Sensors and Materials, Vol. 37, No. 12 (2025) 5447–5464
MYU Tokyo

S & M 4256

Visualization Ultrasound Scanning System for Early Hemodialysis Fistula Stenosis Screening Based on Combining a YOLOv11 Classifier and Mixed Reality

Jun-Yi Lin,[1] Pi-Yun Chen,[1*] Neng-Sheng Pai,[1] Ping-Tzan Huang,[2] Chao-Lin Kuo,[3] Chien-Ming Li,[4] and Chia-Hung Lin[1**]

[1]Department of Electrical Engineering, National Chin-Yi University of Technology, Taichung City 41170, Taiwan
[2]Department of Biomechatronics Engineering, National Pingtung University of Science and Technology, Pingtung 91201, Taiwan
[3]Department of Maritime Information and Technology, National Kaohsiung University of Science and Technology, Kaohsiung City 80543, Taiwan
[4]Tainan Municipal Hospital, Infectious Disease Division of Internal Medicine Department, Tainan City 701, Taiwan

(Received April 24, 2025; accepted December 1, 2025)

Keywords: hemodialysis (HD), arteriovenous fistula (AVF), access stenosis, YOLOv11, degree of stenosis, luminal change rate (LCR)

Chronic kidney disease (CKD) is an irreversible condition that progressively impairs kidney function, leading to the need for hemodialysis or kidney transplantation. For hemodialysis (HD) patients, their dialysis access, such as the arteriovenous fistula, serves as a vital lifeline for efficient blood filtration and waste removal. Because HD patients undergo hemodialysis three times a week, the repeated needle punctures may gradually lead to vascular tissue hyperplasia, vascular intimal thickening, and fibrosis. Underlying chronic conditions in HD patients such as diabetes, cardiovascular diseases, or metabolic syndrome can further contribute to inflammation, infection, or access stenosis or occlusion. Hemodialysis fistula stenosis screening is essential to preserve vascular access, as early detection enables timely intervention, helping to ensure adequate dialysis, and minimize the risk of thrombosis or access failure. Hence, for early HD fistula stenosis detection, in this study, we propose to train a YOLOv11-based classifier with image enhancement, feature extraction, and pattern recognition to enable automatic stenosis detection. With B-mode ultrasound imaging data stream, the proposed classifier performs the automated image segmentation for distinguishing vascular wall contours, vascular lumen, and thrombotic regions and then presents them in a colored visualization pattern. By tenfold cross-validation, the classifier model based on the proposed "Two Dimension (2D) Fractional-Order Convolution Operation (FOCO) + YOLOv11 + Stochastic Gradient Descent Optimizer" achieved the following average performance metrics: Precision of 100.00 ± 0.00%, Recall of 90.00 ± 0.01%, F1 score of 0.9474 ± 0.0032, and Accuracy of 94.74 ± 0.02% for the identification of normal and stenotic regions. On the basis of the segmentation results, the key vascular quantification indexes, such as the degree of stenosis and luminal change rate, are computed to

*Corresponding author: e-mail: chenby@ncut.edu.tw
**Corresponding author: e-mail: eechl53@gmail.com
https://doi.org/10.18494/SAM5707

ISSN 0914-4935
https://myukk.org/

evaluate the access stenosis levels. Furthermore, by utilizing spatial information, we can reconstruct the sequential ultrasound images into a 3D visualization fistula model through Mixed Reality display devices, offering an intuitive and interactive assessment of fistula health and potential blockages.

1. Introduction

According to the Taiwan Society of Nephrology, in 2024, 1 in every 8 persons (approximately 364,000 individuals) suffers from chronic kidney disease (CKD), with a growth rate of approximately 26.4% over the past decade. The number of patients receiving hemodialysis (HD) treatment in Taiwan has exceeded 90000, and the number of affected individuals has been steadily increasing, with the age of onset gradually decreasing. Thus, the highest prevalence rate has earned the nickname "New National Disease".[1,2] Taiwan's HD population now surpasses that of countries such as the United States and Singapore in Asia. HD treatment can be divided into peritoneal dialysis (PD) and HD treatments. The number of people receiving HD treatment in Taiwan is approximately 88000, whereas about 6500 people are undergoing home PD treatment. The most commonly used method for HD treatment is the standard regimen of three times per week, with each time lasting approximately 3 to 4 hours. During HD, a special blood access pathway and circulation loop are required, as shown in Fig. 1(a), which indicates the HD circulation loop and artificial kidney (dialysis machine). Hence, the blood containing uremic toxins can be drawn from the body through this special pathway for filtering the blood, the so-called arteriovenous shunt (AVS), known as an arteriovenous fistula (AVF) and an arteriovenous graft (AVG).[3–5]

With long-term dialysis treatment, HD patients may develop vascular aneurysms (swelling) in their arms. Maintaining a sufficient blood flow is essential to achieving the high blood flow and optimal dialysis dose. When the AVS undergoes pathological changes that will lead to complications such as stenosis (narrowing of the AVS), thrombosis (blood clot formation), and aneurysm formation, these changes affect normal blood flow, resulting in the early symptom of low intra-shunt blood flow. This progressively impacts the normal dialysis blood flow, ultimately resulting in vascular access failure. If vascular abnormalities can be detected in the early stage before stenosis occurs, the risk of interventional treatment can be significantly reduced. Early noninvasive examinations can be performed by using medical physiological evaluation, including inspection, palpation, and auscultation methods.[6–9] In visual inspection, the first step is to observe whether the fistula appears collapsed or flattened, and to check for swelling, redness, or bruising on the hand. The palpation method is used to evaluate the elasticity and thrill within the fistula. If there is a localized, pronounced thrill or warmth, it may indicate narrowing access, stenosis, or infection within the fistula. The auscultation method involves using a regular stethoscope or an electronic stethoscope to examine the arteriovenous anastomosis site and move upward along the AVS access. The presence of murmur, high-pitched bruit, or a diminished blood flow sound may indicate stenosis or occlusion in the fistula.[8,9] However, these physical examination methods may only detect localized narrowing or blockage sites; they cannot precisely identify the anatomical location or severity of the fistula lesion.

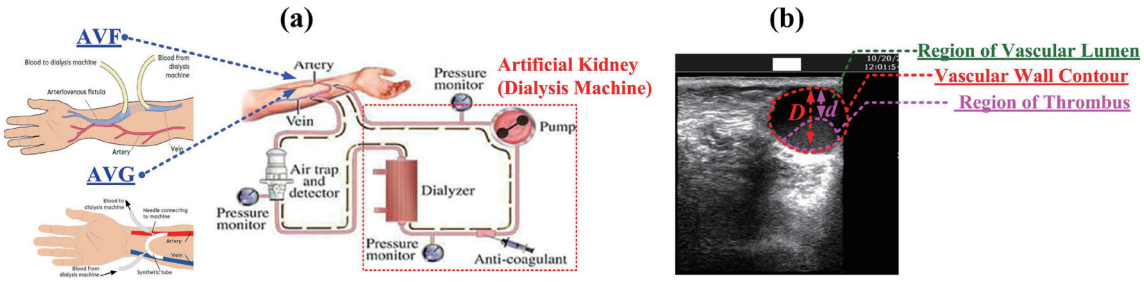

Fig. 1. (Color online) (a) HD circulation loop and AVS types, including AVF and AVG. (b) B-Mode vascular lumen ultrasound image, including vascular wall contours, region of vascular lumen, and region of thrombus.

The aforementioned examination methods can only provide a general indication of potential lesions within the fistula. However, they have some limitations, that is, they are insufficient for quantitatively analysis to evaluate the degree of stenosis (DOS), thrombosis, or complete blockage, and also, they cannot precisely localize the anatomical site of the lesion. Medical imaging techniques also serve as noninvasive methods for directly evaluating the internal condition of the fistula, including ultrasound (US) imaging, computed tomography angiography (CTA), and conventional angiography.[10–12] Li *et al.* proposed an end-to-end deep learning (DL) convolutional neural network (CNN) for media–adventitia (MA) segmentation in intravascular ultrasound (IVUS) images[13] to automatically delineate the media–adventitia border, luminal region, and calcified plaque. A total of 713 grayscale IVUS images were collected from 18 enrolled subjects with atherosclerosis and divided into training and testing image datasets to train and evaluate three classifiers based on U-Nets (Convolutional Networks for Biomedical Image Segmentation, Department of Computer Science, 2015, University of Freiburg).[13–15] This model addressed calcification misclassification and mitigates external pixel interference from plaques, and the Dice, Tversky, and Focus loss functions[13] were employed to optimize and evaluate the segmentation of the proposed performance classifier. The predictive performance was evaluated using three metrics, namely, mean Precision (%), Sensitivity, and Specificity, by comparing the predicted results with the ground truth. However, the IVUS method remained an invasive diagnostic procedure. In addition, the limitations of the U-Nets-based classifier included the following: (1) for training large-scale or complex images, the training stage required a longer time and higher computational resources; (2) owing to pixel-level classification, its model was prone to overfitting problems during image segmentation; (3) its model had a bottleneck in image preprocesses, such as image resizing, rotating, and scaling; (4) data augmentation was required to enhance generalization. Moreover, the U-Nets model was relatively complex, involving multiple convolutional layers and downsampling and upsampling operations, which further increased the computation time and memory requirements in image segmentation.

Among the aforementioned imaging methods, US imaging is the first-line and noninvasive diagnostic method to detect the vessel diameter, blood flow velocity (BFV), and thrombus formation, as shown in Fig. 1(b), owing to its advantages such as low examination costs, affordability, the absence of radiation exposure, and no requirement for contrast agents. Each examination takes approximately 15 min, allowing for a preliminary identification of lesion

locations with >70% diagnostic accuracy, while also recording the scanned US images. By Doppler measurements, key physiological parameters such as BFV, peak systolic velocity (PSV), end diastolic velocity (EDV), and resistance estimation[16,17] can be obtained. Additionally, with gray-scale B-mode US images, 2D imaging can effectively scan the anatomical structure of vascular stenosis or thrombus, providing the visualization views of the vessel depth, vessel lumen, and vascular wall contours. Hence, in this study, we utilized a visualization US scanning system combined with deep learning (DL)-based object detection (OD) and pattern recognition (PR) classifiers to automatically identify key structures within the fistula, including vascular wall contours, the region of vascular lumen, and the region of thrombus. The You Only Look Once (YOLO) (first introduced by Joseph Redmon in 2016)[18–23] family of classifier models, such as YOLOv9, YOLOv10, YOLOv11, and YOLOv12, are capable of performing both OD and PR tasks simultaneously. Among the aforementioned YOLO models, the recent YOLOv12 model[22,23] incorporates a scaling attention module (SAM) in the backbone layer to enhance the recognition of small and distant objects. In the neck layer, a high-resolution feature fusion (HRFF) mechanism is also integrated to improve the fusion of multi-level information and features for extracting objects' spatial and edge features, thereby strengthening OD and PR in complex environments. In the head layer, YOLOv9–YOLOv10 and YOLOv12 have the dynamic label assignment (DLA) mechanism[18,22–24] to adaptively assign learning weights on the basis of current feature maps. This process progressively reduces the probability of incorrect positive and negative sample assignments, thus improving training stability and accuracy. YOLOv10 has a multi-scale feature fusion (MSFF) module[25,26] that integrates feature maps from different levels (small-, medium-, and large-scale objects), which can guide the classifier model to focus on regions of interest (ROIs) in computed tomography (CT) images and to improve its contextual understanding and capacity to capture intricate and diverse structures.[25] The backbone layer uses computationally efficient convolutional operations to reduce the number of network parameters, making the model more lightweight and requiring less memory. YOLOv11[21,22,27] improves the backbone and neck layers to enhance feature extraction for different scale objects, enabling more accurate classification across different object sizes. In addition, its design and training optimization process provide higher computational efficiency, maintaining an optimal balance among training speed, accuracy, and overall performance. Hence, the YOLOv11-based classifier is built upon the integration of "US image enhancement + feature extraction + pattern recognition + learning optimizer". Using these extracted features, we can compute the DOS and luminal change rate (LCR) to evaluate the severity levels of fistula narrowing and stenosis. Additionally, reconstructing a visualized 3D fistula pathway model helps assess the fistula's internal health status.

2. Materials and Methods

2.1. B-mode US imaging and data collection

In this study, we employed handheld US scanning to acquire B-mode imaging datasets. In the B-mode (brightness mode) imaging manner, the echoes generated by US waves interacting with

tissue structures are converted into a grayscale representation, where different brightness levels correspond to varying echo intensities, with stronger echoes appearing brighter (reflective structures such as fibrous tissue) and weaker echoes appearing darker (often from softer tissues or fluid-filled regions). Then, the returning echoes can be detected by the transducer and converted into electrical signals. By systematically scanning a specific region, either by manually sweeping the transducer or through mechanical steering, the scanning system collects echo data from multiple points. Thus, these data points are spatially arranged to construct a 2D cross-sectional image, effectively representing the scanned tissue and its internal structures. The US scanning provides a noninvasive and nonradiative imaging method. In medical US imaging, US probes operate within a frequency range of 3.5–7.5 MHz, producing images with a resolution of 2–3 mm, whereas probes of lower frequency allow for deeper tissue imaging. Hence, the 7.5 MHz operating frequency can penetrate up to 3 cm beneath the skin, capturing arterial lumen and vascular contour details, enabling measurements of vessel wall thickness and the detection of lipid accumulation or tissue proliferation inside the vessels, as seen in Fig. 1(b).

According to the principle of the Doppler effect, the colorized visual representations can be used to display BFV within blood vessels or hemodynamic parameters can be represented using vascular waveform signals, such as PSV and EDV. Additionally, the image format can be converted to the Digital Imaging and Communications in Medicine (DICOM) digital format to facilitate digital image processes, such as resizing and localized magnification, to obtain a clearer view of the vascular lumen. At specific locations within the fistula pathway, the *DOS* in the fistula pathway or arterial vessels (carotid artery) can be further computed using the following general formula:[8–10]

$$DOS\% = (1 - \frac{d^2}{D^2}) \times 100\%, \tag{1}$$

where D is the normal vessel diameter and d is the diameter of the narrowed access, both of which can be obtained from B-mode US images for the quantitative analysis of stenosis levels.

By configuring the DICOM Gateway on the US scanning device, the US images (X-ray, CTA, or magnetic resonance images) can be simultaneously displayed on a monitor and seamlessly transmitted with picture archiving and communication systems (PACS) and hospital information systems (HIS) for further processing and analysis. Through the DICOM Gateway acquisition interface, the data stream of US images can be captured at a rate of 30 frames per second, with each raw image at a resolution of 640 × 480 pixels. The image format is then converted into the DICOM digital format and stored in various formats, such as Joint Photographic Experts Group (JPEG), Portable Network Graphics (PNG), and Graphics Interchange Format (GIF) files, which include skin tissue and vessels within an image for further processing and analysis. With a specific bounding box to exclude the black background and unrelated image data, the remaining region including vessels can be extracted at a resolution of approximately 393 × 450 pixels, as seen in Fig. 2. This preprocessing step helps reduce computational time and memory consumption for subsequent image segmentation processing.

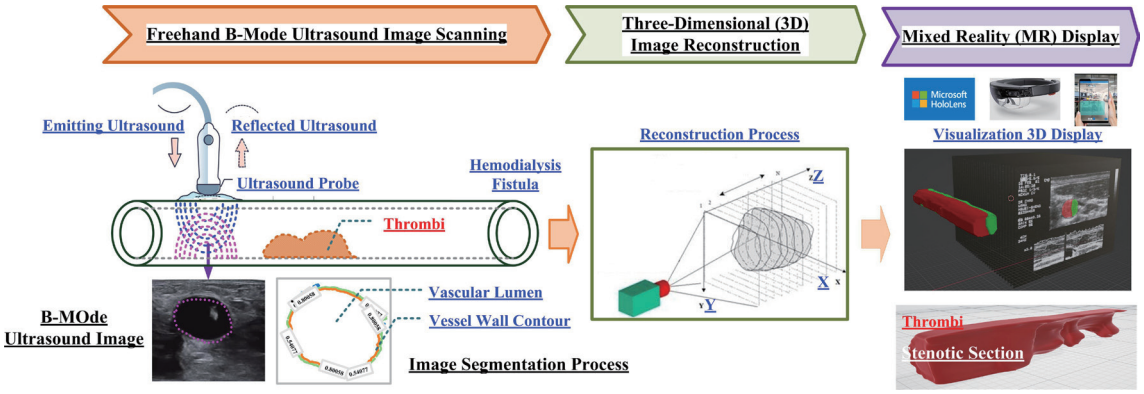

Fig. 2. (Color online) Combining handheld B-mode ultrasound scanning, automated image segmentation, 3D image reconstruction of the fistula model, and 3D model visualization for the stenosis detection workflow.

This study was conducted in a hospital experimental setting, and US images involving 18 enrolled subjects were collected (image source: National Cheng Kung University Hospital, Tainan City, Taiwan). To protect patient privacy, all collected B-mode US images were de-identified. The datasets include the following: (1) the control group ($n = 6$) consisting of subjects with no fistula stenosis ($DOS \leq 30\%$) and (2) the experimental group ($n = 12$) consisting of subjects with varying levels of stenosis [mild stenosis ($30\% < DOS \leq 50\%$) and severe stenosis ($DOS > 50\%$)]. On the basis of manual annotation and verification by experienced clinical physicians, the datasets were divided into three classes: Class I: $DOS \leq 30\%$ (control group), Class II: $30\% < DOS \leq 50\%$, and Class III: $DOS > 50\%$.[8–10] To facilitate stenosis assessment, approximately 650 US images were collected, including 300 images with no stenosis ($DOS \leq 30\%$) and 350 images with varying levels of stenosis ($30\% < DOS \leq 50\%$ and $DOS > 50\%$). These datasets will be used for training and validating the YOLOv11-based classifier[21,22,27] for image segmentation and classification in subsequent analysis.

2.2. YOLOv11-based classifier design

In this study, the YOLOv11 model[21,22,27] was employed to construct the classifier for both image segmentation and classification, as seen in the network architecture in Fig. 3, which has evolved progressively from YOLOv8 to YOLOv10,[18–20,28,29] retaining the advantages of its predecessor models. The proposed YOLOv11's network structure consists of backbone, neck, and head layers. As shown in Fig. 3, the details are outlined as follows.

- The backbone layer serves as the feature extractor with the 2D fractional-order convolution (Conv.) operation and multi-scale Conv. operations, such as applying 2D fractional-order Conv. operation with the appropriate fractional-order parameters ($v = 0.1$–0.3) to enhance the incoming raw medical images (X-ray, CT, or US images).[30–33] Multiple CBS and C3K2 modules are then utilized for multiple Conv. operations and batch normalization (BN) processing, forming a "Conv + BN + SiLU + C3K2" pipeline to extract multi-scale feature patterns. The CBS module ("Conv. + BN + SiLU", CBS) enhances features by using the

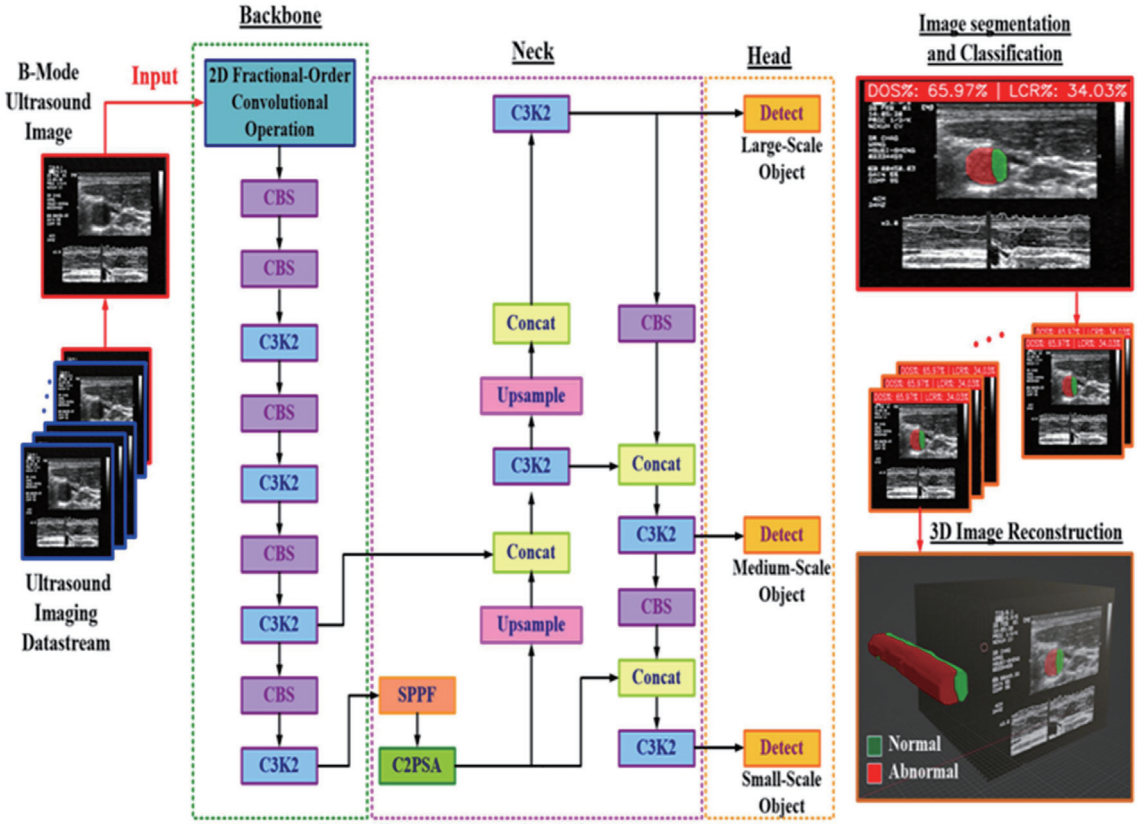

Fig. 3. (Color online) Network architecture of YOLOv11-based classifier consisting of backbone, neck, and head layers.

normalizing process and Sigmoid Linear Unit (SiLU) activation function, which preserves useful gradient information and improves convergence performance. C3K2 [Cross-Stage Partial Network (C3) with Kernel Size Variations (K2)] is designed to extract the local and global features across different scales by leveraging different kernel sizes and multiple receptive fields, which improves the YOLOv11 model's ability to detect fine-grained texture in US images.[21,22] Hence, multi-scale features can be effectively obtained in low-level, mid-level, and high-level feature representations in the backbone layer, which are crucial for achieving precise image segmentation and classification.

- The neck layer functions as an intermediate processing unit, aggregating and enhancing features across different scales by "SPPF + C2PSA + Upsample + C3K2" processing, which is performed to refine and integrate multi-scale features before passing them to the head layer for object segmentation and classification. The spatial pyramid pooling fast (SPPF) technique employs multiple maximum pooling operations with different kernel sizes to extract multi-scale spatial features (both local and global context information) while also reducing computational complexity. The C2 parallel self-attention (C2PSA) technique integrates parallel self-attention (PSA) mechanisms to improve feature refinement and representation learning.[21,22] Upsample feature maps are combined with high-resolution low-level features (lesion, vessel, and tissue within a US image) to refine and integrate multi-scale features for improving object localization and classification capabilities.

- The head layer is employed for object prediction and classification, leveraging the multi-scale feature maps produced by the neck layer. These multi-scale feature maps are divided into large-, medium-, and small-scale detections, enabling automatic object detection for precise specific object localization and classification, thus ensuring accurate detection across various object sizes.

The proposed novel YOLOv11 network architecture is shown in Fig. 3, introducing 2D fractional-order Conv. operation for image enhancement processing and C3K2, SPPF, and C2PSA modules in backbone and neck layers.[21,22] These additional modules significantly enhance the backbone layer's ability to extract multi-scale features while maintaining fast inference and spatial information processing capabilities. In the learning stage, bounding box (BB) regression (BBR), confidence score prediction (CSP), and class probability prediction (CPP) are used to refine box predictions and then to enable multi-class classification in lesion segmentation and classification. The BBR uses the anchor-based mechanisms to predict the target object. The loss function serves as a crucial metric for optimizing the classifier model's performance during the training stage. The BB and segmentation and classification loss functions are summarized as follows.

2.2.1 BB loss function

The BB loss function (BB _LF) is primarily used to estimate the errors between the predicted BB and the ground truth BB in specific object detection, which intends to maximize the overlap region between the predicted BB and the ground truth BB, thereby improving object detection accuracy; the L1 loss (*L1_Loss*) function (mean absolute error) is used in the BBR process to adjust the positional errors between the predicted BB and the ground truth BB, and the general formula is expressed as[21,22]

$$L1\_Loss = \frac{1}{K}\sum_{i=1}^{K} | y_i - \hat{y}_i | , \tag{2}$$

where y_i is the ground truth object, $\hat{y}_i$ is the predicted object, K is the total number of samples, and $| y_i - \hat{y}_i |$ represents the absolute error between each predicted object ($i = 1, 2, 3, ..., K$) and the ground truth object. Then, the Intersection over Union (*IoU*) index is used to measure the overlap between the predicted and ground truth BBs, ranging from 0 to 1, with a higher IoU (IoU $\geq$ 0.5) indicating greater prediction confidence. However, the absolute error function is discontinuity at value zero; thus, during optimization, it may lead to gradient discontinuity, insensitivity to small errors, and low convergence speed.

In this study, the Generalized Intersection over Union (*GIoU*) addresses the shortcomings of IoU by considering the relationship between the BB and its enclosing convex hull, thereby improving learning stability. The general formula of *GIoU* is expressed as[34]

$$GIoU = IoU - \frac{|C - (A \cup B)|}{|C|},\ IoU = \frac{|A \cap B)}{|A \cup B|}, \tag{3}$$

where Regions #*A* and #*B* are the predicted and ground truth BBs, respectively, Region #*C* is the smallest enclosing convex hull that contains both Regions #*A* and #*B*, and $|C - (A \cup B)|$ is the region of *C* excluding $(A \cup B)$, when its value is larger, indicating a greater distance between the predicted and ground truth BBs, as a smaller value means closer alignment. In this study, we also applied the Distribution Focal Loss (*DF_Loss*) function to improve the accuracy of BB predictions, and the general formula of GIoU is expressed as[21,35]

$$DF\_Loss = -\sum_{i=1}^{N} (P_i \log(P_i^*)), \tag{4}$$

where P_i is the ground truth probability distribution, P_i^* is the predicted probability distribution, *N* is the total number of samples, and *DF_Loss* is the loss value. The *DF_Loss* can predict possible BB locations, allowing the classifier model to have a guided target for regressing to correct coordinates, thereby improving the regression accuracy for object detection tasks.

2.2.2 Segmentation and classification loss function

In this study, the Dice loss (*Dice_Loss*) function is applied for image segmentation tasks. The general formula is expressed as[21]

$$Dice\_Loss = 1 - \frac{2|A \cap B|}{|A| + |B|}, \tag{5}$$

where *A* is the predicted segmentation region, *B* is the ground truth region, $|A \cap B|$ represents the intersection between regions *A* and *B*, and $|A|$ and $|B|$ are the numbers of pixels in regions *A* and *B*, respectively. The *Dice_Loss* is a metric for evaluating the image segmentation performance by estimating the overlap ratio between regions *A* and *B*. In addition, the Focal loss (Focal_Loss) function is applied for evaluating the classification performance, and the general formula is expressed as[36]

$$Focal\_Loss = -a_t(1 - p_t)^{\gamma} \log(p_t), \tag{6}$$

where p_t is the predicted probability of the ground-truth class, γ is the modulation factor, and α_t is the balancing factor for different classes, which control the balance between hard-to-classify and easy-to-classify samples during the training stage.

In the training stage, the Stochastic Gradient Descent (SGD) algorithm[21,22] is applied to iteratively update and optimize the YOLOv11 network parameters. The general formula is expressed as

$$w_{t+1} = w_t - \eta \nabla_{Bw} L(w)\,,\ \nabla_{Bw} L(w) = \frac{1}{K_B}\sum_{k=1}^{K_B} \nabla_w L(f(x_k; w), y_k)\,, \tag{7}$$

where x_k is the k-th training dataset, y_k is the k-th ground-truth label, $k = 1, 2, 3, \ldots, K_B$, K_B is the number of mini-batch training datasets, (x_k, y_k) represents the k-th paired training dataset, w_t and w_{t+1} are the network parameters at t-th and $(t + 1)$-th iterations, respectively, η is the learning rate, $L(\bullet)$ is the loss function (segmentation or classification loss function), $f(\bullet)$ is the network output, and $\nabla_{Bw} L(\bullet)$ is the average mini-batch loss. The SGD algorithm is well suited for DL applications with extremely large-scale training datasets. In each iteration computation, it updates the parameters using only a mini-batch of training datasets, which not only significantly improves computational efficiency but also introduces stochasticity into the parameter update process. This stochastic perturbation helps reduce the likelihood of being trapped in local minima, thereby increasing the chance to approach the global optimum. Thus, SGD supports online learning by continuously updating the classifier model with new training datasets, making it well suited for streaming data and online learning tasks.

The proposed classifier, as "2*D fractional-order Conv. operation* + *YOLOv11* + *GIoU_Loss* + *DF_Loss* + *Dice_Loss* + *Focal_Loss* + *SGD*", is capable of performing automatic image enhancement, ROI extraction, image segmentation, and pattern recognition (classification) functions. In the segmentation evaluation, the metrics of mean average precision (*mAP*), including *mAP*_50 and *mAP*_50–95, were used to provide a general measurement for evaluating OD performances. The *mAP* index was a single *IoU* with a threshold of 0.50 to evaluate the overlap between the predicted BB and the ground truth BB. The *mAP_50–95* index with threshold values from 0.50 to 0.95 (i.e., 0.50, 0.55, 0.60, 0.65, …, 0.95)[21,22] was the standard evaluation metric for both localization accuracy and classification performance, and also offered the detailed view of the classifier's overall OD quality. In the classification evaluation for stenosis screening, a tenfold cross-validation method is applied to assess the classifier's performances. On the basis of the confusion matrix generated from the classifier outputs, the numbers of actual and predicted classes are used to estimate four evaluation indexes for classification tasks, namely, true positive (TP), false positive (FP), true negative (TN), and false negative (FN). These indexes are further utilized to estimate the classifier's performance metrics, including Precision (%), Recall (%), F1 score, and Accuracy (%) for evaluating the effectiveness of image classification.[37,38]

2.3. Quantitative indexes for hemodialysis fistula screening

After segmentation and classification processes from the ultrasound imaging data stream, the YOLOv11-based classifier is employed to identify the vascular wall contour and vascular lumen (nonstenostic area), and thrombus (stenostic area), as seen in the green and red regions in Fig. 4, respectively. Through the segmentation results, these distinct regions can be used to compute the quantitative indexes, namely, *DOS* and *LCR*;[8–10,39] the general formulas are defined as

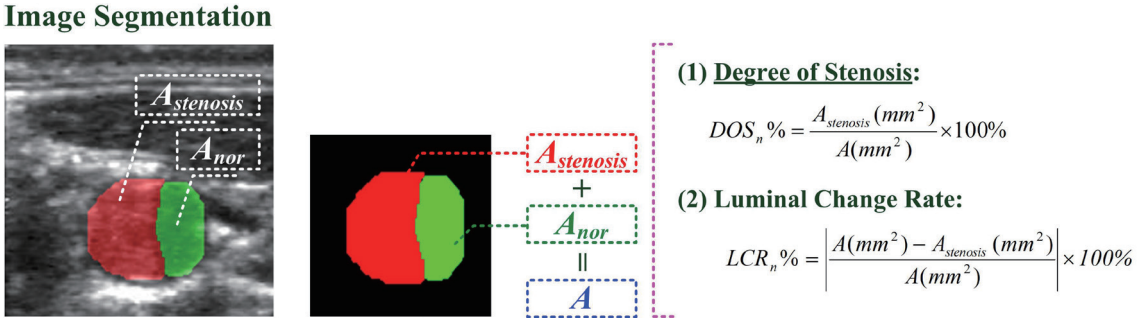

Fig. 4. (Color online) Automatic ultrasound image segmentation and quantification indexes for stenosis severity evaluation.

$$Vas\_A(mm^2) = \frac{Pixels\ of\ Vascular\ Area}{Pixels\ per\ mm^2},\ DOS_n\% = \frac{Vas\_A_{stenosis}(mm^2)}{Vas\_A(mm^2)} \times 100\%, \tag{8}$$

$$LCR_n\% = \left| \frac{Vas\_A(mm^2) - Vas\_A_{stenosis}(mm^2)}{Vas\_A(mm^2)} \right| \times 100\%, \tag{9}$$

where Vas_A(mm²) is the area of the normal vascular lumen, $Vas\_A_{stenosis}$(mm²) is the area of the stenostic region, n = 1, 2, 3, …, N, and N denotes the total number of cross-sectional image frames acquired along the Z-axis direction. On the basis of the identified nonstenotic and stenotic regions (green and red regions, respectively), the number of enclosed pixels is estimated. The actual vascular lumen area is subsequently obtained by converting the pixel count using the calibration factor (1 pixel = 0.2645833 mm²). The *DOS*% is then computed, where a higher *DOS* value indicates a more severe level of vascular stenosis. Therefore, the severity of vascular stenosis can be divided into three classes, namely, Class# I: *DOS*% ≤ 30%, Class# II: 30% < *DOS*% ≤ 50%, and Class# III: *DOS*% > 50%[8–10]. For the severity level reaching Class# II, medical treatment or interventional therapy is suggested, and for the severity level reaching Class# III, balloon angioplasty is performed to dilate the lesion area and restore blood flow within the stenotic section. The *LCR* index is inversely proportional to the *DOS*, indicating that the normal vascular lumen area decreases as the stenosis severity increases. These two indexes provide valuable information for HD patients and clinical physicians to understand the internal conditions within the fistula access, enabling problems to be found at the early stage.

3. Experimental Results and Discussion

In this study, a total of 18 HD subjects were recruited for B-mode ultrasound imaging collection in a hospital experimental field (Image source: National Cheng Kung University Hospital, Tainan City, Taiwan). All patients' images were anonymized to protect patient privacy. Among them, 6 HD subjects without AVF stenosis (control group: *DOS*% ≤ 30%) and 12 subjects with varying *DOS*s (experimental group: 30% < *DOS*% ≤ 50% and *DOS*% > 50%) were

included. The different *DOS* classes were initially selected by experienced clinical physicians by manually identifying the nonstenotic and stenotic regions within each ultrasound imaging frame. The *DOS* levels were preliminarily estimated by using the clinical general formula as Eq. (1), based on the vascular diameter measurements (d and D) as seen in Fig. 1(b). After the manual annotation (biomarking) processes, the collected images were divided into the aforementioned three levels. A total of 650 ultrasound images were subsequently divided into training, validation, and testing datasets for YOLOv11-based classifier establishment and performance evaluation, as seen in Table 1. The performance evaluation of the YOLOv11-based classifier was conducted by using tenfold cross-validation tests.

As shown in Table 2, the performance comparisons of different classifier models for pattern recognition were conducted, including Model#1: "2D Fractional-Order Conv. Operation + YOLOv11 + SGD", Model#2: "2D Fractional-Order Conv. Operation + YOLOv11 + ADAM (adaptive moment estimation)", Model#3: "YOLOv11 + SGD", and Model#4: "YOLOv11 + ADAM". These classifiers were established and evaluated by using identical training and testing datasets to ensure fair performance comparisons. In the 2D spatial domain, the fractional-order Conv. operation with experienced fractional-order parameters, $v = 0.1$ to 0.3,[30–33] was capable of suppressing noises while enhancing edge information in clinical medical images, such as chest X-ray and CT images.[30,33] Therefore, the vessel contour, vascular lumen, and thrombi within the fistula access can be effectively highlighted, thereby facilitating the extraction of critical features and enabling ROI block segmentation within the specific regions.

To speed up the computational performances in image processing, segmentation, and classification operations, a graphics processing unit (GPU) platform (NVIDIA® GeForce® RTX™ series) was utilized in the training, validation, and testing stages. The GPU-based implementation significantly reduced the computational time required for image enhancement, feature extraction, image segmentation (nonstenotic and stenotic regions), and pattern

Table 1
Datasets for YOLOv11-based classifier establishment.

Dataset	Normal (nonstenotic)	Abnormal (stenotic)
Total training dataset	300	350
Total testing dataset	130	80
Training dataset for each fold	180	210
Validation dataset for each fold	120	140
Testing dataset for each fold	130	80

Table 2
Experimental results of tenfold cross validation for different classifier models.

Model	Method	Average precision (%)	Average recall (%)	Average F1 score	Average accuracy (%)
#1	2D Fractional-Order Conv. Operation+ YOLOv11 + SGD	100.00 ± 0.00	90.00 ± 0.01	0.9474 ± 0.0032	94.74 ± 0.02
#2	2D Fractional-Order Conv. Operation+ YOLOv11 + ADAM	99.72 ± 0.01	94.30 ± 0.02	0.9713 ± 0.0032	96.87 ± 0.01
3	YOLOv11 + SGD	100.00 ± 0.00	95.00 ± 0.01	0.9737 ± 0.0010	97.37 ± 0.02
4	YOLOv11 + ADAM	84.56 ± 0.32	76.84 ± 0.38	0.7958 ± 0.0026	87.66 ± 0.12

recognition (normal and abnormal classes) tasks. In training and validation stages, SGD- and ADAM-based optimizers[23,24] were used to iteratively fine-tune and update the YOLOv11 classifier's optimal network parameters, aiming to adjust the segmentation loss to be greater than 0.50 and approach the value of 1.00, and also to minimize the loss function and approach the value of 0.00 for the classification process in the training stage, thereby improving the accuracy of nonstenotic and stenotic region segmentation and classification. For example, for Model#1- and Model#2-based classifiers with SGD (orange solid line) and ADAM (blue solid line) training algorithms, Figs. 5(a) and 5(b) show the training and validation history curves for US image segmentation and classification tasks, respectively. As the number of iterations increased, the classifier's output loss values, including *box_loss*, *seg_loss* (Segmentation_Loss), and *dfl_loss* (Distribution Focal Loss), gradually decreased, with each loss value consistently remaining above 0.50; conversely, the cls_loss (Classification_Loss) gradually decreased to the specific convergence condition, which reached the saturation state after 200 training iterations. As seen in Fig. 5(c), evaluation metrics, including Precision (%), Recall (%), *mAP_50*, and *mAP_50–95*, showed promising performances and quickly converged to the saturation state. Th *mAP_50–95* index reached a value above 70%, promising localization accuracy for correct OD.

With the tenfold cross validation, we could see the promising results for both image segmentation and classification with Model#1 (with 2D Fractional-Order Conv. Operation) and Model#3 (without 2D Fractional-Order Conv. Operation). For example, with Model#1, the average F1 score (0.9474 ± 0.0032) was greater than 0.9000, indicating few FP and FN, and demonstrating good reliability in separating the stenotic (abnormal) regions from the nonstenotic (normal) regions. The average accuracy of 94.74 ± 0.02% was achieved in identifying both normal and abnormal cases. Overall experimental results for four classifier models are shown in Table 2. For example, in three cases, the proposed YOLOv11-based classifier can identify the stenotic and nonstenotic regions from US imaging data stream, which are also marked in red and

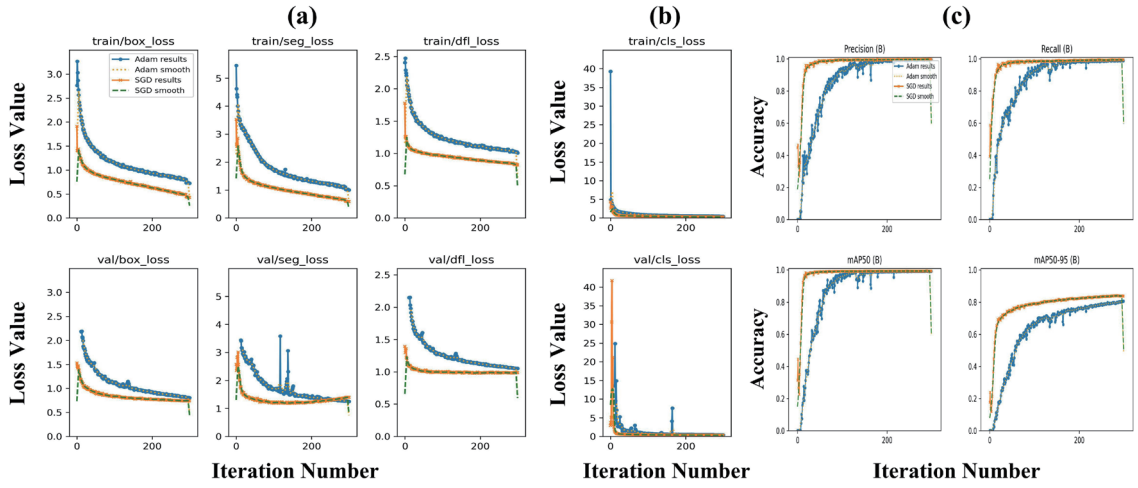

Fig. 5. (Color online) Training and validation history curves. (a) Training and validation history curves of box_loss, seg_loss, and dfl_loss with the SGD and ADAM algorithms. (b) Training and validation history curves of box_loss, seg_loss, and dfl_loss with the SGD and ADAM algorithms. (c) Image segmentation and classification evaluations with the Precision (%), Recall (%), *mAP*_50, and *mAP*_50–95 indexes.

green, respectively. As seen in the detection results of Case#1, *DOS*% = 65.97% and *LCR*% = 34.03%, which can be classified as Class# III level from the individual frame US image in Fig. 6(a). By analyzing the spatial distribution and pixel area of these regions (red and green regions), key quantitative indexes, such as the *DOS*(%) and *LCR*(%) indexes, can be estimated for evaluating the stenotic level as Class III. This frame-level analysis was crucial for identifying localized abnormalities (red regions) and served as a foundation for subsequent sequence-based analysis and 3D AVF reconstruction. For 30-frame US images, as seen in the estimation results in Fig. 7, the stenotic and nonstenotic regions were successfully determined, and then two quantitative indicators, *DOS*(%) and *LCR*(%), were estimated. According to Fig. 7(a), the *DOS*(%) values from Frames #1 to #30 can be used to identify stenotic levels (Classes I–III). For example, for Subject#2, all exceeded 50% (average *DOS*(%) = 58.83%), indicating that the entire examined segment fell under level Class III (*DOS*(%) > 50.0%), as shown in Fig. 7(a), and Fig. 7(b) presents the trend of *LCR*(%) values across 30 US image frames, with an average *LCR*(%) of 34.50% throughout the scanned region.

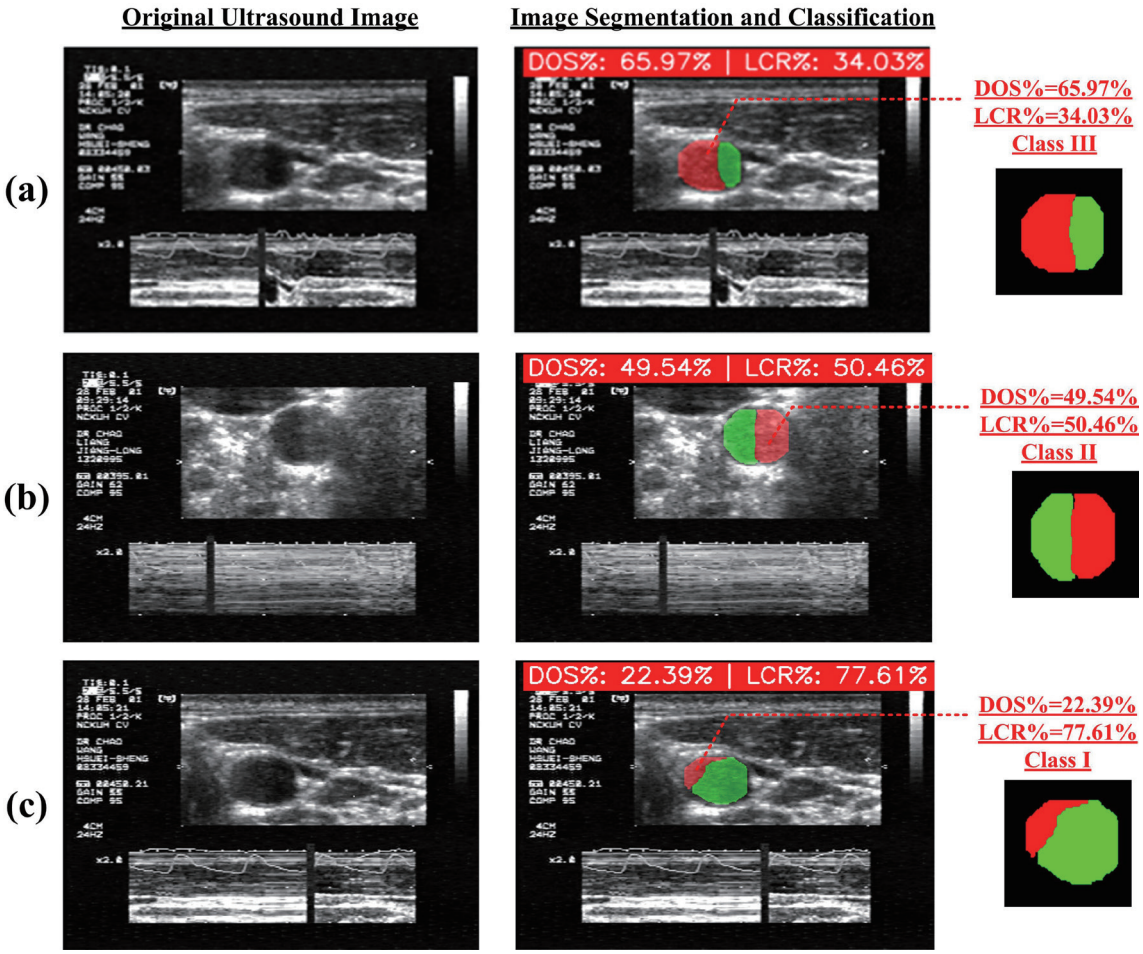

Fig. 6. (Color online) Individual frame US image detection results, indicating (a) *DOS*% = 65.97% and *LCR*% = 34.03% for Class III, (b) *DOS*% = 49.54% and *LCR*% = 50.46% for Class II, and (c) *DOS*% = 22.39% and *LCR*% = 77.61% for Class I.

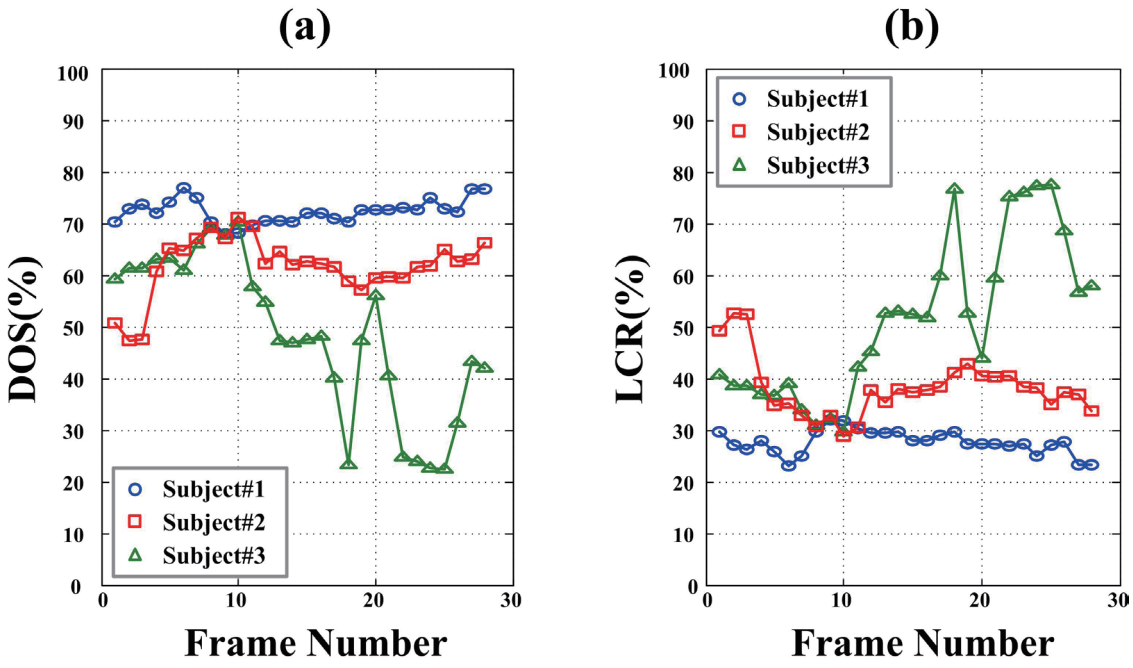

Fig. 7. (Color online) Estimated results of DOS(%) and LCR(%) indicators from 30 US image frames.

On the basis of the experimental results from the automatic US image sequence detection, the feasibility of the proposed YOLOv11-based classifier can be validated. Within the ROI in each frame, the image segmentation successfully identified the stenotic and nonstenotic regions, as illustrated in Fig. 8. With this AVF anatomical information, the *DOS* (%) and *LCR* (%) can be estimated. Therefore, the YOLOv11-based classifier can clearly segment the nonstenotic and stenotic areas and visualize these two regions. By analyzing the size and variation of the red and green regions, the obstructed sections and the narrowed cross-sectional areas within the AVF access can be identified. These results can be applied to evaluate the dialysis functionality of the AVF access. In addition, the spatial information combined with the segmented nonstenotic and stenotic areas derived from the image sequence detection results can be employed for 3D fistula access reconstruction, as seen in Fig. 8. Hence, the spatial information obtained from the imaging data stream was utilized for reconstructing the 3D AVF model. In the 3D AVF model, the red color indicated the stenotic region, whereas the green color represented the nonstenotic region. Hence, this 3D model enabled colored visualization processes for the presented potential lesion segments and their respective locations. In this study, the 3D Slicer image computing software (free and open source software) was used to integrate and interpolate the image sequence data to reconstruct a 3D model. The Blender software (free and open source 3D computer graphics tool) was used for dynamic visualization, enabling the 360° viewing of the AVF model or focusing on specific suspected lesion regions. This provided a realistic and real-time 3D visualization manner to assist clinical diagnosis for both patients and physicians and to assess the AVF conditions. Thus, the colored visualization model enabled the generation of a clinically meaningful 3D structure without the need for endoscopy and offered a noninvasive and intuitive tool for evaluating the AVF internal conditions.

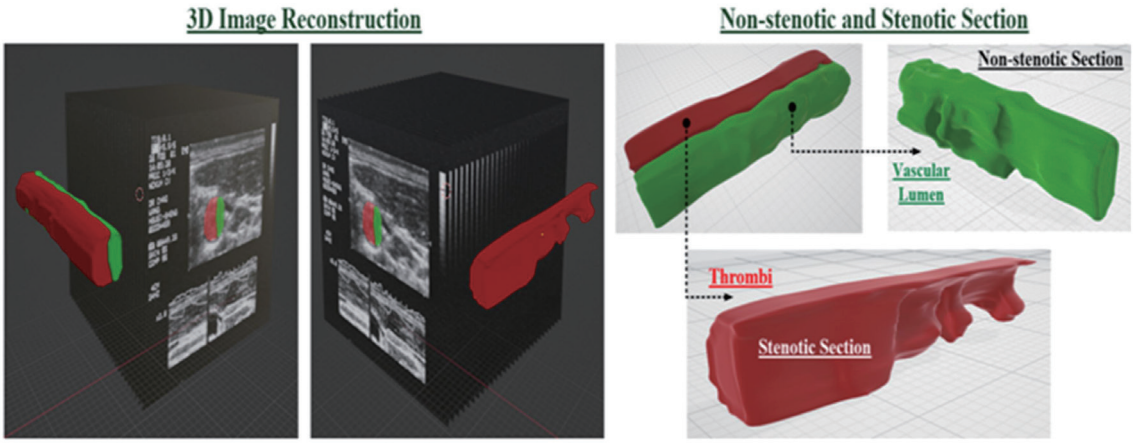

Fig. 8. (Color online) 3D image reconstruction and visualization of dialysis 3D AVF model for indicating nonstenotic and stenotic regions.

In addition, the 3D reconstructed model can also be used for *in vitro* experiments for hemodynamic simulations,[8,40] allowing further analysis of blood flow dynamics and vascular properties under various clinical conditions. By analyzing the 3D AVF structure, the internal flow field, blood flow volume, and blood flow velocity within the AVF access can be simulated at the stenotic region above 50% by using computational fluid dynamics (CFD) numerical analysis. The CFD simulation software can be applied to solve the governing equations of fluid mechanics at the lesion site using numerical analysis methods. By considering the blood viscosity variations and vascular wall elasticity, the blood flow field, flow rate, and velocity within the AVF access can be estimated. Furthermore, the shear stress exerted on the vascular walls, as well as physical characteristics such as wall elasticity and compliance, can also be analyzed. The proposed noninvasive automated image detection combined with hemodynamic simulation analysis contributes significantly to the enhancement of the healthcare of dialysis AVFs.

4. Conclusion

In this study, we developed an automated US image sequence detection tool by constructing YOLOv11-based classifiers for dialysis AVF stenosis screening. A total of 650 US images were collected from 18 subjects and used for the proposed classifier's training and validation. The "2D Fractional-Order Conv. Operation + YOLOv11 + SGD" and "YOLOv11 + SGD" was established, and an average precision of 100.00%, average recall of >90.00%, and average accuracy of >90.00% were obtained for identifying the stenotic (abnormal) and nonstenotic (normal) regions. Moreover, both classifiers had an average F1 score of > 0.9000, indicating that the models performed very well in terms of both precision and recall (harmonic mean of precision and recall). The proposed classifier can accurately identify both stenotic and nonstenotic regions with minimal errors, demonstrating high reliability in image segmentation and classification tasks. The experimental results also showed that the proposed YOLOv11-

based classifier models were clinically feasible for screening AVF access stenosis and can also serve as effective supportive tools in the routine healthcare for HD patients. In addition, the YOLOv11-based classifier's generalization is ensured by training on diverse image datasets with varied patients, scanning angles, and image qualities, and by applying extensive US image augmentation to improve the robustness and adaptability to unseen clinical data (untrained datasets).

Acknowledgments

This work was supported by the National Science and Technology Council (NSTC) under contract numbers NSTC 113-2813-C-167-013-E and NSTC 114-2813-C-167-008-E (duration: July 1, 2024 –February 28, 2026).

References

1 S.-M. Ou, C.-T. Chao, M.-Y. Lin, C.-C. Hsu, M.-H. Lin, C.-L. Chou, Y.-H. Hsu, S.-J. Hwang, M-S. Wu, and M-Y. Wu: Acta Nephrologica **38** (2024) 201. https://doi.org/10.6221/AN.202412_38(4).0001
2 M.-Y. Tsai, C.-Y. Chin, W.-C. Lee, Y.-C. Huang, and Y.-C. Cheng: BMC Health Serv. Res. **24** (2024) 1. https://doi.org/10.1186/s12913-024-11912-1
3 M. Moini, M. R. Rasouli, G. M. Williams, S. Najafizadeh, and G. Sheykholeslami: European Soc. Vasc. Surg. **17** (2009) 7. https://doi.org/10.1016/j.ejvsextra.2008.11.001
4 R. Stanziale, M. Lodi, E. D'Andrea, F. Sammartino, and V. Dl Luzio: Hemodial. Int. **15** (2011) 100. https://doi.org/10.1111/j.1542-4758.2010.00503.x
5 M. Moini, M. R. Rasouli, G. M. Williams, S. Najafizadeh, and G. Sheykholeslami: European J. Vascular and Endovascular Surg. **37** (2009) 495. https://doi.org/10.1016/j.ejvs.2008.11.016
6 D. Tian, W. Chen, D. Xu, L. Xu, G. Xu, Y Guo, and Y. Yao: Comput. Biol. Med. **170** (2024) 1. https://doi.org/10.1016/j.compbiomed.2024.108074
7 H. Takeda, F. Kato, K. Kamishima, and H. Iwata: Measurement: Sensors **38** (2025) 1. https://doi.org/10.1016/j.measen.2025.101846
8 W.-L. Chen, C.-D. Kan, C.-H. Lin, and Y.-C. Mai: IEEE Access **6** (2018) 10381. https://doi.org/10.1109/ACCESS.2018.2802479
9 C.-H. Lin, J.-X. Wu, C.-D. Kan, P.-Y. Chen, and W.-L. Chen: Biomed. Signal Process. Control **66** (2021) 1. https://doi.org/10.1016/j.bspc.2021.102461
10 J.-Xing W., P.-Y. Chen, H.-C. Liu, C.-H. Lin, S. Chen, and K. K. Shung: IEEE Access **7** (2019) 171765. https://doi.org/10.1109/ACCESS.2019.2955742
11 X. Hu, B. Li, J. Mao, X. Hu, J. Zhang, H. Guo, D. Wang, Y. Zhang, J. He, N. Zhao, H. Zhang, and P. Pang: Annals Vasc. Surg. **84** (2022) 286. https://doi.org/10.1016/j.avsg.2022.01.023
12 S. Chang and A. K. Bowden: J. Biomed. Optics. **24** (2019) 090901-1. https://doi.org/10.1117/1.JBO.24.9.090901
13 Y.-C. Li, Thau-Yun Shen, Chien-Cheng Chen, Wei-Ting Chang, Po-Yang Lee, and Chih-Chung Huang: IEEE Trans. Ultrason. Ferroelectr. Freq. Control **68** (2021) 1762. https://doi.org/10.1109/TUFFC.2021.3052486
14 W. Weng and X. Zhu: IEEE Access **9** (2021) 16591. https://doi.org/10.1109/ACCESS.2021.3053408
15 Z. Lin, P.-H. Tsui, Y. Zeng, G. Bin, S. Wu, and Z. Zhou: 2022 IEEE Int. Ultrasonics Symp. (IUS) Venice, Italy (2022). https://doi.org/10.1109/IUS54386.2022.9958784
16 B. Rosengarten, C. Aldinger, A. Kaufmann, and M. Kaps: Ultrasound Med. Biol. **27** (2002) 1499. https://doi.org/10.1016/s0301-5629(01)00464-1
17 S.-Y. Chuang, C.-H. Bai, H.-M. Cheng, J.-R. Chen, W.-T. Yeh, P.-F. Hsu, W.- L. Liu, and W.-H. Pan: Eur. J. Prev. Cardiol. **23** (2016) 116. https://doi.org/10.1177/2047487315571888
18 R. Hosney, F. M. Talaat, E. M. El-Gendy, and M. M. Saafan: Neural Comp. Appl. **36** (2024) 17199. https://doi.org/10.1007/s00521-024-09966-7
19 J. Joniwan, G. Firmasyah, A. M. Widodo, and B. Tjahjono: J. Social Res **2** (2023) 3291. https://doi.org/10.55324/josr.v2i9.1403
20 J. Zou and H. Wang: IEEE Access **12** (2024) 124160. https://doi.org/10.1109/ACCESS.2024.3453931

21 L. He, Y. Zhou, L. Liu, and J. Ma: Buildings **14** (2024) 1. https://doi.org/10.3390/buildings14123777
22 R. Khanam and M Hussain: arXiv (2024) 1. https://doi.org/10.48550/arXiv.2410.17725
22 J. E. Garza and M. F. Islam: IEEE Access **13** (2025) 138813. https://doi.org/10.1109/ACCESS.2025.3596039
23 Y. Tian, Q. Ye, and D. Doermann: arXiv (2025). https://doi.org/10.48550/arXiv.2502.12524
24 C.-T. Tsai, C.-H. Lin, H.-Y. Lai, Y.-E. Cheng, P.-T. Huang, N.-S. Pai, and C. M. Li: IEEE Access **13** (2025) 80559. https://doi.org/10.1109/ACCESS.2025.3566108
25 D. Yang, H. Jiang, S. Ai, T. Yang, Z. Zhi, D. Jing, J. Gao, K. Yue, H. Cheng, and Y. Xu: Eng. Appl. Artif. Intell. **138** (2024) 1. https://doi.org/10.1016/j.engappai.2024.109442
26 S. Wang, Y. Wang, Y. Peng, and X. Chen: Compu. Electr. Eng. **120** (2024) 1. https://doi.org/10.1016/j.compeleceng.2024.109654
27 Q. Song, Y. Liu, H. Sun, Y. Chen, and Z. Zhou: IEEE Access **12** (2024) 104503. https://doi.org/10.1109/ACCESS.2024.3435386
28 M. Hussain: Machines **11** (2023) 1. https://doi.org/10.3390/machines11070677
29 M. Hussain: IEEE Access **12** (2024) 42816. https://doi.org/10.1109/ACCESS.2024.3378568
30 C.-H. Lin, C.-D. Kan, W.-L. Chen, and P.-T. Huang: J. X-Ray Sci. Technol. **27** (2019) 517. https://doi.org/10.3233/XST-180473
31 Y.-F. Pu, J.-L. Zhou, and X. Yuan: IEEE Trans. Image Process. **19** (2010) 491. https://doi.org/10.1109/TIP.2009.2035980
32 J.-L. Chen, C.-H. Huaang, Y.-C. Du, and C.-H. Lin: IET Image Process. **8** (2014) 354. https://doi.org/10.1049/iet-ipr.2012.0660
33 P.-Y. Chen, J.-X. Wu, C.-H. Lin, J.-C. Hsu, and N.-S. Pai: IEEJ Trans. Electr. Electron. Eng. **17** (2022) 132. https://doi.org/10.1002/tee.23495
34 C. Tong, X. Yang, Q. Huang, and F. Qian: Appl. Sci. **12** (2022) 1. https://doi.org/10.3390/app122412785
35 J. Zhu, H. Tang, and L. Chai: 2024 36th Chinese Control and Decision Conf. (CCDC), Xi'an, China, (2024).
36 Z. Liu, Y. Gao, and Q. Du: IEEE Access **11** (2023) 109179. https://doi.org/10.1109/ACCESS.2023.3321828.
37 I. R. I. Haque and J. Neuber: Inf. Med. Unlocked **18** (2020) 1. https://doi.org/10.1016/j.imu.2020.100297
38 M. Madhikermi, S. Kubler, J. Robert, A. Buda, and K. Främling: Expert Syst. Appl. **63** (2016) 145. https://doi.org/10.1016/j.eswa.2016.06.043.
39 C.-H. Lin, C. C. Pai, C.-D. Kan, J.-X. Wu, P.-Y. Chen, and N.-S. Pai: IEEJ Trans. Electr. Electron. Eng. **18** (2023) 254. https://doi.org/10.1002/tee.23726
40 C.-H Lin, C.-D. Kan, W.-L. Chen, M.-J. Wu, and F.-M. Yu: Technol. Health Care **24** (2016) 295. https://doi.org/10.3233/THC-161130.