will likely contribute to the potential performance degradation when AI developments in ophthalmic data analysis are implemented in real-world clinical settings.

FP imaging is widely used for documentation and screening of retinal diseases, but evaluation of vitreous status, including Weiss ring detection, remains primarily a clinical endeavor. Weiss rings can present in various forms on FP, including complete or partial rings, ball-like opacities, and holes without a ring, with their morphology influenced by refractive status and typically remaining stable over time after PVD onset[8]. In most FP-based screening programs, vitreous opacities are either not reported or are described in nonspecific terms such as "floaters" without further classification. While PVD has been studied using optical coherence tomography (OCT)[9], which visualizes the vitreoretinal interface, no comparable work has been attempted with FPs. There have been studies diagnosing PVD using OCT images that directly capture the vitreous and retinal borders, but no such attempts have been made using FPs. Deep learning has demonstrated state-of-the-art performance in identifying a wide range of retinal pathologies from FP[10,11], yet its application to vitreoretinal interface abnormalities remains unexplored. Weiss ring detection may present challenges for AI models because the opacity is often subtle, semi-transparent, variably shaped, and can overlap with other peripapillary changes or photographic artifacts, similar to other subtle FP findings such as cotton wool spots. Consequently, automated recognition of Weiss rings on FPs has remained largely unexplored, despite its potential value for retrospective documentation, quality-assurance flagging, and research phenotyping where ultrawide-field retinal imaging/OCT may be unavailable.

In this study, we present a deep learning framework for Weiss ring detection and segmentation in FP images derived from open datasets (Fig. 1). We developed a convolutional neural network (CNN) to classify images

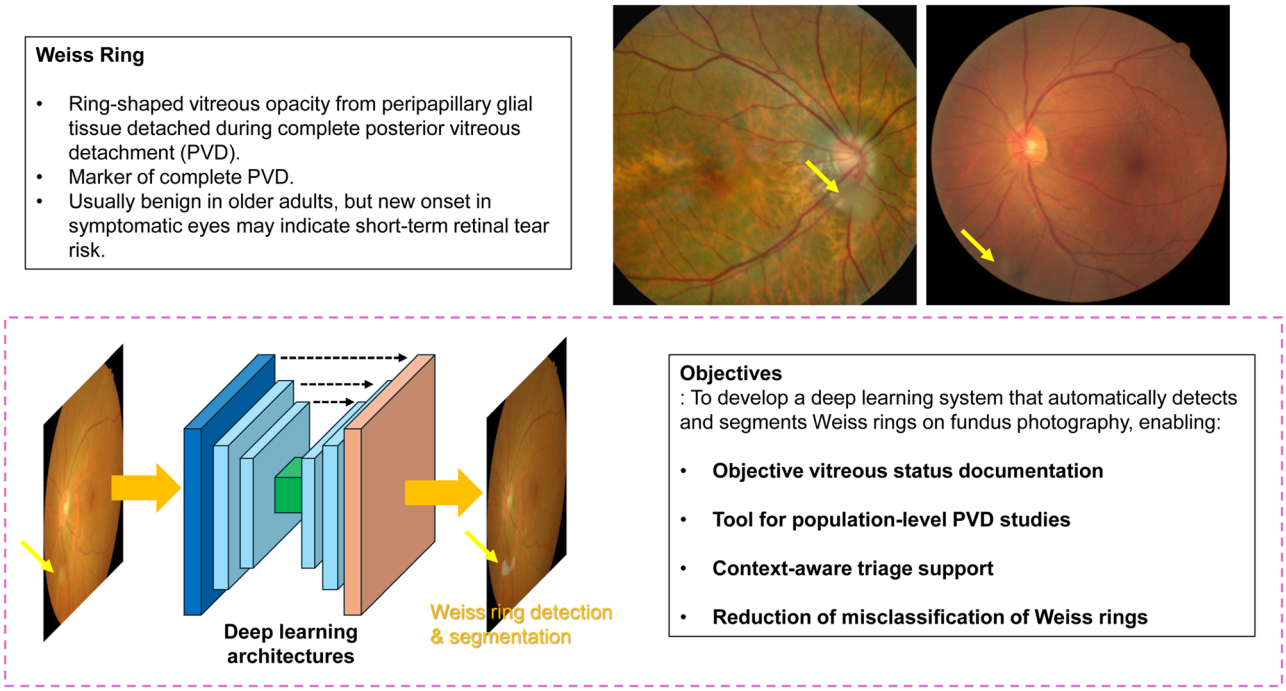

Fig. 1. Example Weiss ring appearance on fundus photography and concept of the deep learning–based detection and segmentation model. Weiss rings are ring-shaped vitreous opacities resulting from complete posterior vitreous detachment (PVD).

as containing or not containing a Weiss ring. We also developed a U-Net segmentation model trained on ophthalmologist-annotated masks to localize the Weiss ring within the image. By combining two deep learning models, we enhanced the Weiss ring detection performance and explainability. To our knowledge, this is the first study to develop and evaluate deep learning models for the identification of Weiss ring from color FPs.

Methods

Dataset

This study was conducted using anonymized data from publicly available FP datasets; therefore, informed consent was waived, and institutional review board (IRB) approval was not required. All procedures adhered to the tenets of the Declaration of Helsinki.

FP images were obtained from multiple publicly accessible datasets encompassing diverse acquisition settings, camera types, and disease categories. The included datasets were ODIR-5 K[12], RFMiD (ISBI 2021 Retinal Image Analysis Challenge)[13], RFMiD 2.0[14], 1000 Fundus (Joint Shantou International Eye Centre, JSIEC)[10], PALM (ISBI 2019 Pathologic Myopia Challenge)[15], and REFUGE (ISBI 2018 Retinal Fundus Glaucoma Challenge)[16], as shown in Table 1. A single ophthalmologist initially screened the entire dataset, after which three ophthalmologists independently reviewed the screened subset to identify and manually extract images containing Weiss rings. Discrepant cases were resolved by majority vote. Uncertain cases (for example, faint ring-like shadows or prominent peripapillary changes mimicking a ring) were excluded from the curated cohort to prioritize label specificity. Supplementary Table S1 lists the IDs for the extracted Weiss ring images. To construct the negative class while preserving source composition and minimizing sampling-induced domain bias, no–Weiss ring images were randomly sampled within each dataset source at a 2:1 ratio relative to Weiss ring–positive images. In total, 582 images were included, comprising 194 Weiss ring–positive and 388 Weiss ring–negative images. Because subject identifiers and fellow-eye linkage were not consistently available across all public datasets, partitioning was performed at the image level rather than the patient level. Accordingly, residual within-subject correlation or fellow-eye leakage cannot be fully excluded and is acknowledged as a limitation.

The full cohort (N = 582) was split at the image level into a development set (80%) and a locked held-out test set (20%) using a source-stratified and class-stratified random procedure to preserve the per-source distribution and the 2:1 class ratio across splits (Supplementary Figure S1). The locked test set contained 39 Weiss ring–positive and 78 Weiss ring–negative images (N = 117) and was defined before model training. It was not used for model selection, hyperparameter tuning, or threshold selection. The remaining images were used for model development, including five-fold cross-validation. Final performance was reported once on the locked test set using the frozen model and a prespecified classification threshold of 0.5.

Model validation

Model performance was evaluated for both Weiss ring detection (classification) and Weiss ring localization (segmentation). To ensure an unbiased final assessment, we predefined a held-out locked test set (20%) prior to any model development. All remaining images (development set; 80%) were used for five-fold cross-validation (CV) to estimate model stability and reduce variance in performance estimates.

For five-fold CV, folds were constructed using a source-stratified and class-stratified procedure to preserve similar proportions of images from each dataset source and the 1:2 class ratio within each fold. In each CV iteration, four folds were used for training and the remaining fold served as the validation fold. Model training was performed using only the training folds. Model selection decisions (e.g., backbone selection and early stopping) were made without reference to the locked test set and were based on CV performance within the development set.

The locked test set was not used for model selection, hyperparameter tuning, or decision-threshold determination. Final performance metrics for the primary evaluation were computed once on the locked test set using the frozen model configuration and a fixed operating threshold determined a priori from the development set.

Dataset	Description	Total N of original dataset	Curated cohort (Weiss / No-Weiss)	Development set (Weiss / No-Weiss)	Locked test set (Weiss / No-Weiss)
ODIR-5 K[12]	Ocular Disease Intelligent Recognition	8,000	84 / 168	67 / 134	17 / 34
RFMiD[13]	ISBI 2021 Retinal Image Analysis for multi-Disease Detection Challenge	3,200	17 / 34	14 / 27	3 / 7
RFMiD 2.0[14]	Multilabel Dataset with 49 rare and frequently observed diseases	860	10 / 20	8 / 16	2 / 4
1000 Fundus[10]	Fundus photographs covering 39 disease classes from JSIEC	1,000	19 / 38	15 / 30	4 / 8
PALM[15]	ISBI 2019 Pathologic Myopia Challenge	1,200	61 / 122	49 / 98	12 / 24
REFUGE[16]	ISBI 2018 Retinal Fundus Glaucoma Challenge	1,200	3 / 6	2 / 5	1 / 1
Total	Combined public FP sources used in this study	15,460	194 / 388 (N = 582)	155 / 310 (N = 465)	39 / 78 (N = 117)

Table 1. Dataset sources and distribution for this study. ISBI = International Symposium on Biomedical Imaging; JSIEC = Joint Shantou International Eye Centre; PALM = Pathologic Myopia; ODIR = Ocular Disease Intelligent Recognition; REFUGE = Retinal Fundus Glaucoma Challenge; RFMiD = Retinal Fundus Multi-disease Image Dataset. For model development, the number of randomly sampled no-Weiss ring images was set to be twice the number of extracted Weiss ring images (1:2 ratio).

Segmentation annotations

Segmentation of Weiss rings was performed using the Computer Vision Annotation Tool (CVAT), which allows precise polygonal annotation and export in machine learning–ready formats[17]. Three board-certified ophthalmologists independently annotated all identified Weiss ring images using polygonal contours. Each annotation precisely delineated the boundary of the Weiss ring. To establish the final ground truth masks, a majority voting strategy was applied: pixels labeled as part of a Weiss ring by at least two ophthalmologists were retained in the reference mask, while others were excluded. This consensus approach minimized inter-observer variability and ensured a robust standard for model evaluation.

Model development

For preprocessing, input resolution was matched to each network's native input size. EfficientNet-B0 classification used 224 × 224 images, whereas the U-Net segmentation model used 256 × 256 images to preserve spatial detail while remaining computationally efficient.

We developed a dual-model framework for automated detection and localization of Weiss rings in FPs (Fig. 2). The workflow integrated both classification and segmentation tasks to enhance robustness. Specifically, two EfficientNet-B0 classifiers and one U-Net segmentation model were trained independently, and their outputs were combined using a logistic regression concatenation meta-classifier to produce the final prediction score. This architecture was designed to provide not only a model score (0–1) for Weiss ring presence but also precise anatomical localization. We implemented model development and training in MATLAB R2025a (MathWorks, Natick, MA, USA), which offers a comprehensive environment with access to various state-of-the-art deep learning architectures and toolboxes for streamlined experimentation (Supplementary Code S1–S2).

For the segmentation task, a U-Net architecture was implemented to localize Weiss rings at the pixel level[18], All fundus images were resized to 256 × 256 pixels and intensity-normalized. To address the strong class imbalance between background and Weiss-ring pixels, class weights were computed from the training masks as the inverse of each class's pixel frequency (stabilized with a small epsilon) and normalized so that rarer class (Weiss ring) received proportionally higher penalty in the loss. The network was optimized with Adam (initial learning rate 1×10^{-3}, L2 regularization 1×10^{-4}, 60 epochs, mini-batch 16), using a class-weighted pixel-wise cross-entropy objective. Standard data augmentation was applied on-the-fly, including random rotations (±15°), horizontal/vertical flips, isotropic scaling (0.9–1.1 ×), and brightness adjustments, to improve generalization across cameras and illumination conditions. Model selection was performed on a held-out validation split with periodic monitoring to prevent overfitting. For U-Net training, we used a fixed learning rate (no decay schedule) and did not apply early stopping.

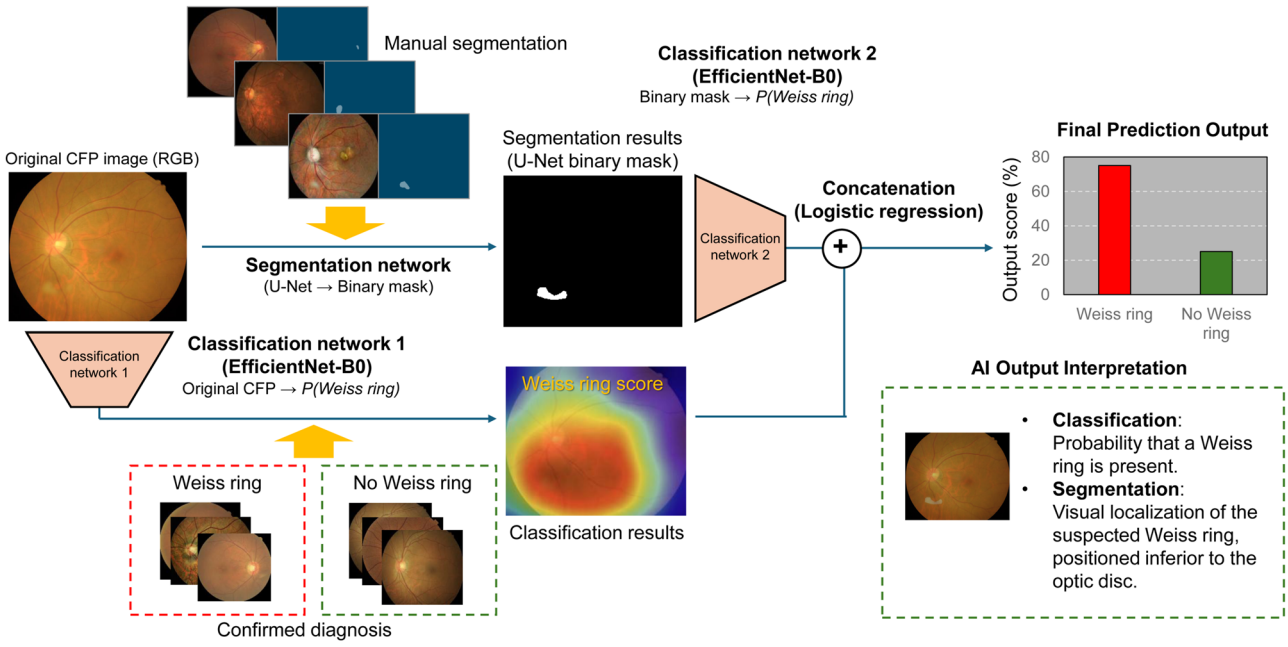

Fig. 2. Dual-model workflow for Weiss ring detection and localization. Each color fundus photograph (RGB) was analyzed by two complementary components. First, Classification Network 1 (EfficientNet-B0) directly processed the original image to output a model score P of Weiss ring presence. Second, a Segmentation Network (U-Net) generated a binary mask to localize candidate Weiss ring regions using ophthalmologist-annotated reference masks. The U-Net binary mask was then provided as the sole input to Classification Network 2 (EfficientNet-B0) to produce an additional score. Finally, the two scores from Network 1 and Network 2 were concatenated and combined using a logistic regression concatenation meta-classifier to yield the final prediction score. Classifier 2 receives only the U-Net–generated binary mask as input and does not use a masked RGB image.

We developed two independent image-level classifiers using an EfficientNet-B0 backbone initialized with ImageNet weights[19], The first classifier was trained directly on labeled FPs (Weiss ring vs. no Weiss ring) to generate image-level scores of Weiss ring presence. The second classifier, trained independently, received U-Net segmentation outputs as input and was optimized on the same labels (Weiss ring vs. no Weiss ring) to determine whether the segmented regions represented true Weiss rings or false positives. For both classifiers, inputs were resized to 224 × 224 pixels and intensity-normalized. To improve robustness, we applied on-the-fly data augmentation consisting of random rotations (± 30°), horizontal reflections, x/y shearing (± 10°), x/y translations (± 30 pixels), and isotropic scaling (0.9–1.1 ×). Grayscale images were converted to three channels to match the backbone's input. Training used stochastic gradient descent with momentum (SGDM) with an initial learning rate of 3×10^{-4}, mini-batch size of 16, and 20 epochs. The objective of optimization was cross-entropy loss. Model selection was monitored on a held-out validation set at regular intervals; the final weights correspond to the epoch with the best validation performance. The learning rate was held constant (no decay schedule), and no early-stopping rule was applied.

To generate the final prediction, we concatenated the outputs of Classification Network 1 (image-level score) and Classification Network 2 (segmentation-derived score) and used them as inputs to a logistic regression meta-classifier (Fig. 2). The concatenation meta-classifier was trained on the development set using the ground truth presence or absence of a Weiss ring as the dependent variable, enabling logistic regression to learn the optimal weighting of these complementary signals. This approach ensured interpretability while improving classification stability, particularly in borderline or noisy cases, and allowed the ensemble to integrate both image-level and segmentation-derived information, thereby enhancing prediction accuracy and robustness[20]. For the i-th fundus photograph x_i, let $p_{1,i}$ denote the probability score from the first classifier and $p_{2,i}$ denote the probability score from the second classifier. These two model outputs were concatenated into a feature vector $\boldsymbol{z}_i = [p_{1,i}, p_{2,i}]^\top$. The final probability of Weiss ring presence was then modeled using logistic regression:

$$P(y_i = 1 \mid \boldsymbol{z}_i) = \sigma(\beta_0 + \beta_1 p_{1,i} + \beta_2 p_{2,i})$$

where $y_i \in \{0,1\}$ is the ground-truth label, β_0 is the intercept, β_1 and β_2 are regression coefficients, and σ is the logistic sigmoid function. The regression coefficients were estimated from the development data by maximizing the likelihood of the observed labels. This logistic regression model therefore learned how to combine the two deep learning prediction scores into a single final probability for Weiss ring detection. The fitted meta-classifier was then applied unchanged to the locked test set. Because out-of-fold predictions were not used for meta-classifier training, this fusion strategy should be interpreted as a simple ensemble approach rather than formal stacked generalization.

Metrics and statistical analysis

Primary segmentation outcomes were evaluated using the Dice similarity coefficient and intersection-over-union (IoU) on the independent test set[21], with precision and recall. Because Weiss ring pixels comprise only a small minority of all pixels, global accuracy can be high even when class-specific precision and recall are modest. Therefore, we interpret global accuracy as a secondary metric alongside Dice/IoU and precision/recall. Dice and IoU were calculated on the Weiss-ring-positive test images only, because these overlap metrics are defined for images containing a target structure. Negative images were excluded from Dice/IoU computation and were instead reflected in the classification analysis. For classification, diagnostic performance was assessed using receiver operating characteristic (ROC) curves and the area under the curve (AUC). We used a prespecified threshold of 0.5 for all binary classification tasks to avoid data-driven threshold selection. Statistical differences between AUCs of competing models were tested using the DeLong method[22], with p-values reported.

Segmentation performance of the proposed method was compared against established architectures, including DeepLabv3 + and SegNet[23,24]. Classification performance was benchmarked against ResNet-50, VGG-19, and ConvNeXt-V2[25,26]. Final predictions from the primary classification network and the concatenated output were reported as scores across the test set. Grad-CAM visualization was applied to qualitatively confirm that the regions highlighted by the models corresponded to clinically relevant features, such as Weiss rings.

Results

In the curated cohort, presence or absence labeling for Weiss ring was fully concordant across the three graders (100% agreement, $\kappa = 1.0$). Inter-rater agreement was quantified using pairwise IoU and Dice among the three graders and versus the majority-vote reference mask (Fig. 3), showing moderate pairwise agreement (IoU 0.722–0.811; Dice 0.837–0.895) and slightly higher agreement against the reference (IoU 0.805–0.822; Dice 0.892–0.901). Qualitatively, most discrepancies occurred in low-contrast or peripapillary rings and near structures that can mimic ring edges (vessel reflexes/peripapillary atrophy).

Representative segmentation examples from the U-Net are shown in Fig. 4. In typical cases, the models produced masks that closely overlapped the ophthalmologist annotations. Errors were most common for small, faint, or peripapillary rings, where models either under-segmented (missed a portion of the ring) or produced false-positive blobs from vessel reflexes, peripapillary atrophy, or photographic artifacts.

Performance on the independent test set was summarized in Table 2. All models achieved very high global pixel-wise accuracy (> 0.99), reflecting consistent background classification (Supplementary Table 2); however, overlap-based metrics provided more discriminative insights. Among the architectures, U-Net achieved the highest overlap, with a Dice score of 0.578 (95% CI 0.484–0.673) and IoU of 0.421 (95% CI 0.324–0.518). U-Net also demonstrated the highest sensitivity, with Recall 0.917 (95% CI 0.859–0.975), although this came at the cost of lower Precision (0.436, 95% CI 0.336–0.535), consistent with a tendency toward over-segmentation. DeepLabv3 + showed comparable performance, with a Dice score of 0.561 (95% CI 0.468–0.654), IoU of 0.404

Fig. 3. Inter-observer agreement for Weiss-ring segmentation. (**A**) Pairwise Intersection-over-Union (IoU; Jaccard index) between the three ophthalmologist graders and the consensus standard reference mask (majority voting). (**B**) Corresponding Dice coefficients for the same pairs.

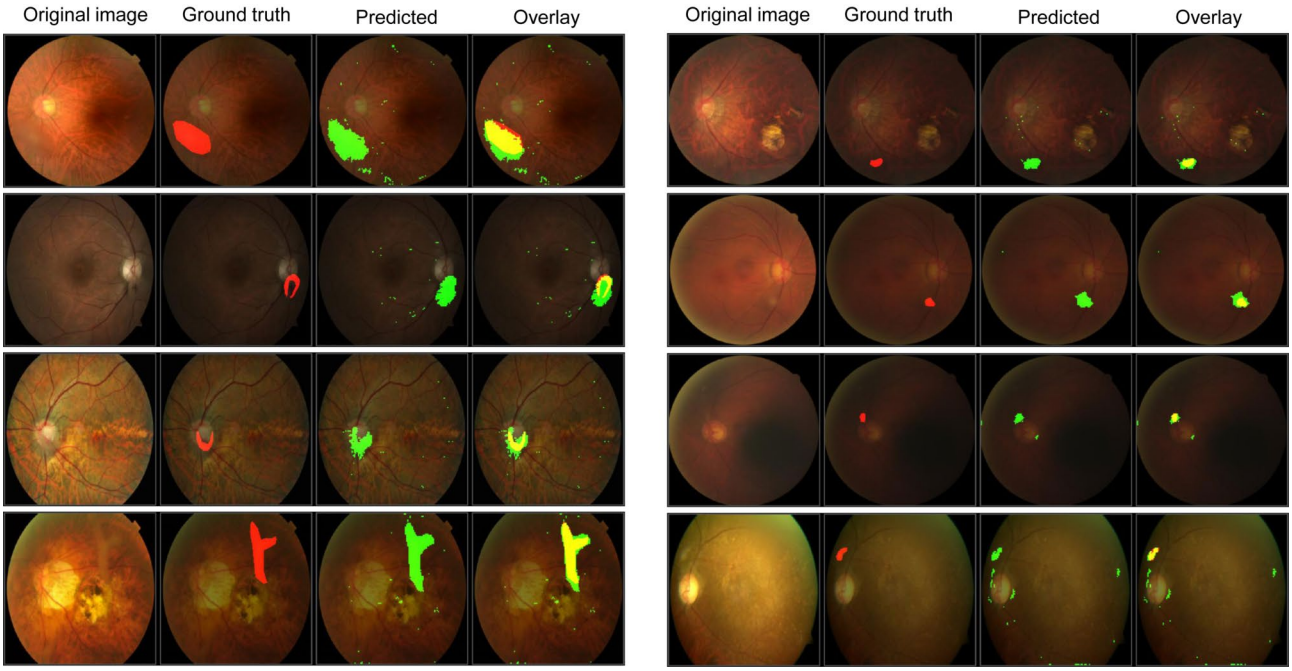

Fig. 4. Qualitative examples of Weiss ring segmentation. For each case, columns display (from left to right) the original color fundus photograph, the ground truth mask (red), the model-predicted mask (green), and an overlay on the original image highlighting agreement and errors (yellow indicates overlap between ground truth and prediction).

(95% CI 0.310–0.497), Recall of 0.894 (95% CI 0.837–0.950), and Precision of 0.422 (95% CI 0.324–0.520). SegNet produced similar results to DeepLabv3+, with Dice 0.555 and IoU 0.398, Recall 0.876 (95% CI 0.819–0.932), and Precision 0.420 (95% CI 0.319–0.520).

In five-fold cross-validation on the development set (Supplementary Table S3), the concatenation meta-classifier achieved the best overall discrimination for Weiss ring detection (AUC 0.830, 95% CI 0.795–0.864) with an accuracy of 0.770 (0.733–0.803), sensitivity of 0.753 (0.686–0.812), and specificity of 0.778 (0.734–0.819). Among single-backbone models, EfficientNet-B0, ResNet-50, and ConvNeXt-V2 showed similar performance (AUC 0.809–0.812; accuracy 0.747–0.751), whereas VGG-19 was slightly lower (AUC 0.797; accuracy 0.735).

Model	Dice (95% CI)	IoU (95% CI)	Precision (95% CI)	Recall (95% CI)
U-Net	0.5783 (0.4837–0.6728)	0.4212 (0.3244–0.5180)	0.4355 (0.3362–0.5348)	0.9172 (0.8590–0.9754)
DeepLabv3 +	0.5609 (0.4677–0.6541)	0.4035 (0.3099–0.4970)	0.4220 (0.3240–0.5200)	0.8936 (0.8368–0.9504)
SegNet	0.5545 (0.4590–0.6500)	0.3979 (0.3022–0.4935)	0.4195 (0.3186–0.5204)	0.8755 (0.8185–0.9324)

Table 2. Segmentation performance metrics of different CNN models. CI = confidence interval; IoU = Intersection over Union. This evaluation was performed on the locked test set (n = 117; 39 Weiss ring–positive and 78 Weiss ring–negative images). Reported values are point estimates with 95% confidence intervals (CIs) computed using nonparametric bootstrap resampling.

Model	AUC	SE	95% CI	P-value (DeLong)	P-value (Holm adjusted)
Concatenation (meta-classifier)	0.903	0.0296	0.845–0.961	Reference	Reference
EfficientNet-B0	0.847	0.0387	0.771–0.923	0.2048	0.473
ResNet-50	0.839	0.0379	0.765–0.914	0.1578	0.473
VGG-19	0.837	0.0379	0.762–0.911	0.1717	0.473
ConvNeXt-V2	0.829	0.0376	0.755–0.902	0.1019	0.407

Table 3. Comparison of areas under the ROC curve for different deep learning models. AUC = area under the receiver operating characteristic curve; CI = confidence interval; SE = standard error. This evaluation was performed on the locked test set (n = 117; 39 Weiss ring–positive and 78 Weiss ring–negative images). Reported values are point estimates with 95% confidence intervals (CIs) computed using nonparametric bootstrap resampling.

Fig. 5. Receiver operating characteristic (ROC) curves for Weiss ring detection. (**A**) Comparison of different convolutional neural network (CNN) architectures, including EfficientNet-B0, ResNet-50, VGG-19, and ConvNeXt-V2, against the concatenated ensemble model. (**B**) ROC curve of the concatenated ensemble model with 95% confidence interval (shaded area).

Across models, PPV was modest (0.578–0.629) while NPV remained high (0.859–0.864), indicating reliable exclusion of Weiss ring absence in this setting.

As shown in Table 3 and Fig. 5, the concatenation meta-classifier achieved the highest discrimination for Weiss ring detection (AUC 0.903, 95% CI 0.845–0.961). However, compared with the individual CNN backbones, the AUC differences were not statistically significant by DeLong testing after Holm adjustment (Holm-adjusted p = 0.407–0.473). Among the single models, EfficientNet-B0 showed the largest AUC (0.847, 95% CI 0.771–0.923), followed by ResNet-50 (0.839, 95% CI 0.765–0.914), VGG-19 (0.837, 95% CI 0.762–0.911), and ConvNeXt-V2 (0.829, 95% CI 0.755–0.902). The ROC curves illustrated the clear separation achieved by the concatenation meta-classifier, with the shaded 95% CI band in Fig. 5 confirming the robustness of the ensemble across resampling.

Table 4 summarized fixed-threshold (0.5) classification performance. The concatenation model provided the best overall accuracy (0.812, 95% CI 0.729–0.878), with sensitivity 0.692 (0.524–0.830), specificity 0.872 (0.777–0.937), PPV 0.730 (0.559–0.862), and NPV 0.850 (0.753–0.920). EfficientNet-B0 was the strongest single model (accuracy 0.778, 0.692–0.849; AUC 0.847), while ResNet-50 and VGG-19 showed comparable performance, and ConvNeXt-V2 exhibited the lowest overall accuracy (0.718, 0.627–0.797) and PPV (0.562, 0.412–0.705). On the test set, the confusion matrices (Fig. 6) showed that the concatenation meta-classifier produced the fewest overall errors (27 true positives, 68 true negatives) compared with the individual backbones at the fixed threshold of 0.5. On the test set, reliability diagrams (Supplementary Figure S2) suggested that the concatenation

Model	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Concatenation (meta-classifier)	0.812 (0.729–0.878)	0.692 (0.524–0.830)	0.872 (0.777–0.937)	0.730 (0.559–0.862)	0.850 (0.753–0.920)
EfficientNet-B0	0.778 (0.692–0.849)	0.641 (0.472–0.788)	0.846 (0.747–0.918)	0.676 (0.502–0.820)	0.825 (0.724–0.901)
ResNet-50	0.761 (0.673–0.835)	0.692 (0.524–0.830)	0.795 (0.688–0.878)	0.628 (0.467–0.770)	0.838 (0.734–0.913)
VGG-19	0.761 (0.673–0.835)	0.641 (0.472–0.788)	0.821 (0.717–0.898)	0.641 (0.472–0.788)	0.821 (0.717–0.898)
ConvNeXt-V2	0.718 (0.627–0.797)	0.692 (0.524–0.830)	0.731 (0.618–0.825)	0.562 (0.412–0.705)	0.826 (0.716–0.907)

Table 4. Classification performance metrics of different convolutional neural network models (based on the threshold of 0.5). CI = confidence interval; NPV = negative predictive value; PPV = positive predictive value. This evaluation was performed on the locked test set (n = 117; 39 Weiss ring–positive and 78 Weiss ring–negative images). Reported values are point estimates with 95% confidence intervals (CIs) computed using the Wilson score method for proportions (accuracy, sensitivity, specificity, PPV, NPV).

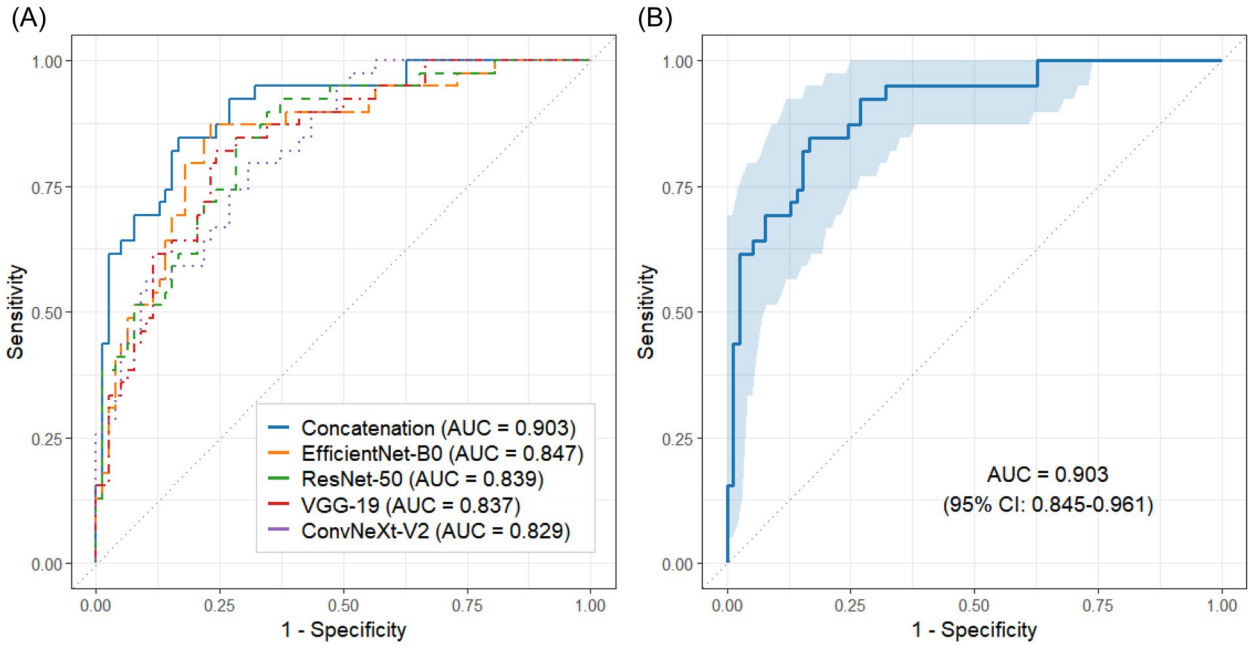

Fig. 6. Confusion matrices for five classification models on the locked test set. Confusion matrices are shown for the concatenation meta-classifier, EfficientNet-B0, ResNet-50, VGG-19, and ConvNeXt-V2. Rows indicate the actual labels (Weiss vs No-Weiss) and columns indicate the predicted labels. Cell values represent the number of test images in each category (true positives, false negatives, false positives, and true negatives). A fixed threshold of 0.5 was applied to binarize model outputs; darker shading indicates higher counts.

meta-classifier was the best calibrated overall, with the lowest Brier score (0.0956) and observed event rates most closely tracking the diagonal compared with the backbone models (Brier 0.1373–0.1661).

Figure 7 shows representative examples, including the original fundus image, a magnified view of the Weiss ring, the ophthalmologist-annotated ground truth, the model-predicted segmentation, and the Grad-CAM visualization. The predicted masks demonstrated strong agreement with expert annotations, while the heatmaps highlighted clinically relevant regions associated with the Weiss ring. In these cases, classification scores from Network 1 ranged from 89.05% to 99.40% and were further improved with concatenation (97.61%–99.99%). Integration of segmentation and classification enhanced prediction reliability, particularly in challenging cases involving faint or peripapillary rings.

Qualitative inspection of representative errors (Fig. 8) showed that U-Net failures were dominated by false-positive over-segmentation, in which broad regions of low-contrast background, peripheral vignetting, or illumination artifacts were labeled as Weiss ring. Additional errors occurred when peripapillary structures (e.g., optic disc margin or adjacent texture changes) were mistaken for the target, producing large confluent masks that only partially overlapped the annotated ring.

In held-out source evaluation (Supplementary Table S4), we assessed cross-dataset generalizability by training on all sources except one and testing on the entire held-out dataset (several sources were small). AUCs varied across datasets, with the highest discrimination on RFMiD 2.0 (AUC 0.930, 95% CI 0.815–1.000) and strong performance on RFMiD (AUC 0.860, 0.747–0.948) and 1000FP (AUC 0.859, 0.747–0.949). Lower

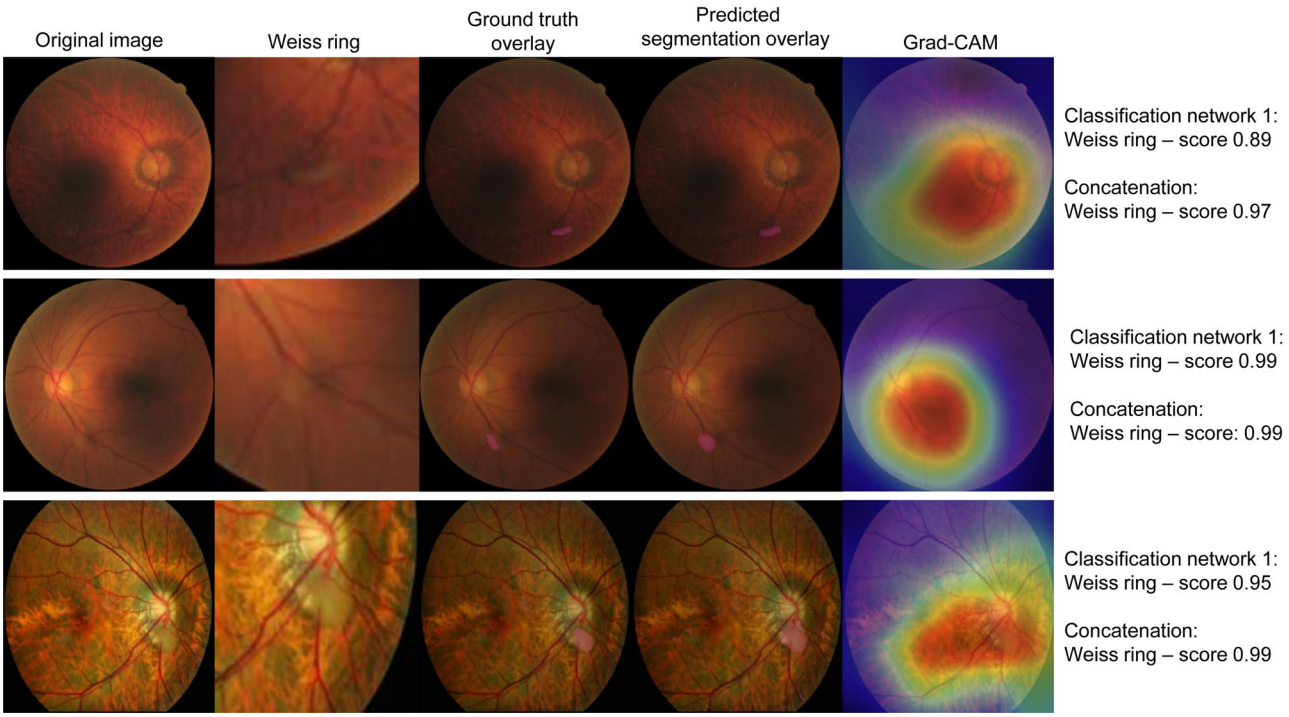

Fig. 7. Segmentation results, Grad-CAM visualization, and final prediction scores. Each row shows the original fundus image, a zoomed-in view of the Weiss ring, the ophthalmologist-annotated ground truth overlay, the model-predicted segmentation overlay, and the Grad-CAM heatmap highlighting regions contributing to classification. The final report displays classification scores from Classification Network 1 (EfficientNet-B0) and the concatenated output (logistic regression).

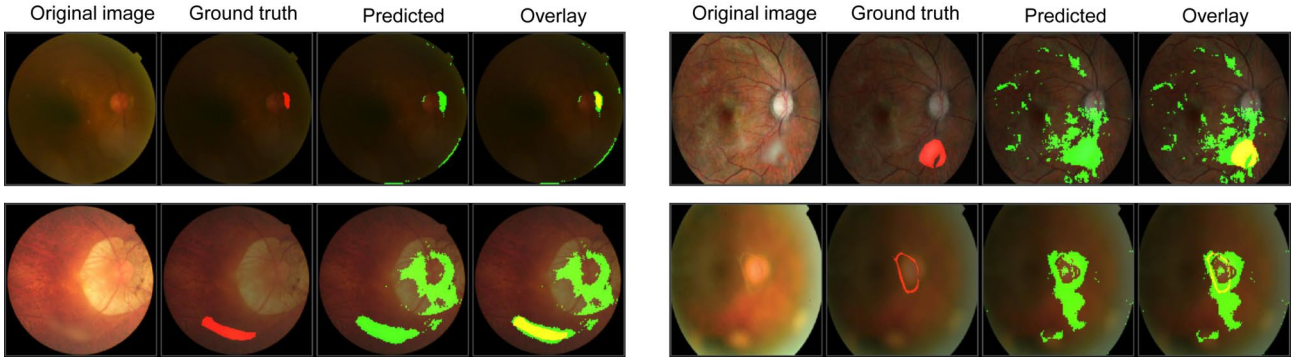

Fig. 8. Representative segmentation failure modes of the U-Net for Weiss ring localization. Each example is shown as a four-panel sequence: original color fundus image, ophthalmologist-annotated ground truth mask (red), model-predicted mask (green), and overlay on the original image (yellow indicates overlap between ground truth and prediction).

discrimination was observed on ODIR-5 K (AUC 0.723, 0.659–0.784) and PALM (AUC 0.693, 0.610–0.776), indicating a larger domain shift for these sources.

Discussion

In this study, we developed and validated a dual-model deep learning framework for the automated detection and segmentation of Weiss rings from FPs. Our results demonstrate that combining a U-Net segmentation backbone with image-level classification networks and logistic regression significantly improves performance compared to individual models. The ensemble achieved an AUC of 0.903, with high accuracy, sensitivity, and specificity, while also providing anatomical localization. Segmentation performance, although challenged by subtle or peripapillary rings, showed consistent superiority of U-Net over DeepLabv3 + and SegNet in Dice and IoU metrics. Clinically, this approach may support fundus photograph review by highlighting subtle or peripapillary Weiss rings that might otherwise be overlooked, although its effect on posterior vitreous detachment recognition

and downstream clinical decision-making remains unproven. For residents and general practitioners, a Weiss ring is a clinically relevant sign of complete posterior vitreous detachment in the appropriate context. Automated detection and visualization may therefore be useful for education and image review, but their impact on early recognition and clinical management remains uncertain. These findings should be interpreted cautiously because the study was retrospective, used heterogeneous public datasets, and lacked prospective external validation. Accordingly, this framework should be regarded as feasibility-level rather than clinically deployable.

We position this system as an adjunct for documentation and research, not as a replacement for clinical PVD assessment. In line with current standards of care, this tool is not intended for triage of symptomatic PVD nor to replace slit-lamp examination, scleral depression, or ultrawide-field imaging/OCT when indicated. Instead, we position it as an adjunct for archival/retrospective review and quality-assurance flagging of subtle peripapillary rings, education and standardized documentation, and research phenotyping where ultrawide-field imaging/OCT may be unavailable. This reframing acknowledges that Weiss rings are not universal in clinically confirmed PVD and may be absent or outside the field in standard 30–55° FPs.

To our knowledge, this is the first study to apply deep learning to Weiss ring detection in FPs. Prior investigations of PVD have primarily relied on OCT, which directly images the vitreoretinal interface[3], or ultrawide-field imaging for peripheral retinal changes[27]. These modalities have successfully enabled automated assessment of PVD or PVD-related peripheral lesions, but none have specifically addressed Weiss rings, which represent the hallmark of complete PVD. Because Weiss rings are often invisible in OCT scans and have been largely overlooked in FP analysis, their diagnostic potential remains underexplored. Previous AI applications in fundus imaging have predominantly focused on vascular and retinal diseases such as diabetic retinopathy, age-related macular degeneration, and glaucoma[10,28]. Our work extends AI-based ophthalmic image analysis to the vitreous domain, demonstrating that vitreoretinal interface findings can also be robustly identified from standard FP datasets.

Segmentation improves Weiss ring detection by constraining learning to clinically relevant regions. Several prior studies have shown that integrating segmentation with CNN-based classification improves performance compared to using classification networks alone[29,30]. Segmentation constrains the learning to clinically relevant regions, reduces background noise, and enhances feature extraction for subtle lesions. For example, in retinal vein occlusion detection, segmentation-guided CNNs have consistently outperformed pure image-level classifiers by focusing on vascular or lesion-related regions rather than entire images[28]. Our findings align with these reports, as combining U-Net segmentation with CNN classifiers yielded higher accuracy and AUC than standalone CNN models (ResNet-50, VGG-19, ConvNeXt-V2). This synergy suggests that segmentation not only improves interpretability but also acts as a form of spatial attention, guiding networks toward pathologically meaningful areas and thereby strengthening robustness in challenging cases such as faint or peripapillary Weiss rings.

Further data collection is needed to improve accuracy, particularly in cases that are difficult to interpret. Figure 8 shows that most errors occur in low-signal conditions. False negatives arise when the Weiss ring is faint, incomplete, or peripapillary, especially when blur or media opacity reduces boundary contrast. False positives are mainly driven by artifacts (peripheral illumination fall-off, reflections, and textured peripheral retina) that are misclassified as ring-like regions. Overall, these cases suggest that model robustness is primarily limited by image quality and peripapillary confounding. Because inter-observer agreement among expert graders was high (pairwise Dice ≈ 0.84–0.90), the substantially lower model Dice (~0.58) relative to this human baseline indicates that the current segmentation output should be interpreted as supportive anatomical localization rather than human-equivalent delineation.

The presence of a Weiss ring carries clinical implications beyond its role as a marker of complete PVD. One important consideration is its relationship with the occurrence of new floaters after cataract surgery[31]. Cataract extraction can accelerate vitreous changes due to the removal of the natural lens barrier, increased vitreous mobility, and shifts in intraocular dynamics. Patients with a pre-existing Weiss ring are more likely to experience symptomatic floaters postoperatively, as the detached vitreous opacity becomes more noticeable when vision clears following lens replacement. Automated detection of Weiss rings from fundus images may therefore provide prognostic information to clinicians counseling patients on the risk of postoperative floaters after cataract surgery.

In addition, the identification of a Weiss ring has implications for vitreoretinal surgical planning. During pars plana vitrectomy, the decision of whether to induce a PVD intraoperatively depends on the pre-existing vitreous status[32]. In eyes where a Weiss ring is already present, complete PVD has typically occurred, and induction may not be necessary. Conversely, in eyes without a Weiss ring, especially in younger patients or those with vitreoretinal interface disease, surgeons may need to actively induce PVD, a step that carries potential risks such as iatrogenic retinal breaks. Automated Weiss ring detection could thus serve as a preoperative marker to stratify surgical difficulty and anticipate intraoperative complications.

Although this system is not intended to diagnose PVD or determine acute risk, a "Weiss ring–positive" flag provides an actionable indexing signal in large image archives. First, the flag enables rapid phenotyping of thousands of fundus photographs for retrospective PVD epidemiology and risk-factor analyses, supporting scalable correlations with age, refractive error, systemic factors, or coexisting retinal pathology. Second, it facilitates cohort discovery and enrichment by prioritizing Weiss ring–positive images for targeted chart review and outcomes linkage (e.g., documented retinal tear/detachment) without manual screening of entire databases. Third, it can function as a QA tag to identify images with prominent vitreous opacities that may degrade or confound other retinal AI models, enabling exclusion or stratified analyses in research pipelines. In a diabetic screening archive, a Weiss ring flag can triage records for subsequent clinician chart review for retinal tear/detachment history.

In this study, relatively simple architectures (U-Net and EfficientNet-B0) demonstrated superior performance compared to more complex models. One plausible explanation is the limited dataset size, which constrains

the generalizability and stability of deeper or more parameter-heavy networks. Complex architectures such as ConvNeXt-V2 typically require large-scale training data to fully exploit their representational capacity, whereas lighter models and hybrid approaches are less prone to overfitting in small-data regimes[33]. This observation is consistent with prior reports in medical imaging, where simpler models or ensemble strategies often outperform advanced networks when the available dataset is modest[26,34].

Although we evaluated the model on a locked independent test set, the number of test cases, particularly the limited number of Weiss ring–positive images, remains modest. Small validation samples yield imprecise performance estimates (wider confidence intervals for sensitivity/specificity and AUC) and can make point estimates appear overly strong due to sampling variability, especially for rare findings[35]. Accordingly, our high point estimates should be interpreted cautiously, and larger, multi-center external validation with more positive cases will be necessary to confirm generalizability and obtain tighter uncertainty bounds.

This study has several important limitations related to its retrospective design, possible dataset bias, and generalizability. First, the dataset was aggregated from multiple public sources, introducing heterogeneity in camera type, field of view, resolution, compression, illumination, and patient mix. Because subject identifiers were not consistently available, data were split at the image level rather than the patient level, and possible fellow-eye leakage cannot be fully excluded. While this diversity may improve robustness, it also induces domain shift, with performance varying across held-out sources (Supplementary Table S4), particularly for low-contrast or peripapillary rings where segmentation is sensitive to image quality. Accordingly, source-specific quality control/harmonization and prospective external validation are needed to confirm generalizability across acquisition environments. Second, the relatively small number of Weiss ring–positive images (n = 194) limits precision and subgroup inference for rare or challenging phenotypes, such as faint, incomplete, or peripapillary rings, and for cases with media opacity or artifacts. Although transfer learning and augmentation mitigate overfitting, a larger curated cohort with more positive cases and explicit enrichment of difficult presentations will be required to obtain tighter confidence intervals, support reliable stratified analyses, and improve generalization in real-world screening images. Clinical readiness is limited by the absence of validation on a fully external, prospectively collected cohort acquired under real-world conditions. Third, segmentation ground truth relied on manual ophthalmologist annotations, which, despite consensus voting, remain subject to inter-observer variability. Fourth, the test set size (39 Weiss ring and 78 no-Weiss ring images) restricts definitive conclusions regarding model generalization to real-world clinical populations. Fifth, because several source datasets were challenge-style collections enriched for retinal pathology, the overall image distribution is not representative of population-based screening cohorts. Finally, while our framework improves explainability using segmentation overlays and Grad-CAM, further validation in prospective clinical datasets and workflow integration studies is warranted before clinical translation. Since FPs provide a limited field of view, and ring visibility is variable, prospective validation against ultrawide-field imaging or OCT reference standards is needed in the future.

Conclusion

We present a hybrid FP-based system that couples U-Net segmentation with CNN classification to detect and localize Weiss rings, achieving high discrimination while providing spatial interpretability. Performance gains over standalone classifiers suggest that explicit localization functions as effective spatial attention. Importantly, this tool is not intended to triage symptomatic PVD or replace slit-lamp examination, scleral depression, or ultrawide-field imaging/OCT. Instead, its most appropriate roles are archival/retrospective review, education and quality-assurance flagging, and research phenotyping, particularly where wider-field imaging is unavailable. Segmentation results should be interpreted against human inter-observer variability and the limited field-of-view/visibility of Weiss rings on FPs. Further work is required before clinical deployment. This includes prospective multi-site validation using explicit reference standards (e.g., OCT and/or ultrawide-field imaging when available) and evaluation on larger datasets enriched for difficult presentations with image-quality stratification. Workflow/usability studies are also needed to determine whether automated Weiss ring flags improve archival phenotyping and reviewer efficiency, without substituting for clinical examination or dictating intervention.

Data availability

The datasets utilized in this study are publicly accessible. All methodologies have been detailed within this manuscript. The full MATLAB training and inference scripts are provided in Supplementary Materials.

Received: 21 October 2025; Accepted: 12 March 2026
Published online: 17 March 2026

References

1. Johnson, M. W. Posterior vitreous detachment: Evolution and complications of its early stages. *Am. J. Ophthalmol.* **149**, 371-382.e1 (2010).
2. Hollands, H. et al. Acute-onset floaters and flashes: Is this patient at risk for retinal detachment?. *JAMA* **302**, 2243–2249 (2009).
3. Abraham, J. R. & Ehlers, J. P. Posterior vitreous detachment: Methods for detection. *Ophthalmol. Retina* **4**, 119–121 (2020).
4. Nixon, T. R. W., Davie, R. L. & Snead, M. P. Posterior vitreous detachment and retinal tear – A prospective study of community referrals. *Eye* **38**, 786–791 (2024).
5. Mamou, J. et al. Ultrasound-based quantification of vitreous floaters correlates with contrast sensitivity and quality of life. *Invest. Ophthalmol. Vis. Sci.* **56**, 1611–1617 (2015).
6. Xiong, L., Li, H. & Xu, L. An approach to evaluate blurriness in retinal images with vitreous opacity for cataract diagnosis. *J. Healthc. Eng.* **2017**, 5645498 (2017).
7. Li, T. et al. Applications of deep learning in fundus images: A review. *Med. Image Anal.* **69**, 101971 (2021).
8. Akiba, J., Ishiko, S. & Yoshida, A. Variations of Weiss's ring. *Retina* **21**, 243 (2001).

9. Li, A. L. et al. Automated detection of posterior vitreous detachment on OCT using computer vision and deep learning algorithms. *Ophthalmol. Sci.* **3**, 100254 (2023).
10. Cen, L.-P. et al. Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks. *Nat. Commun.* **12**, 4828 (2021).
11. Du, R. et al. Deep learning approach for automated detection of myopic maculopathy and pathologic myopia in fundus images. *Ophthalmol. Retina* **5**, 1235–1244 (2021).
12. Bhati, A., Gour, N., Khanna, P. & Ojha, A. Discriminative kernel convolution network for multi-label ophthalmic disease detection on imbalanced fundus image dataset. *Comput. Biol. Med.* **153**, 106519 (2023).
13. Pachade, S. et al. Retinal fundus multi-disease image dataset (RFMiD): A dataset for multi-disease detection research. *Data* **6**, 14 (2021).
14 Panchal, S. et al. Retinal fundus multi-disease image dataset (RFMiD) 2.0: A dataset of frequently and rarely identified diseases. *Data* **8**, 29 (2023).
15. Fang, H. et al. Open fundus photograph dataset with pathologic myopia recognition and anatomical structure annotation. *Sci. Data* **11**, 99 (2024).
16. Orlando, J. I. et al. REFUGE Challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Med. Image. Anal.* **59**, 101570 (2020).
17. Aljabri, M., AlAmir, M., AlGhamdi, M., Abdel-Mottaleb, M. & Collado-Mesa, F. Towards a better understanding of annotation tools for medical imaging: A survey. *Multimed. Tools. Appl.* **81**, 25877–25911 (2022).
18. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (eds Navab, N. et al.) 234–241 (Springer International Publishing, 2015). https://doi.org/10.1007/978-3-319-24574-4_28.
19. Shim, V. J., Shim, H. & Roh, S. EfficientNet-B0 outperforms other CNNs in image-based five-class embryo grading: A comparative analysis. *J. Anim. Reprod. Biotechnol.* **39**, 267–277 (2024).
20. Kumar, M., Bajaj, K., Sharma, B. & Narang, S. A comparative performance assessment of optimized multilevel ensemble learning model with existing classifier models. *Big Data* **10**, 371–387 (2022).
21. Yoo, T. K. et al. Simple code implementation for deep learning–based segmentation to evaluate central serous chorioretinopathy in fundus photography. *Transl. Vis. Sci. Technol.* **11**, 22 (2022).
22. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* **44**, 837–845 (1988).
23. Saranya, S. M. et al. Enhancing segmentation accuracy of diabetic retinal images using transform techniques and deepLabV3+ with ensemble classification. In *Emerging Trends and Technologies on Intelligent Systems* (eds Noor, A. et al.) 179–190 (Springer Nature, 2025). https://doi.org/10.1007/978-981-97-5703-9_14.
24. Ryu, J., Rehman, M. U., Nizami, I. F. & Chong, K. T. SegR-Net: A deep learning framework with multi-scale feature fusion for robust retinal vessel segmentation. *Comput. Biol. Med.* **163**, 107132 (2023).
25. Mmileng, O. P., Whata, A., Olusanya, M. & Mhlongo, S. Application of ConvNeXt with transfer learning and data augmentation for malaria parasite detection in resource-limited settings using microscopic images. *PLoS ONE* **20**, e0313734 (2025).
26. Ryu, S. Y., Choi, J. Y. & Yoo, T. K. Automated detection of retinal artery occlusion in fundus photography via self-supervised deep learning and multimodal interpretability using a multimodal AI chatbot. *Med. Biol. Eng. Comput.* https://doi.org/10.1007/s11517-025-03353-7 (2025).
27. Tang, Y.-W. et al. Automatic detection of peripheral retinal lesions from ultrawide-field fundus images using deep learning. *Asia-Pacific J. Ophthalmol.* **12**, 284–292 (2023).
28. Choi, E. Y. et al. Predicting branch retinal vein occlusion development using multimodal deep learning and pre-onset fundus hemisection images. *Sci. Rep.* **15**, 2729 (2025).
29. Shyamalee, T. & Meedeniya, D. Glaucoma detection with retinal fundus images using segmentation and classification. *Mach. Intell. Res.* **19**, 563–580 (2022).
30. De Fauw, J. et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat. Med.* **24**, 1342–1350 (2018).
31. Morano, M. J. et al. Incidence and risk factors for retinal detachment and retinal tear after cataract surgery: IRIS® Registry (Intelligent Research in Sight) analysis. *Ophthalmol. Sci.* **3**, 100314 (2023).
32. Rahman, R., Chaudhary, R. & Anand, N. Verification of posterior hyaloid status during Pars Plana Vitrectomy, after preoperative evaluation on Optical Coherence Tomography. *Retina* **32**, 706 (2012).
33. Chowa, S. S. et al. A low complexity efficient deep learning model for automated retinal disease diagnosis. *J. Healthc. Inform. Res.* **9**, 1–40 (2025).
34. Tian, G. et al. A deep ensemble learning-based automated detection of COVID-19 using lung CT images and Vision Transformer and ConvNeXt. *Front. Microbiol.* https://doi.org/10.3389/fmicb.2022.1024104 (2022).
35. Hajian-Tilaki, K. Sample size estimation in diagnostic test studies of biomedical informatics. *J. Biomed. Inform.* **48**, 193–204 (2014).

Author contributions

Conceptualization & Methodology: H Kim, and T K Yoo Data collection: H Kim, S Y Ryu, and T K Yoo Data processing: H Kim, S Y Ryu, T K Yoo, and D D Hwang Investigation & Visualization: H Kim, S Y Ryu, and T K Yoo Supervision: T K Yoo and D D Hwang Writing—original draft: H Kim, S Y Ryu, and T K Yoo Writing—review & editing: T K Yoo and D D Hwang.

Declarations

Competing interests

T K Yoo is an advisory board member of MediWhale and has received consultant fees as part of the standard compensation package. The remaining authors declare no conflicts of interest.

Ethics approval and consent to participate

As the study exclusively utilized publicly available data, IRB approval was not required. The study complied with the principles outlined in the Declaration of Helsinki, ensuring ethical standards were maintained throughout.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-44593-3.

Correspondence and requests for materials should be addressed to T.K.Y.