Scientific Reports

https://doi.org/10.1038/s41598-026-38272-6

Article in Press

Deep learning-based detection of retinal detachment with vitreous hemorrhage in ocular ultrasound images

Received: 23 August 2025

Accepted: 29 January 2026

Published online: 11 February 2026

Cite this article as: Toyama N., Hidaka T., Tamura H. *et al.* Deep learning-based detection of retinal detachment with vitreous hemorrhage in ocular ultrasound images. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-38272-6

Naoki Toyama, Takako Hidaka, Hiroki Tamura & Yasuhiro Ikeda

【Title】

Deep learning–based detection of retinal detachment with vitreous hemorrhage in ocular ultrasound images

【Authors】

Naoki Toyama[1], Takako Hidaka[1], Hiroki Tamura[2], Yasuhiro Ikeda[1]

1. Department of Ophthalmology, Faculty of Medicine, University of Miyazaki, Japan
2. Electrical and Electronic Engineering Program, Faculty of Engineering, University of Miyazaki, Japan

Corresponding Author: Yasuhiro Ikeda, MD, PhD.

Department of Ophthalmology, Faculty of Medicine, University of Miyazaki, 5200 Kihara, Kiyotake, Miyazaki, 889-1692, Japan.

Tel: 81-985-85-2806

Fax: 81-985-84-2065

e-mail address: ymocl@med.miyazaki-u.ac.jp

and

Hiroki Tamura, PhD.

Electrical and Electronic Engineering Program, Faculty of Engineering, University of Miyazaki, 1-1Gakuen Kibanadai-nishi, Miyazaki, 889-2192, Japan.

Tel: +81-985-58-7409, Fax: +81-985-58-7409

e-mail address: htamura@miyazaki-u.ac.jp

[Key Words]

YOLOv5, Deep Learning, retinal detachment, vitreous hemorrhage, ocular ultrasonography

Abstract

Retinal detachment (RD) is a serious ocular disease that can lead to permanent vision loss. In cases with fundus-obscuring vitreous hemorrhage (VH), it is difficult to detect RD even using ocular ultrasonography. We developed a convolutional neural network (CNN) based on the You Only Look Once version 5 (YOLOv5) architecture to detect RD and VH on B-scan ultrasound images. The model was trained using 2,188 images and validated using 1,042 images. We applied image enhancement techniques, including unsharp masking (UM), to improve the detection accuracy. The final model (Incorporating 5-fold cross-validation along with previous techniques) achieved overall accuracies of 96.6%, 99.2%, and 98.0% for RD, VH, and RD with VH, respectively. Our deep-learning algorithm showed high accuracy in detecting RD and VH on ocular ultrasound images. In cases with fundus-obscuring VH, our deep-learning algorithm might be useful for detecting RD as a supportive tool on ocular ultrasound images.

Introduction

Retinal detachment (RD) and vitreous hemorrhage (VH) are serious ocular diseases that can lead to permanent vision loss if not promptly diagnosed and treated[1,2]. RD can usually be identified on direct fundus examination. However, dense VH that obscures the fundus may markedly limit visualization of the posterior pole. In such cases, ocular ultrasonography is a valuable diagnostic tool, particularly when direct fundus examination is not possible owing to media opacity[3,4]. However, the presence of VH can complicate the interpretation of ultrasound findings. Hemorrhagic vitreous opacities frequently generate linear or membranous echo patterns that mimic or obscure RD, leading to diagnostic uncertainty, particularly for less experienced clinicians[5].

Since the third artificial intelligence (AI) revolution around 2010, AI has grown rapidly and is supporting humans in a wide range of industries. This technological progress is driven by a machine learning technique called deep learning, which learns by itself based on huge amounts of data and grows by absorbing human actions and behavior. Recent advances in AI and deep learning have yielded promising results in various areas of ophthalmology[6,7]. Deep-learning algorithms have shown high accuracy in detecting diabetic retinopathy from fundus images and diagnosing multiple retinal diseases[8,9]. However, the application of these techniques to ocular ultrasonography for the detection of RD and VH has been limited[10,11].

In parallel, machine-learning approaches have been widely applied to ocular disease analysis using retinal fundus images, particularly for

glaucoma detection, including analytical comparisons of models, feature-based classification, and multi-stage deep-learning frameworks[6,12,13]. Broader AI reviews further highlight trends toward task-specific validation and robustness[14]. However, these fundus-based approaches rely on clear optical media and are not directly transferable to ocular ultrasonography, where image interpretation is dominated by acoustic echoes, speckle noise, and overlapping membranous structures, especially in the presence of VH.

YOLO is a computer vision algorithm for real-time object detection. Conventional object detection systems use techniques such as sliding windows and region proposals to identify the location of potential objects in an image. However, YOLO simultaneously classifies and localizes objects by looking at an image only once, enabling fast and highly accurate object detection.

Unlike prior studies that primarily focus on detecting a single retinal pathology or relying on fundus imaging, this study addresses the clinically challenging task of differentiating RD, VH, and their coexistence using ocular ultrasound images. Furthermore, rather than proposing a fixed detection pipeline, we systematically evaluate the impact of image enhancement, threshold-based processing, and detection-based modeling within a unified framework. This approach enables task-specific validation of model design choices for diagnostically ambiguous ultrasound cases.

The present study developed and validated a deep-learning algorithm that can detect RD with VH from ocular ultrasound images. We hypothesized that such an algorithm could assist clinicians in the rapid and accurate diagnosis of RD, especially in cases with media opacities such as

severe cataract and fundus-obscuring VH and thereby improve patient outcomes in the emergency setting. Accordingly, this study focuses on supporting differential interpretation of RD, VH, and their coexistence in diagnostically challenging ocular ultrasound images obtained from patients in whom pathology is already suspected, rather than on screening normal cases.

Methods

Ethical approval and study design

This study was approved by the Miyazaki University Ethics Review Board (approval number: O-1065) and adhered to the tenets of the Declaration of Helsinki. The need to obtain written informed consent was waived by the Miyazaki University Ethics Review Board owing to the retrospective nature of the study and the use of anonymized data. All patients were provided with the opportunity to refuse participation in the data analysis through an opt-out method.

Data collection and preprocessing

We retrospectively reviewed the clinical records of patients at the University of Miyazaki Hospital between 2017 and 2023. We collected 3,773 B-scan ocular ultrasound images from patients suspected of having VH or RD. Because the purpose of this study was to develop a diagnostic support tool for cases in which RD or VH was clinically suspected, the dataset consisted exclusively of images obtained from such patients rather than from a general screening population. All images were anonymized prior to analysis and labeled as RD (1,571), VH (1,832), or RD with VH (370). The diagnosis of RD and/or VH was based on comprehensive clinical evaluation, including ophthalmic examinations and intraoperative findings, as determined by a team of board-certified ophthalmologists at our institution. Image-level annotations were then performed by a single experienced ophthalmologist based on the corresponding clinical records and surgical findings. Multiple ultrasound images obtained from the same

patient were included in the dataset. The dataset was divided into a training set (2,188 images, 58.0%), validation set (1,042 images, 27.6%), and test set (543 images, 14.4%). Given the limited availability of annotated ocular ultrasound images, this split was selected to preserve sufficient data for model training while maintaining an independent test set for unbiased performance evaluation. Corresponding bounding box annotations were determined using the same expert-based ground truth labeling process. Representative examples of ocular ultrasound images for each class (RD, VH, and RD with VH) are provided in Supplementary Figure S1.

Image enhancement

Unsharp masking was applied as an external image-enhancement preprocessing step, independent from the model's internal processing, to improve the quality of ultrasound images. UM is a technique widely used in medical imaging and photo editing to improve image clarity and contrast[15-17]. The process involves creating a blurred version of the original image, subtracting it from the original image to enhance the edges, and then restoring the edge-enhanced image to the original image. The technique is controlled by three parameters: radius (which determines the extent of blurring), amount (which adjusts the strength of the edge enhancement), and threshold (which sets the minimum contrast difference to be considered an edge). A single parameter configuration was selected and applied consistently across all images, based on preliminary visual assessment to enhance structural edges without introducing noticeable

artifacts.

We compared UM with other enhancement methods, including Contrast Limited Adaptive Histogram Equalization (CLAHE) and Histogram Equalization (HE). The comparison used metrics such as Contrast Improvement Index (CII), Edge Intensity (EI), and Image Sharpness (IS). The same fixed parameter set for UM was applied uniformly to the training, validation, and test datasets using the same fixed parameter set, and these image-quality metrics were used to characterize enhancement behavior rather than to directly optimize model performance.

Deep learning model

We developed a CNN based on the YOLOv5 architecture for detecting RD and VH. YOLO is a real-time object detection algorithm that processes images in a single pass and simultaneously predicts the bounding boxes and class probabilities. Although YOLO is primarily designed for object detection, it was adopted in this study because RD and VH appear as localized pathological regions in ocular ultrasound images. Accordingly, the images were annotated with region-level labels, and classification was derived from the class confidence scores of the detected regions. Bounding boxes were manually annotated to enclose visible pathological regions corresponding to RD and/or VH on each ocular ultrasound image. This detection-based approach enables localization-aware classification, which is clinically relevant for interpreting RD and VH in ultrasound images. This approach makes it particularly suitable for applications requiring real-time processing, such as in our study aimed at practical clinical implementation.

The YOLO algorithm divides the image into a grid and predicts multiple bounding boxes and class probabilities for each grid cell. Each bounding box is associated with an "objectness score" indicating the probability of containing an object, and "class probabilities" indicating the likelihood of the contained object belonging to specific classes. The final detection results are determined by combining these scores and applying a confidence threshold. Here, the model training followed the standard YOLOv5 configuration, using the default composite loss function and built-in data augmentation, without task-specific modification of the network architecture or loss design.

We also conducted a comparative study of each type of YOLO. In this study, we used YOLO, but we had to choose the best one for this study from the three types: YOLOv2-tiny, YOLOv5, and YOLOv7. Developments are observed in the following relationship: YOLOv2-tiny < YOLOv5 < YOLOv7. YOLOv2-tiny enables high-speed object detection and real-time processing, but its accuracy is limited to basic applications. This model features a network with reduced depth and width, making it suitable for use in environments with limited computing resources. Conversely, YOLOv5 significantly improves the balance between accuracy and speed. Equipped with advanced features, such as the introduction of a new architecture, optimization of the training process, and automatic learning rate adjustment, it achieves higher accuracy than YOLOv2-tiny while maintaining high-speed processing. YOLOv5 also excels in the flexibility and scalability of the model, making it adaptable to a variety of applications. Furthermore, YOLOv7 has demonstrated strong performance

improvements over previous versions on general object detection benchmarks. It incorporates the latest research results, such as refining the architecture, introducing a new loss function, and evolving data augmentation techniques. In particular, there have been significant improvements in the accuracy of small object detection, shortening training time, model versatility, and hardware optimization. With these developments, YOLOv7 shows very high performance in the latest object detection tasks. The development from YOLOv2-tiny to YOLOv7 clearly shows the trade-off between computational power and accuracy along with the progress of object detection technology. The selection of each version should be made carefully depending on the performance of the target application.

In this study, the YOLO model was decided based on the results below. In this section, data comprising a total of 1527 images were used. The distribution and hit rate of each YOLO model are shown in Table 1. From the above results, the accuracy rates for VH were 97%, 100%, and 100% for YOLOv2-tiny, YOLOv5, and YOLOv7, respectively. In other words, the verification revealed that regardless of which model is used, the accuracy rate for VH will be approximately 100%. Therefore, the YOLO model to be used was decided based on the results of RD. From the results, YOLOv2-tiny had an accuracy of 61%, YOLOv5 had an accuracy of 77.1%, and YOLOv7 had an accuracy of 50.8%, and the experiment showed that the optimal model was YOLOv5. Therefore, in this study, YOLOv5 will be used for the rest of the study.

Although YOLOv7 has demonstrated strong performance on general

object detection benchmarks, its RD classification performance was inferior to YOLOv5 in this study. This may be related to the higher architectural complexity of YOLOv7, which can require larger and more diverse datasets for stable generalization, whereas YOLOv5 showed more robust performance under the available data conditions. These results indicate that model selection for medical ultrasound analysis should be guided by task-specific validation rather than benchmark-level superiority. All model training and evaluation were conducted on a workstation equipped with an NVIDIA GeForce RTX 4090 GPU.

Four methods for validation experiments

We trained the YOLOv5 model on the enhanced images using data augmentation techniques such as brightness adjustments. The model was improved through four iterations.

1. Initial Approach: Basic YOLOv5 implementation

   We tried to distinguish between VH, RD, and RD with VH using only YOLOv5. First, we prepared three types of images taken with an ophthalmic ultrasound device, and divided the images into train, valid, and test at a ratio of 6:3:1 for learning. We then used the weights obtained by training for the images left for testing to calculate the accuracy rate. The formula used in the previous chapter was used for the accuracy rate. Images of the two symptoms are shown in Figure 1-1.

2. Image Enhancement: Applying UM to emphasize RD and VH features

   The second point is that if you simply input images directly into YOLOv5,

the accuracy rate will only increase to a certain extent, even if there is a lot of training data. Therefore, before inputting the images into YOLOv5, we added an image processing approach called UM to clarify the symptoms of RD. Figure 1-2 shows one image after UM has been applied to the left photo and the other shows the image after UM has been applied to the right photo. Although both RD and VH are slight, the symptoms are clearly visible.

3. Combined Approach: Integrating iterative learning and image enhancement techniques

   In proposed method 2, image processing was added to make the two symptoms clearer, but in order to accurately diagnose the concurrent symptoms, which is the purpose of this study, it is necessary to make it easier to find RD hidden by the symptoms of VH. Therefore, the third method focuses on the fact that RD is connected in a line, while VH is a spread of white particles, like dots. After image processing with UM, threshold processing is performed, and then binarization processing is applied to a specific pixel grid, and if the ratio of white pixels in each grid is less than a specific ratio, a process is added to fill the grid with black, and then machine learning is performed. Figure 1-3 shows the binarized image and the image after threshold processing, and Figure 1-4 shows the image after the above processing is applied to the two symptoms.

   The graphical user interface (GUI) elements provided by the ultrasound device were retained and underwent the same binarization process as the ultrasound image region. This was done to maintain consistency

with images obtained in actual clinical operation, and the presence of the GUI did not adversely affect model training or detection performance.

The threshold was set to suppress VH-related patterns while preserving RD structures, as shown in Figure 1-3. The threshold value (77) was determined empirically during model development by evaluating multiple candidate values on the training and validation data, and selecting the value that provided the best balance between suppression of VH-related noise and preservation of RD structures. This threshold was then fixed and applied consistently across all images.

4. Final Model: Incorporating five-fold cross-validation along with previous techniques.

   Images were unsharp masked (proposed method 2), after which five-fold cross-validation was applied._In this study, cross-validation was performed using the combined training and validation data, while the test dataset was excluded from the cross-validation process. In each fold, a subset of the combined data was used for validation, and the remaining data were used for training, and this procedure was repeated five times. The data used for cross-validation were the same as those in proposed method 3. The fixed test dataset, consisting of 247 RD images, 246 VH images, and 50 comorbidity images, was used only for final performance evaluation. This approach was adopted to assess model performance on unseen data under limited data conditions. Accuracy was defined as class-wise, image-level classification accuracy calculated on this fixed test dataset. Although the YOLO framework

predicts both bounding boxes and class confidence scores, performance evaluation was conducted at the image level using the predicted class outcomes. Accordingly, the proposed framework performs localization-aware classification, while the evaluation focuses on diagnostic classification performance rather than detailed localization metrics.

For all experiments, the validation accuracy stabilized within fewer than 500 training epochs, indicating good convergence and temporal stability of the learning process. Model training was performed for up to 1000 epochs with a batch size of 16, and convergence was observed earlier in all experiments.

In addition to the staged validation experiments described above, an ablation study was conducted to analyze the individual contribution of each preprocessing component. This analysis was performed under a simplified and controlled setting that differs from the validation experiments. Four configurations were evaluated using the same fixed test dataset and evaluation criteria: (i) YOLOv5 without preprocessing, (ii) YOLOv5 with UM, (iii) YOLOv5 with thresholding only, and (iv) YOLOv5 with combined UM and thresholding. Performance was evaluated using class-wise, image-level accuracy.

Model evaluations

The performance of the model was evaluated using a test set of 543 images (247 RD, 246 VH, and 50 RD with VH). We calculated the accuracy, sensitivity, and specificity for each condition.

To assess the model's accuracy, we used the following metrics:

1. True Positive (TP): Number of cases correctly predicted as positive by the model.
2. True Negative (TN): Number of cases correctly predicted as negative by the model.
3. False Positive (FP): Number of cases incorrectly predicted as positive by the model.
4. False Negative (FN): Number of cases incorrectly predicted as Negative by the model.

The accuracy was then calculated using the following formula:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Results

Model performance and development process

We developed our deep learning model through several iterations, each of which improved after the last iteration.

- Initial approach (Iteration 1)

At the initial stage of the research, a total of 3388 images were available for use. At this point, the breakdown of the images was 153 images of RD and VH and 79 images of RD with VH. These data were used until the redistribution of proposed method 2. In this study, the model was trained using a total of 3388 images, divided in a 6:4 ratio into train and valid data sets. The training was performed 1000 times. The details of the test were 64 images of RD with VH, 121 images of RD, and 185 images of VH.

The first iteration achieved accuracies of 89.2%, 90.3%, and 82.8% for RD, VH, and RD with VH, respectively. This initial performance was

promising but indicated the potential for further improvement (Table 2).

- Image enhancement with UM (Iteration 2)

In this experiment, we applied the UM image processing method to determine whether or not it could improve accuracy. The number of training runs was 1000. This test included 247 RD, 246 images VH images, and 50 complication images. The training data are listed in Table 3.

Applying UM as an image-enhancement technique improved our results, and this iteration achieved accuracies of 96.3% for RD, 97.6% for VH, and 92.0% for RD with VH (Table 2).

- Combined approach (Iteration 3)

We combined iterative learning with UM enhancement and implemented additional image processing techniques.

In this experiment, before applying the image processing shown in proposed iteration 2, binarization processing was performed after threshold processing to suppress the symptoms of VH, which is a hindrance when diagnosing RD. Looking at the image, the white shadows of RD cover a certain area, while VH has small, sparse areas. Therefore, binarization processing was performed to remove the small white areas, and threshold processing was attempted based on a certain degree of brightness. Here, a python program was used to threshold the image after UM so that pixels with a brightness of ≤77 were converted to black, while pixels with a brightness >77 were converted to white. Binarization processing was then applied to the processed image in a 10×10-pixel grid, and if the ratio of white pixels in each grid was less than 30%, the grid was filled in black.

Figure 1-4 shows the RD image after processing. The thread-like lines are RD, and VH appears in the left half of the eyeball. As can be seen from the image, the thread-like lines are highlighted in white after processing. The number of training runs was 1000, and the training data are summarized in Table 3. However, to improve the probability of diagnosing RD, the train and valid sets contained images that were difficult to diagnose as RD, as in proposed iteration 2.

This approach yielded an accuracy of 96.3% for RD, 95.9% for VH, and 98.0% for RD with VH, with an overall accuracy of 96.1% (Table 4).

- Final model (Iteration 4)

Our final model incorporated 5-fold cross-validation along with techniques from previous iterations. The training data used was the same as in Table 3. This model achieved accuracies of 96.6% for RD detection, 99.2% for VH detection, and 98.0% for RD with VH, with an overall accuracy of 97.9%, representing a significant improvement over the initial iterations (Table 4).

Comparisons of each validation experiment

The proposed methods were compared. Proposed method 1 was excluded from the comparison because it was included in proposed methods 2, 3, and 4. The test data included 247 RD images, 246 VH images, and 50 RD with VH images. The results are summarized in Table 4. Compared with the results of proposed method 3, the overall accuracy rate for RD and VH in proposed method 3 was 96.1%, while that of proposed method 2 was 97.0%, and that of proposed method 4 was 97.9%.

However, for the RD with VH data, proposed method 2 had the lowest accuracy, whereas proposed methods 3 and 4 had similar accuracy rates. The corresponding confusion matrices for Methods 2–4 are shown in Figure 2, providing a detailed breakdown of image-level predictions and misclassification patterns across the three classes.

Ablation study of preprocessing components

Table 5 summarizes the results of the ablation study evaluating the effect of individual preprocessing components. The baseline YOLO-only configuration achieved the highest overall image-level accuracy 95.8%. Introducing UM resulted in a slight reduction in overall accuracy 95.0%, while thresholding alone and the combination of UM and thresholding further reduced overall accuracy to 92.4% and 91.9%, respectively.

These results indicate that preprocessing does not universally improve overall classification accuracy under the simplified ablation setting. However, the preprocessing steps were designed to enhance structural interpretability and to stabilize detection behavior in visually complex cases, particularly when RD features are obscured by VH related patterns. As such, the ablation analysis serves to clarify the functional role of each preprocessing component rather than to represent the final optimized performance achieved in the validation experiments.

Discussion

Ocular ultrasonography has long been used as a diagnostic tool and is particularly effective when the fundus is not visible owing to media

opacity. However, in actual clinical practice, it is often difficult to accurately diagnose RD based on ocular ultrasound findings. Therefore, we investigated the potential of AI-assisted diagnosis and developed and validated a deep-learning algorithm with a high accuracy for detecting RD and VH.

Chen et al. previously reported a similar study[11]. In that study, DCNN1 was first used to filter out unqualified images, and U-net++ was then used to segment the eyeball (DCNN2). DCNN3 was used to distinguish abnormal images from normal asymptomatic images. DCNN4, DCNN5, DCNN6, and DCNN7 were used to identify PVD, VH, RD, and other pathologies. Except for the U-net++ used in DCNN2, Chen et al. developed their own models. In this study, we developed and validated a deep-learning algorithm for detecting RD and VH on ocular ultrasound images. Our study showed an improvement in accuracy of approximately 3%–7% compared with previous studies, with values of 94% (RD) and 92% (VH).

Other recent studies have also reported high performance in AI-based analysis of ocular B-scan images. Wei et al. used an Automated Machine Learning (AutoML) framework to detect multiple pathologies (e.g., RD, VH), achieving an Area Under the Precision–Recall Curve (AUPRC) of 0.9880 and accuracy of 95.1% for RD[18]. Ye et al. developed Dual-Path Lesion Attention Network (DPLA-Net) for fine-grained classification, reaching an AUC of 0.997 for RD and improving junior ophthalmologists' diagnostic accuracy from 0.696 to 0.919[19]. Caki et al. reported a segmentation-assisted classification model with a 96.5% F-

score, correctly identifying 98.1% of PVD as non-RD—an important distinction for treatment planning[20]. The high accuracy of our model (96.6% for RD and 99.2% for VH) is comparable to or exceeds that of previous reports. Notably, our model achieved 98.0% accuracy in detecting RD with VH, which is clinically important because these cases often present diagnostic difficulties.

The ablation study further clarifies the role of image preprocessing in the proposed framework. While the baseline YOLO-only configuration achieved the highest overall image-level accuracy under a simplified ablation setting, the preprocessing steps were not designed to maximize aggregate accuracy alone. Instead, UM and thresholding were introduced to enhance the visibility of RD-related linear structures and suppress VH related diffuse patterns, thereby improving interpretability and contributing to more stable detection behavior in visually complex cases. As shown by the quantitative comparisons in Table 5, these preprocessing steps do not maximize overall image-level accuracy in isolation but contribute to improved robustness in challenging RD-with-VH scenarios, which helps explain why their benefit is more apparent in the validation experiments than in isolated ablation analysis.

Several limitations should be noted. The dataset was retrospectively collected from a single tertiary referral institution and does not represent a screening population; all ultrasound images were obtained from patients with clinical suspicion of RD and/or VH. Images were acquired during routine clinical examinations using the institution’s standard ocular ultrasonography equipment, without explicit evaluation across different

device manufacturers or standardized acquisition protocols. Although the dataset encompassed a broad range of disease severity, from subtle to advanced cases confirmed intraoperatively, these factors may limit generalizability to other clinical settings.

In addition, all image selection and labeling were performed by a single experienced ophthalmologist, which may introduce observer-related bias. The absence of an independent external validation set means that the reported performance may be influenced by institution-specific image characteristics, and the dataset imbalance may have affected model training. The present study did not include a direct comparison between the diagnostic performance of the proposed algorithm and that of experienced ophthalmologists. Therefore, the findings should be interpreted as demonstrating technical feasibility rather than immediate clinical equivalence. In routine practice, the diagnosis of RD is based on integrated clinical information beyond B-scan ultrasonography alone. Future studies should evaluate clinician–algorithm comparisons and real-time performance within clinical workflows. Furthermore, although UM preprocessing is computationally fast (<0.02 s per image) and compatible with real-time YOLO inference, reliance on enhanced images during live scanning may affect workflow efficiency in busy emergency settings. Finally, certain visually complex ocular ultrasound images remain challenging, particularly when features of RD and VH overlap; representative examples are shown in Supplementary Figure S2.

In conclusion, we developed and validated a YOLO-based deep-

learning algorithm for detecting RD and VH from ocular ultrasound images. By combining YOLO's efficient object detection framework with external UM preprocessing, the system enables rapid image analysis while maintaining high diagnostic accuracy. This approach has the potential to support timely and reliable decision-making in emergency settings, particularly when prompt intervention is essential to preserving vision. Future work should include multi-institutional validation, seamless integration into clinical ultrasound workflows, and prospective evaluation of its impact on patient outcomes.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgements

We would like to express our gratitude to the following undergraduate students from the Faculty of Engineering at the University of Miyazaki for their valuable contribution to the data analysis in this research: Yuichiro Uchida, Ruon Kanda, Taiyo Nagayama. Their assistance in analyzing the data was crucial to the success of this study.

Funding Declaration

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

Compliance with ethical standards

This study was approved by the Institutional Review Board of University of Miyazaki (approval number: O-1065).

Conflict of interest

About this work, all authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements) or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

The authors declare that they have no conflicts of interest regarding the publication of this paper.

1. Mitry, D., Charteris, D. G., Fleck, B. W., Campbell, H. & Singh, J. The epidemiology of rhegmatogenous retinal detachment: geographical variation and clinical associations. *Br. J. Ophthalmol.* **94**, 678–684 (2010). doi:10.1136/bjo.2009.157727
2. Spraul, C. W. & Grossniklaus, H. E. Vitreous Hemorrhage. *Surv. Ophthalmol.* **42**, 3–39 (1997). doi:10.1016/S0039-6257(97)84041-6.
3. Kendall, C. J. *et al.* Diagnostic Ophthalmic Ultrasound for Radiologists. *Neuroimaging Clin. N. Am.* **25**, 327–365 (2015). doi:10.1016/j.nic.2015.05.001
4. Parchand, S., Singh, R. & Bhalekar, S. Reliability of ocular ultrasonography findings for pre-surgical evaluation in various vitreo-retinal disorders. *Semin. Ophthalmol.* **29**, 236–241 (2014). doi:10.3109/08820538.2013.821506
5. Lahham, S. *et al.* Point-of-Care Ultrasonography in the Diagnosis of Retinal Detachment, Vitreous Hemorrhage, and Vitreous Detachment in the Emergency Department. *JAMA Netw Open* **2**, e192162 (2019). doi:10.1001/jamanetworkopen.2019.2162
6. Ting, D. S. W. *et al.* Artificial intelligence and deep learning in ophthalmology. *Br. J. Ophthalmol.* **103**, 167–175 (2019). doi:10.1136/bjophthalmol-2018-313173
7. Schmidt-Erfurth, U., Sadeghipour, A., Gerendas, B. S., Waldstein, S. M. & Bogunović, H. Artificial intelligence in retina. *Prog. Retin. Eye Res.* **67**, 1–29 (2018). doi:10.1016/j.preteyeres.2018.07.004
8. Gulshan, V. *et al.* Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus

Photographs. *JAMA* **316**, 2402–2410 (2016). doi:10.1001/jama.2016.17216

9. De Fauw, J. *et al.* Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat. Med.* **24**, 1342–1350 (2018). doi:10.1038/s41591-018-0107-6
10. Adithya, V. K. *et al.* Development and validation of an offline deep learning algorithm to detect vitreoretinal abnormalities on ocular ultrasound. *Indian J. Ophthalmol.* **70**, 1145–1149 (2022). doi:10.4103/ijo.IJO_2119_21
11. Chen, D. *et al.* A Deep Learning Model for Screening Multiple Abnormal Findings in Ophthalmic Ultrasonography (With Video). *Transl. Vis. Sci. Technol.* **10**, 22 (2021). doi:10.1167/tvst.10.4.22
12. Singh, L. K., Garg, H., Pooja, N. A. & Khanna, M. Performance analysis of machine learning techniques for glaucoma detection based on textural and intensity features. *International Journal of Innovative Computing and Applications.* **11**, 216–230 (2020). doi:10.1504/IJICA.2020.111230
13. Singh, L.K., Khanna, M., Garg, H. *et al*. A three-stage novel framework for efficient and automatic glaucoma classification from retinal fundus images. *Multimed Tools Appl* **83**, 85421–85481 (2024). doi:10.1007/s11042-024-19603-z
14. Singh, L. K. & Khanna, M. Introduction to artificial intelligence and current trends. *Innovations in Artificial Intelligence and Human-Computer Interaction in the Digital Era.* Academic Press; 31–66 (2023). doi:10.1016/B978-0-323-99891-8.00001-2

15. Sorenson, J. A., Niklason, L. T. & Nelson, J. A. Photographic unsharp masking in chest radiography. *Invest. Radiol.* **16**, 281–288 (1981). doi: 10.1097/00004424-198107000-00007
16. Panetta, K., Zhou, Y., Agaian, S. & Jia, H. Nonlinear unsharp masking for mammogram enhancement. *IEEE Trans. Inf. Technol. Biomed.* **15**, 918–928 (2011). doi:10.1109/TITB.2011.2164259
17. Edla, D. R., Simi, V. R. & Joseph, J. A Noise-robust and Overshoot-free Alternative to Unsharp Masking for Enhancing the Acuity of MR Images. *J. Digit. Imaging* **35**, 1041–1060 (2022). doi:10.1007/s10278-022-00585-z
18. Wei, Q., Chen, Q., Zhao, C. & Jiang, R. Performance of automated machine learning in detecting fundus diseases based on ophthalmologic B-scan ultrasound images. *BMJ Open Ophthalmol.* **9**, (2024). doi:10.1136/bmjoph-2024-001873
19. Ye, X. *et al.* Ocular disease detection with deep learning (fine-grained image categorization) applied to ocular B-scan ultrasound images. *Ophthalmol. Ther.* **13**, 2645–2659 (2024). doi:10.1007/s40123-024-01009-7
20. Caki, O. *et al.* Automated detection of retinal detachment using deep learning-based segmentation on ocular ultrasonography images. *Transl. Vis. Sci. Technol.* **14**, 26 (2025). doi:10.1167/tvst.14.2.26 doi:10.1167/tvst.14.2.26

Figure and Table legends

Figure 1-1: Images of Retinal Detachment and Vitreous Hemorrhage. (a) Ocular ultrasound image showing RD. (b) Ocular ultrasound image showing VH. These images demonstrate the characteristic ultrasound appearance of these two distinct ocular pathologies.

Figure 1-2: Unsharp Masking Applied to RD and VH Images. (a) Original RD image. (b) RD image after applying unsharp masking. (c) Original VH image. (d) VH image after applying unsharp masking. This image processing step enhances overall image clarity, improving the visibility of key features in both RD and VH cases.

Figure 1-3: Binarized and Thresholded Images. Demonstration of thresholding and binarization techniques applied to ocular ultrasound images. (a) Original RD with VH image. (b) The same image after applying thresholding and binarization. (c) RD image after thresholding/binarization. (d) VH image after thresholding/binarization. These techniques help in segmenting and highlighting specific features within the ultrasound images.

Figure 1-4: The Comparison of Original and Enhanced Retinal Images. (a) Original retinal ultrasound image. (b) Enhanced version after processing. The thread-like lines are RD, and VH appears in the left half of the eyeball. As can be seen from the image, the thread-like lines are highlighted in white after processing. This enhancement significantly improves the visibility of the RD, making it easier for clinicians to identify and assess the extent of the detachment.

Figure 2: Figure 2. Confusion matrix analysis of the final classification

results.

Confusion matrices for (a) Method 2, (b) Method 3, and (c) Method 4, showing the distribution of image-level predictions across retinal detachment (RD), vitreous hemorrhage (VH), and RD with VH categories. The numbers in each cell indicate the count of images for the corresponding actual and predicted classes. These results correspond to the class-wise accuracy values reported in Table 4.

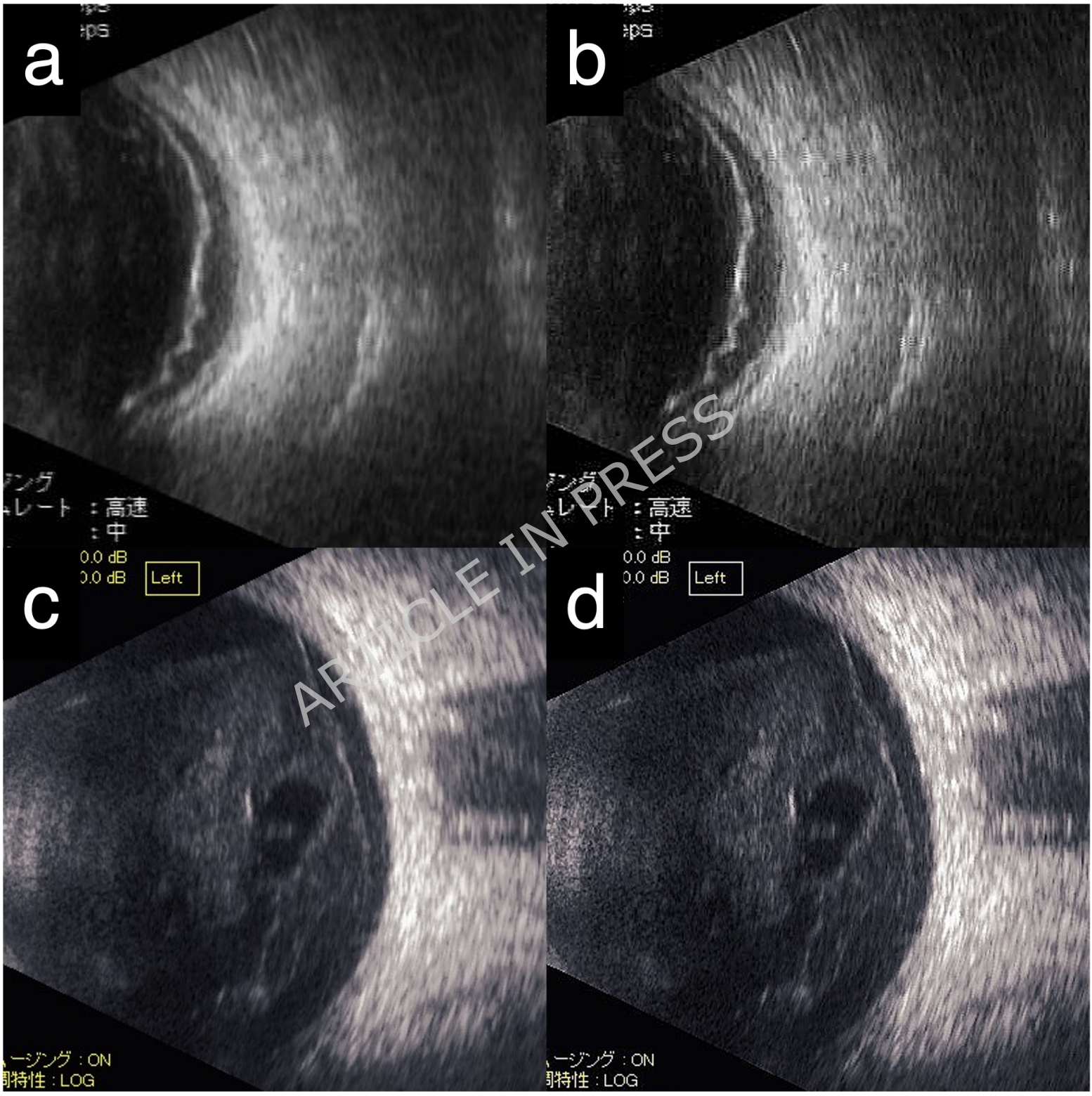
Table 1: An Accuracy Comparison of YOLO Models for RD and VH Detection. This table presents a comparative analysis of various YOLO (You Only Look Once) model versions in detecting retinal detachment (RD) and vitreous hemorrhage (VH).

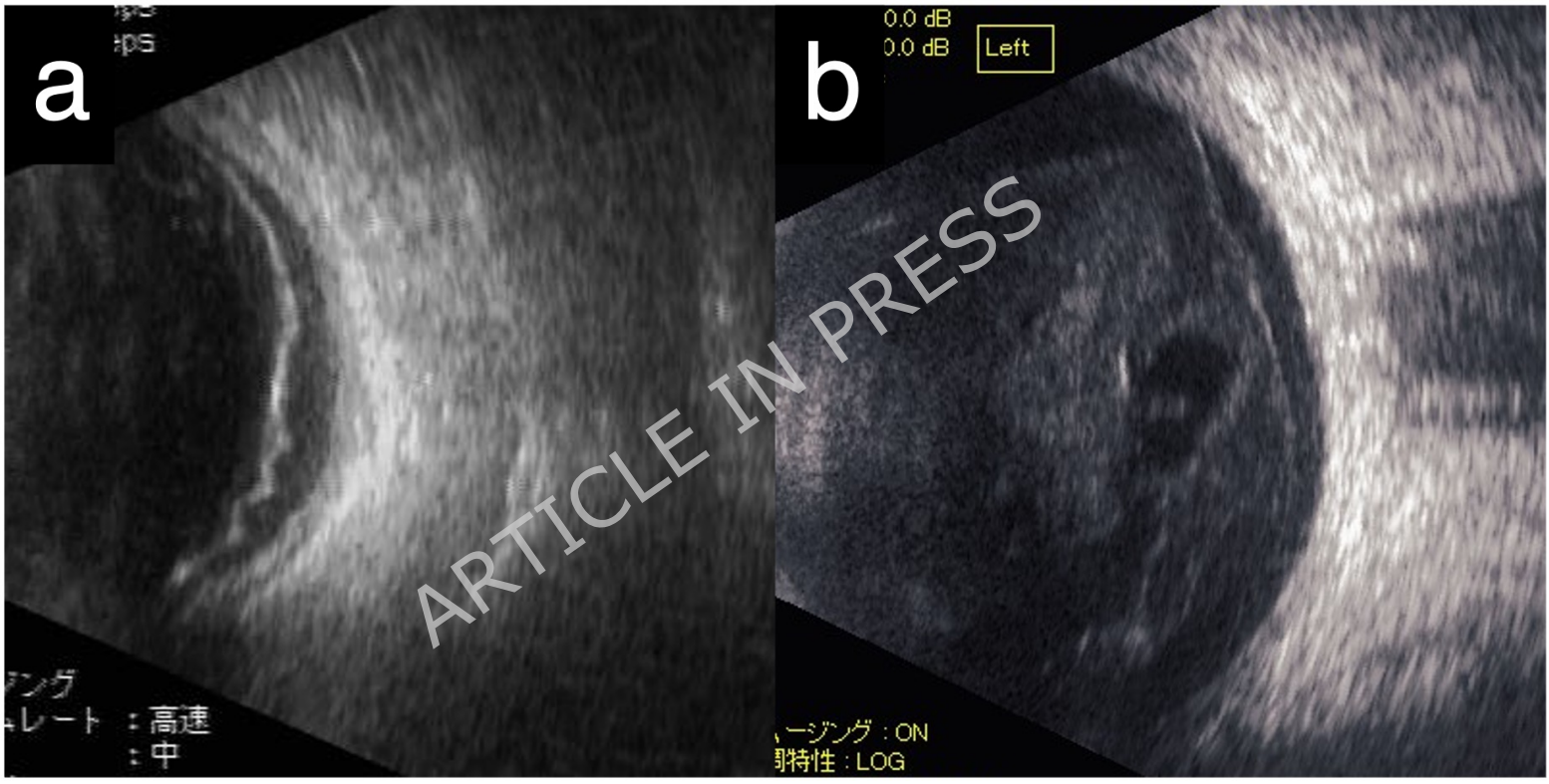
Table 2: Accuracy Rates of Proposed Methods 1, 2, and 3 for RD, VH, and RD with VH, respectively. This table summarizes the accuracy rates of the three proposed methods in detecting RD, VH, and RD with VH.

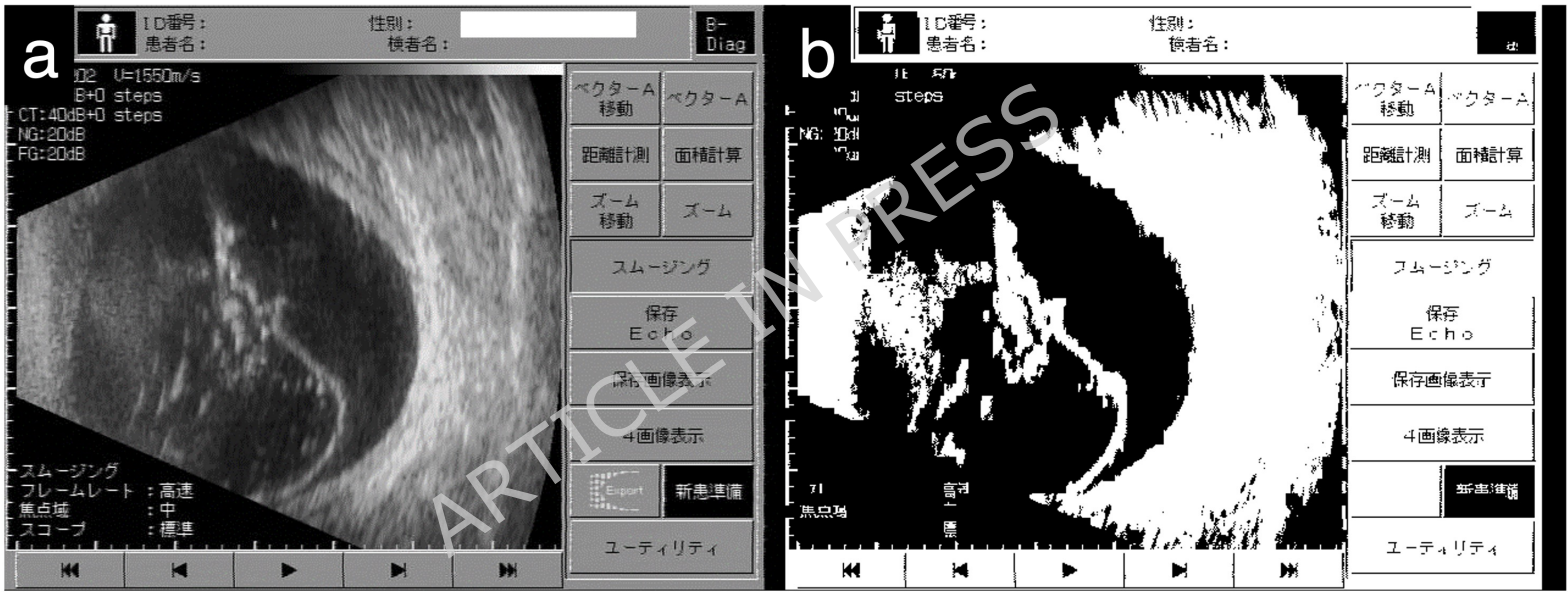
Table 3: Original Training Data for Proposed Methods 2 and 3. This table outlines the composition of the original training dataset used for Proposed Methods 2 and 3.

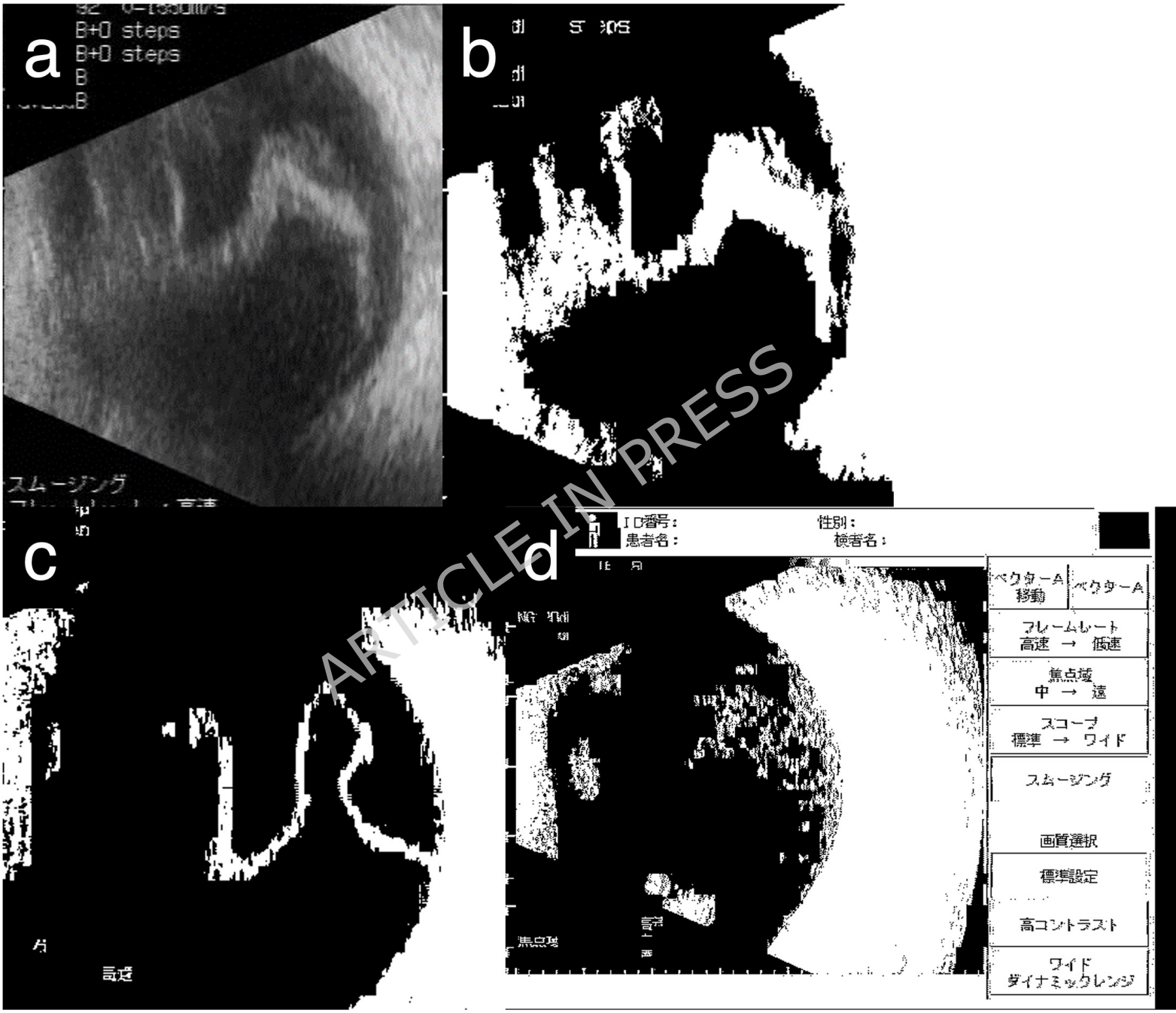
Table 4: Combined Accuracy Rates for Proposed Methods 2, 3, and 4. This table presents the combined accuracy rates achieved by Proposed Methods 2, 3, and 4.

Table 5: Ablation study of preprocessing components. Overall image-level classification accuracy for YOLOv5 under different preprocessing configurations. Performance is shown for YOLOv5 without preprocessing (YOLO only), with unsharp masking (UM), with thresholding only, and with the combination of UM and thresholding. All configurations were evaluated using the same fixed test dataset and identical evaluation criteria.

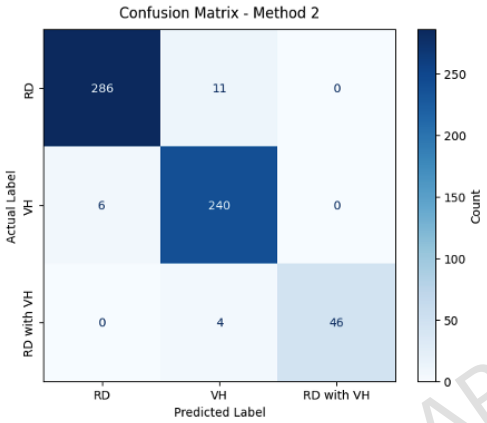




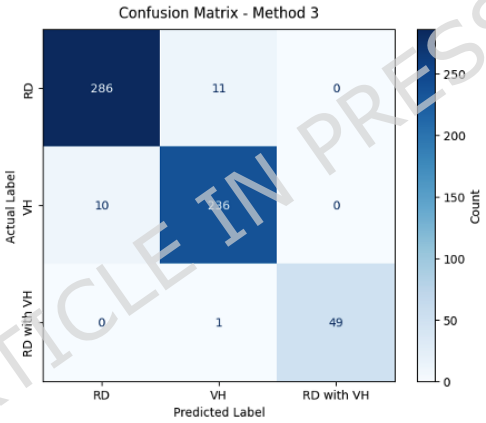




(a)



(b)



(c)

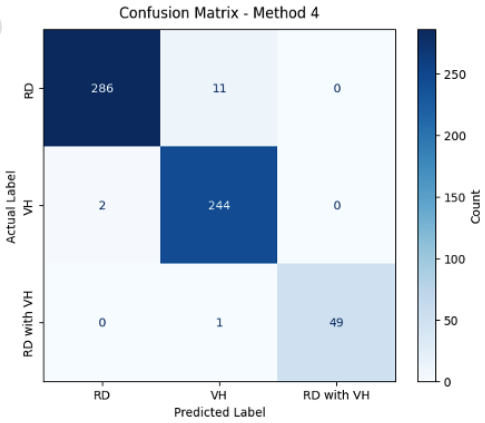


Table 1: Accuracy Comparison of YOLO Models for RD and VH Detection

Model	**Condition**	**Correct (Count)**	**Correct (%)**	**Incorrect (Count)**	**Incorrect (%)**	**Undetected (Count)**	**Undetected (%)**	**Total (Count)**	**Total (%)**
YOLOv2-tiny	RD	72	61	46	39	-	-	118	100
	VH	31	97	0	0	1	3	32	100
YOLOv5	RD	91	77.1	27	22.9	-	-	118	100
	VH	32	100	0	0	-	-	32	100
YOLOv7	RD	60	50.8	58	49.2	-	-	118	100
□	VH	32	100	0	0	-	-	32	100

RD: Retinal Detachment, VH: Vitreous Hemorrhage

Table 2: Accuracy Rates of Proposed Method 1,2 and 3 for RD, VH, and RD with VH

	Condition	**Correct (Count)**	**Correct (%)**	**Incorrect (Count)**	**Incorrect (%)**	**Total (Count)**	**Total (%)**
Method 1	RD	165	89.2	20	10.8	185	100
	VH	167	90.3	18	9.7	185	100
	RD with VH	53	82.8	11	17.2	64	100
Method 2	RD	286	96.3	11	3.7	297	100
	VH	240	97.6	6	2.4	246	100
	RD with VH	46	92.0	4	8.0	50	100
Method 3	RD	286	96.3	11	3.7	297	100
	VH	236	95.9	10	4.1	246	100
	RD with VH	49	98.0	1	2.0	50	100

RD: Retinal Detachment, VH: Vitreous Hemorrhage

Table 3: Original Training Data for Proposed Method 2 and 3

	Category	**Count**	**Percentage (%)**
Method 2	Training Data	2,000	54.2
	Validation Data	1,143	31.0
	Test Data	543	14.7
	Total	3,686	100
Method 3	Training Data	2148	58.3
	Validation Data	995	30.0
	Test Data	543	14.7
	Total	3686	100

Table 4: Combined Accuracy Rates for Proposed Methods 2, 3, and 4

Condition	**Method 2 Accuracy (%)**	**Method 3 Accuracy (%)**	**Method 4 Accuracy (%)**
RD	96.3	96.3	96.6
VH	97.6	95.9	99.2
RD with VH	92.0	98.0	98.0
Overall Accuracy	97.0	96.1	97.9

Table 5: Ablation Study of Preprocessing Components

Configuration	**Overall accuracy (%)**
YOLO only	95.8
YOLO + UM	95.0
YOLO + Thresholding only	92.4
YOLO + UM + Thresholding	91.9