

Deep Learning Models for Brachial Lumen Segmentation

Modelos de Aprendizado Profundo para Segmentação do Lúmen Braquial

Breno Luiz Nunes da Silva^{1*}, Alexandre Gonçalves Silva², Lucas Ziza Nastari³, Rangel Arthur³

Abstract: Cardiovascular diseases remain one of the leading causes of global mortality. Flow-mediated dilation of the brachial artery, assessed by ultrasonography, is a non-invasive method to evaluate endothelial function; however, manual analysis is time-consuming and prone to variability. This study proposes the automatic segmentation of the brachial artery lumen using computer vision techniques based on deep learning. We investigated three segmentation architectures: UNet++, MultiResUNet, and TransUNet, applied to ultrasound images enhanced with contrast adjustment through CLAHE. Model training was conducted using K-fold cross-validation, with performance assessed through established metrics such as Dice, IoU, and Accuracy, while employing the Focal Tversky Loss function, suitable for imbalanced scenarios. Among the evaluated approaches, UNet++ achieved the best results, with Dice of 0.950, IoU of 0.905, and Accuracy of 0.980. These findings demonstrate the potential of deep segmentation models to reduce the subjectivity of clinical analysis and improve diagnostic accuracy in ultrasound-based vascular assessments.

Keywords: lumen segmentation — UNet++ — MultiResUNet — TransUNet — deep learning

Resumo: As doenças cardiovasculares permanecem entre as principais causas de mortalidade no mundo. A dilatação mediada por fluxo, obtida pela ultrassonografia da artéria braquial, é um método não invasivo para estimar a função endotelial; contudo, a análise manual é demorada e suscetível à variabilidade. Este estudo propõe a segmentação automática do lúmen braquial por meio de técnicas de visão computacional baseadas em aprendizado profundo. Foram avaliadas três arquiteturas de segmentação: UNet++, MultiResUNet e TransUNet, aplicadas a imagens de ultrassom com realce de contraste obtido por CLAHE. O treinamento foi conduzido com validação cruzada do tipo K-Fold, utilizando métricas consolidadas como Dice, IoU e Acurácia, além da função de perda Focal Tversky Loss, apropriada para cenários desbalanceados. Entre as arquiteturas, a UNet++ apresentou o melhor desempenho, com Dice de 0,950, IoU de 0,905 e Acurácia de 0,980. Os resultados demonstram o potencial de modelos de segmentação profunda para reduzir a subjetividade da análise clínica e aprimorar a precisão diagnóstica em exames ultrassonográficos.

Palavras-Chave: segmentação do lúmen — UNet++ — MultiResUNet — TransUNet — aprendizado profundo

¹ Departamento de Automação e Sistemas, Universidade Federal de Santa Catarina (UFSC), Brazil

² Departamento de Informática e Estatística, Universidade Federal de Santa Catarina (UFSC), Brazil

³ Escola de Tecnologia, Universidade Estadual de Campinas (Unicamp), Brazil

*Corresponding author: brenluiz13@gmail.com

DOI: <http://dx.doi.org/10.22456/2175-2745.150923> • Received: 14/10/2025 • Accepted: 03/11/2025

CC BY-NC-ND 4.0 - This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

1. Introduction

Cardiovascular diseases (CVD) remain the leading cause of global mortality, having been responsible for 17.9 million deaths in 2019 (32% of the total) [1]. Flow-mediated dilation (FMD) of the brachial artery is widely used as a non-invasive measure of endothelial function. However, manual lumen analysis is time-consuming and susceptible to both inter- and intra-observer variability, especially in low-contrast images, as occurs in B-mode ultrasound (also known as Brightness Mode). This mode is the most common form of acquisition in ultrasound examinations, where the intensity of the reflected

echo is represented in two-dimensional grayscale levels, resulting in grayscale ultrasound images of the tissue.

In this context, automatic brachial lumen segmentation emerges as a promising alternative to standardize FMD measurements, reduce subjectivity, and increase clinical reliability. The present study conducts a systematic comparison of three modern segmentation architectures applied to FMD examination in ultrasound.

The text is organized as follows: Section 2 reviews previous relevant works on the topic, situating this study in the state of the art. Section 3 describes in detail the clinical scenario

and the adopted methodology, covering the definition of FMD (Section 3.1), data preparation (Section 3.2), image registration (Section 3.3), the compared architectures (Section 3.4), the loss function used (Section 3.5), the training procedures (Section 3.6) and the thresholding and post-processing steps (Section 3.7). Section 4 gathers the quantitative (Section 4.1) and qualitative (Section 4.2) results, which are critically discussed in Section 4.3. Finally, Section 5 synthesizes the main results and presents perspectives for future work.

2. Related Works

Image segmentation has been widely studied over the last decades, with approaches ranging from classical methods to modern deep learning techniques. Among traditional methods, edge and region-based algorithms stand out, such as the Canny detector, active contours (snakes) [2], level set models, watershed segmentation, and graph-based techniques, including graph cut. These methods are grounded in solid mathematical principles and provided the foundation for the development of biomedical segmentation, although they present limitations in images with low contrast or high noise levels.

In the domain of vascular ultrasound, segmentation evolved from these classical methods to specific image processing strategies. Probabilistic models, such as particle filters applied in temporal sequences [3], have also been explored. However, both active contours and probabilistic models face significant difficulties in the presence of speckle noise and low contrast, characteristic of B-mode ultrasound.

In this context, preprocessing plays a central role in contrast improvement. The use of Contrast Limited Adaptive Histogram Equalization (CLAHE) has proven particularly efficient for enhancing edges and attenuating speckle, favoring more robust segmentation [4].

With the advent of deep learning, architectures like U-Net transformed biomedical analysis, by enabling multi-scale representations and direct learning of discriminative features from data [5]. Recent reviews highlight the impact of these architectures in ultrasound and biomedical applications [6, 7], consolidating U-Net and its variants as central references.

Applied studies in vascular ultrasound confirm this trend. U-Net++ was employed in carotid artery segmentation, achieving improvements in atherosclerotic plaque detection [8]. In FMD examinations, post-processing strategies proved effective in reducing false positives [9]. Comparative evaluations of different CNN architectures also highlighted the challenges of generalization between heterogeneous clinical databases [10]. Other recent proposals include multi-scale models with spatial attention applied to carotid segmentation [11] and hybrid CNN–Transformer architectures that showed consistent gains in peripheral artery scenarios [12].

Even more recent works expand this scope. For example, promising results in functional ultrasound images were reported in investigations that proposed new vascular segmentation tools [7], while hybrid techniques in 3D reinforced the trend of adaptive architectures to handle the high struc-

tural variability of ultrasound [13]. These examples show that vascular ultrasound segmentation follows a trajectory of increasing sophistication, with hybrid and attention-based approaches emerging as robust alternatives.

Beyond the U-Net family, multi-resolution architectures and attention mechanisms, including Transformers [14], have been expanding the repertoire of available techniques. Hybrid models prove particularly effective for capturing long-range dependencies and handling the structural variability typical of ultrasound images.

Despite these advances, studies that conduct standardized comparisons of architectures in flow-mediated dilation (FMD) examinations in B-mode, under a reproducible and transparent pipeline, are still scarce. The present work seeks to fill this gap by proposing a systematic evaluation of three representative architectures (UNet++, MultiResUNet and TransUNet), conducted under a unified protocol of preprocessing, training, and validation, ensuring reproducibility and fairness in the comparison between methods.

3. Methodology

We anchor the methodology in a robust, comparable between models and reproducible pipeline. We start with the clinical scenario and acquisition parameters of the FMD examination in B-mode ultrasound. Then, we describe the database formation and preprocessing. The reference masks were produced by two evaluators with consensus for disagreements. The comparison between UNet++, MultiResUNet and TransUNet was conducted under a single training protocol to ensure experimental fairness.

3.1 FMD Definition

For the determination of FMD, the lumen diameter of the artery is necessary, which, in turn, must be estimated from its segmentation in the image. FMD is calculated by:

$$\text{FMD}(\%) = 100 \cdot \frac{D_{\text{peak}} - D_{\text{basal}}}{D_{\text{basal}}} \quad (1)$$

where D_{basal} corresponds to the mean lumen diameter during the basal period (rest), and D_{peak} is the largest diameter observed in the post-hyperemia period.

Segmentation errors at the edges directly impact the FMD calculation. Underestimation of the basal or peak diameter can introduce systematic bias, affecting the clinical interpretation of endothelial function. Thus, the evaluation of architectures should consider not only segmentation metrics but also the potential effect on the derived physiological measure.

3.2 Data and Preparation

We worked with 935 image–mask pairs (1,870 files), organized in two folders (each with `img/` and `mask/`). The masks come from segmentations done manually or semi-assisted. To contextualize this evaluation process, Figure 1 presents two ultrasound cases accompanied by their respective reference masks. These masks, serving as the gold standard in

this study, were produced by two physicians from UNICAMP with extensive experience in vascular ultrasound examinations. The term gold standard refers to the set of manual references considered the most reliable and accurate form of segmentation available, against which automatic methods are compared. The process involved consensus in cases of disagreement, ensuring greater reliability of the reference set. Thus, they constitute the training base and serve as a criterion for quantitative validation, functioning as an essential comparison point to assess the quality of automatic segmentations.

The images were initially cropped by an automated region of interest (ROI), defined from the pixel intensity distribution. Figure 2 illustrates this cropping process guided by the intensity distribution, used to exclude irrelevant regions (annotations, borders, and background) and preserve the lumen. Then, Contrast Limited Adaptive Histogram Equalization (CLAHE) [4] was applied, normalization to the range [0,1] (followed by standardization when indicated), resizing to 256×256 , and a set of data augmentations (rotations, translations, elastic distortions, Gaussian noise, brightness and contrast variation) implemented with Albumentations and OpenCV [15, 16]. Finally, the database was stratified by patient into training (70%), validation (15%), and test (15%), ensuring that images from the same individual did not appear in more than one partition, thus avoiding information leakage. It is worth noting, however, that the test set still comes from the same patient database. The use of an external and completely independent set would constitute the most robust evaluation, but was not available in this study, being pointed out as a methodological limitation.

It is worth noting that the ultrasound frames have low SNR and strong presence of speckle. CLAHE was chosen over more sophisticated techniques, such as BM3D or median filters, due to its balance between simplicity and computational cost, maintaining good efficacy in edge enhancement and enabling use in near real-time scenarios.

3.3 Image Registration

A grayscale image I is represented by intensities $I(x, y)$ in the range $[0, L - 1]$, for each pixel with $x \in [0, H - 1]$ and $y \in [0, W - 1]$, where H is the height, W the width of the matrix and $L = 2^b$, where b corresponds to the number of bits. In the present study, $b = 8$, which results in 256 possible graylevels. The sums of intensities on the vertical (S_{vert}) and horizontal (S_{horiz}) axes are defined as:

$$S_{vert}(j) = \sum_{i=0}^{H-1} I(i, j), \quad j = 0, 1, \dots, W - 1 \quad (2)$$

$$S_{horiz}(i) = \sum_{j=0}^{W-1} I(i, j), \quad i = 0, 1, \dots, H - 1 \quad (3)$$

The projections S_{vert} and S_{horiz} highlight the cumulative distribution of intensities on each axis, functioning as profiles that highlight regions of higher signal density.

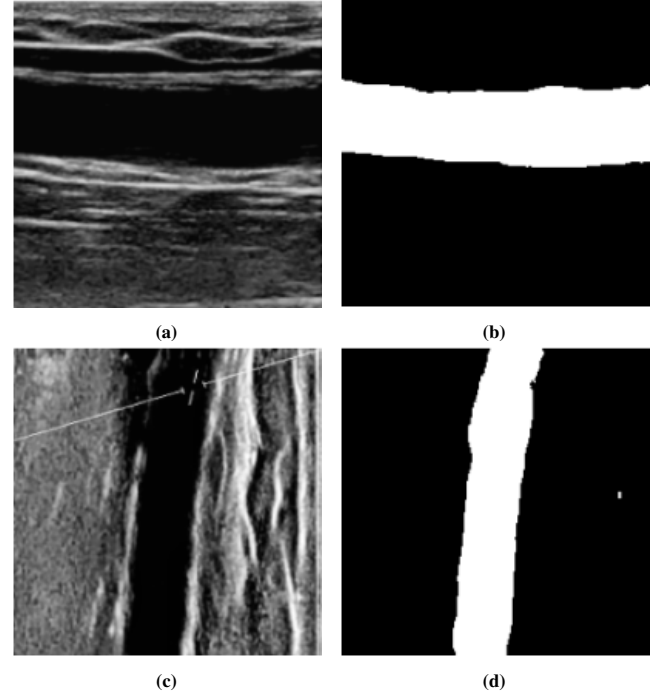


Figure 1. Examples of original ultrasound images (a,c) and their respective reference masks (gold standard) (b,d).

A regional maximum M of a grayscale image I is defined by the set of connected pixels with a certain intensity h , such that each of its neighboring pixels has a strictly lower value [17]. An operator O is said to be anti-extensive if, for any image I , it satisfies the property that the resulting image is a subset of the original image, i.e., $O(I) \leq I$. The extinction value $E_{atr}(M)$ of a maximum M consists of the successive application of operations O by a growing attribute atr [18], typically filtering by height, (relative to intensities), area (relative to the number of connected pixels) or volume (relative to the sum of intensities), until M ceases to be a regional maximum in the image. One can, therefore, order the regional maxima by their extinction values, i.e.: the regional maximum that requires the most filtering to disappear, followed by the regional maximum that requires the second most filtering to disappear, and so on. For the registration of ultrasound images, M_1 and M_2 are used as the two regional maxima with the highest volume extinction E_{vol} [19, 18]. The signals S_{vert} and S_{horiz} are morphologically reconstructed [17, 18], Rec , from M_1 and M_2 (in the case of S_{vert} only M_1 is used). Then, the average between the intensities of the base and the top of the Rec signal is used as the threshold value (yellow dashed line in Figure 2), i.e., for $k = 0, 1, \dots, N - 1$ ($N = W$ in the vertical, and $N = H$ in the horizontal):

$$S_{bin}(k) = \begin{cases} 1, & \text{if } Rec(k) \geq \frac{\max(Rec(k)) - \min(Rec(k))}{2}, \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

The start and end columns or rows (green and magenta arrows in Figure 2) correspond, respectively, to the first index

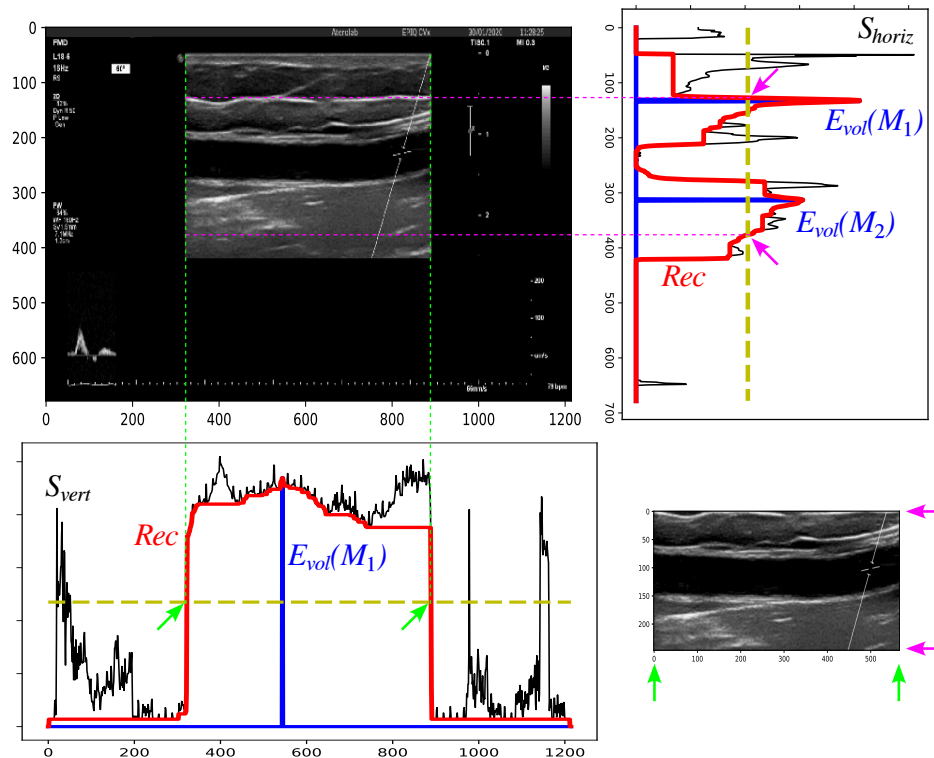


Figure 2. Registration of ultrasound images guided by pixel intensity distribution.

where S_{bin} takes value 1 and the last index where it returns to 0. These transition points automatically define the region of interest (ROI) boundaries, and are used to crop the original image and generate the final version in the lower right corner of Figure 2.

In the context of ultrasound images, the goal of registration is not to align different frames, as in classical computer vision applications, but to automatically crop a consistent ROI, removing dark borders and equipment artifacts. The notion of volume extinction (E_{vol}) allows quantifying the structural relevance of regional maxima: the higher the E_{vol} , the greater the persistence of the structure against filtering and, therefore, the more relevant its contribution to the global contrast. We select the two regional maxima with the highest E_{vol} (M_1 and M_2) because they generally correspond to the main intensity transitions associated with the lumen edges. From them, we reconstruct the projected signals (S_{vert} and S_{horiz}), which function as smoothed profiles capable of robustly locating the vessel boundaries even in the presence of speckle noise, guiding the final cropping shown in Figure 2.

3.4 Deep Learning Architectures

U-Net [5] has established itself as one of the most widely used architectures in biomedical image segmentation, due to its ability to represent different scales and preserve spatial details. For artery lumen segmentation, three variations of this architecture were investigated:

UNet++: uses nested skip connections that perform multi-scale fusion, promoting greater edge stability in low-contrast

scenarios [20];

MultiResUNet: employs blocks composed of 3×3 , 5×5 and 7×7 convolutions, enabling the capture of patterns at multiple scales and better handling of heterogeneous textures typical of ultrasound [21];

TransUNet (light variant): combines initial convolutions with a channel attention module (squeeze–excitation) and a shallow decoder, reducing complexity and the number of parameters compared to the original model with transformer encoder, but preserving the ability to model contextual dependencies [22].

All architectures were configured with a single input channel (grayscale images), base of 32 convolutional filters, and sigmoid output to map the pixel-wise probability [20, 21, 23, 5].

The rationale for selecting these architectures lies in their complementary strengths: UNet++ enhances multi-scale fusion and improves contour segmentation under noise; MultiResUNet increases robustness against texture variations; and the light TransUNet reduces computational cost while maintaining contextual attention. This comparison allows evaluating different trade-offs between accuracy, robustness, and computational efficiency.

3.5 Loss Function

To handle class imbalance and the subtle edges typical of ultrasound images, we employed the Focal Tversky Loss (FTL) [24], defined as:

$$\text{FTL} = (1 - T(\alpha, \beta))^\gamma, \quad T(\alpha, \beta) = \frac{\text{TP} + \varepsilon}{\text{TP} + \alpha \text{FP} + \beta \text{FN} + \varepsilon}. \quad (5)$$

where TP, FP and FN denote, respectively, the numbers of true positives, false positives and false negatives, defined pixel-wise as follows:

$$\text{TP} = \sum_{i \in \Omega} p_i g_i, \quad \text{FP} = \sum_{i \in \Omega} (1 - g_i) p_i, \quad \text{FN} = \sum_{i \in \Omega} g_i (1 - p_i), \quad (6)$$

with $p_i \in [0, 1]$ representing the predicted probability and $g_i \in \{0, 1\}$ the pixel label i . The discrete image domain is denoted by Ω . The term ε corresponds to a numerical stability factor (fixed at 10^{-7} in this work), which prevents divisions by zero in degenerate cases. The parameters α , β and γ control, respectively, the relative penalty of false positives, false negatives and the degree of focalization of the function.

In this study, the hyperparameters were adjusted to more heavily penalize false negatives, since the omission of contours tends to underestimate the arterial diameter and, consequently, compromise the FMD estimate. Thus, the loss function favors the preservation of vascular edges even in low-contrast regions, promoting more consistent segmentations.

3.6 Training

The comparison between architectures was performed with $K = 2$ cross-validation (shuffled KFold, seed=42) [25]. Optimization employed AdamW [26], with dynamic learning rate adjustment via ReduceLROnPlateau (factor 0.5) and early stopping monitoring the Dice, as defined in Section 3.8. Training explored mixed precision via Automatic Mixed Precision (AMP, via autocast + GradScaler) on GPU [27], implemented in PyTorch [28].

All experiments were conducted with deterministic cuDNN configuration (when applicable), ensuring reproducibility. The early stopping criterion was defined as no improvement in validation Dice within the stipulated patience (Section 3.8), while the learning rate reduction followed a standardized protocol for all architectures. This arrangement ensures fair comparability between models and stability in the obtained results.

3.7 Thresholding and Post-processing

The probabilistic output $p(x)$ for a pixel x was binarized with a threshold value τ specific to each model ($\tau = 0.45$ for UNet++; $\tau = 0.50$ for MultiResUNet; $\tau = 0.48$ for TransUNet):

$$\hat{y}(x) = \begin{cases} 1, & \text{if } p(x) \geq \tau, \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

The τ values were defined empirically from the validation set. For each architecture, we varied τ in increments of 0.01

in the interval $[0, 1]$, selecting the point that maximized the balance between validation Dice and IoU. This procedure led to the choice of $\tau = 0.45$ for UNet++, $\tau = 0.50$ for MultiResUNet and $\tau = 0.48$ for TransUNet. It is worth noting that the adjustment was performed exclusively on the validation set, preserving the independence of the test set and ensuring fair comparability between models.

Then, morphological closing 3×3 and selection of the largest connected component [29] were applied:

$$y_{\text{close}} = (\hat{y} \oplus B) \ominus B, \quad \hat{y}^* = \arg \max_{C \in \mathcal{C}(y_{\text{close}})} |C|, \quad (8)$$

where B is the 3×3 structuring element; $\oplus$ and $\ominus$ denote, respectively, dilation and erosion. Morphological closing consists of a dilation followed by erosion, an operation useful for eliminating small holes or irregularities within the segmented regions, making the contours smoother and more continuous. The selection of the largest connected component aims to discard small isolated regions (noise), keeping only the main structure of the artery, that is, the largest connected region in the binary mask. Formally, $\mathcal{C}(\cdot)$ represents the set of connected components, and $|C|$ indicates the cardinality (number of pixels) of each component.

It is important to mention that all aggregated metrics were calculated before applying Equation (8), ensuring that the reported values directly reflect the model's prediction, without influence from morphological post-processing.

3.8 Evaluation Metrics

For the evaluation of results, we report Dice, IoU and Accuracy (ACC) [30] as quality metrics:

$$\text{Dice} = \frac{2|P \cap G|}{|P| + |G|}, \quad \text{IoU} = \frac{|P \cap G|}{|P \cup G|}, \quad \text{ACC} = \frac{\text{TP} + \text{TN}}{|\Omega|}. \quad (9)$$

Definitions: $P = \{x : \hat{y}(x) = 1\}$ represents the set of pixels predicted as belonging to the lumen, while $G = \{x : g(x) = 1\}$ corresponds to the reference pixels (gold standard). TP (true positives) and TN (true negatives) indicate, respectively, the correct detections of positive and negative pixels. And $|\Omega|$ indicates the cardinality, i.e., the total number of pixels in the image.

The Dice metric measures the relative overlap between the prediction and the gold standard, ranging from 0 (no intersection) to 1 (perfect overlap). IoU (Intersection over Union) is the ratio between the area of intersection and the area of union between the two masks, being mathematically equivalent to the Jaccard index. Accuracy (ACC) quantifies the total proportion of correctly classified pixels, considering both background and lumen.

Additionally, more sensitive contour metrics can be used: Hausdorff Distance: measures the largest distance between the predicted and reference contours, indicating extreme local discrepancies. Average Symmetric Surface Distance (ASSD): calculates the average distance between predicted and real

surfaces in both directions, providing a balanced view of the error. Boundary F1-score: evaluates the overlap between predicted edges and real edges, combining precision and recall specifically in the contour regions.

In this work we adopt the nomenclature IoU, as it is the most usual in computer vision. Furthermore, the values were reported with three decimal places, sufficient to show relevant differences without conveying false numerical precision.

3.9 Environment and Implementation

The execution of the experiments was performed on Google Colab (with GPU, whenever available). The pipeline was implemented in Python/PyTorch and covered all steps presented in this work: upload and extraction; dataset preparation with registration, CLAHE and normalization; definition of models and FTL; training with AdamW, scheduler, early stopping and AMP; and quantitative and qualitative evaluation.

The hyperparameters adopted for each architecture were kept under the same training set, and are summarized in Table 1.

4. Results and Discussion

The results presented below combine quantitative evaluations, expressed by segmentation metrics (Dice, IoU and Accuracy), and qualitative examples in ultrasound images. This approach allows an integrated evaluation of both the performance of the architectures in numerical terms and the practical impact of the segmentations on the FMD calculation.

4.1 Quantitative Results

Table 2 presents the average performance of each model on the validation and test sets, while Figure 3 comparatively illustrates the metrics (ACC, Dice and IoU). A consistent advantage of UNet++ over the other architectures is observed in the evaluated set.

All metrics were calculated directly from the raw predictions, before the morphological post-processing described in Section 3.7, ensuring that the reported results reflect only the intrinsic output of each architecture. Morphological closing and selection of the largest connected component were employed only for qualitative visualization uniformity. Additional tests showed that the average quantitative impact of these operations on Dice and IoU was marginal (less than 0.5%), not altering the comparative conclusions between models.

4.2 Qualitative Results

Figure 4 shows representative examples with the predicted contours (in red) for each architecture in low contrast and speckle scenarios, illustrating the behavior of the models in the segmentation task.

Speckle refers to a granular noise pattern typical of ultrasound images, which reduces edge clarity. Acoustic shadows correspond to dark regions behind dense structures that block the sound beam. Specular artifacts are bright echoes produced

Table 1. Hyperparameters by architecture.

Model	LR	WD	Batch	ES	Sched
UNet++	7.5×10^{-5}	5×10^{-6}	8/4	6	2
MultiResUNet	1.0×10^{-4}	1×10^{-5}	8/4	9	3
TransUNet	8.0×10^{-5}	5×10^{-6}	8/4	6	2

LR: initial learning rate

WD: weight decay

Batch: training batch size

ES: early stopping patience

Sched: scheduler patience

by intense reflections on smooth surfaces, which can distort the perception of real edges.

Despite the good overall performance, we observed failures in adverse situations, such as these degraded image conditions, where the predictions were less stable. In these cases, MultiResUNet approached UNet++, while TransUNet exhibited greater contour variability. These examples reinforce the need to evaluate the models under realistic conditions, common in clinical examinations.

4.3 Discussion

The observed advantage of UNet++ is consistent with the literature, which associates its nested skip connections with multi-scale fusion and greater edge stability in low-contrast images [20]. Similar results have been reported in studies of carotid artery segmentation [31] and brachial artery segmentation in FMD examinations [9], both highlighting the superiority of UNet++ over conventional architectures. In the experiments conducted in this work, this trend was confirmed empirically: UNet++ outperformed MultiResUNet on average by 1.8 percentage points in Dice and 2.3 points in IoU, and was superior to TransUNet by about 3.0 points in Dice and 2.7 in IoU. These quantitative gains therefore support not only a theoretical expectation, but a practical evidence that UNet++ is more suitable for handling subtle edges in vascular ultrasound.

MultiResUNet presented benefits derived from the composition of kernels of different sizes (3×3, 5×5, 7×7), favoring the capture of patterns at multiple scales, especially in heterogeneous ultrasound textures [21]. The light TransUNet reduced computational cost and the number of parameters, but showed limitations in modeling long-range dependencies, behavior consistent with comparative analyses in ultrasound [10] and recent findings of hybrid architectures combining CNNs and Transformers [12]. These works reinforce that the superiority of UNet++ observed here is not isolated, but fits into a broader body of evidence about the differentiated performance of multi-scale architectures in vascular ultrasound.

From a clinical perspective, the balance between recall and precision is determining. Recall indicates the proportion of true positives detected among all relevant elements, reflect-

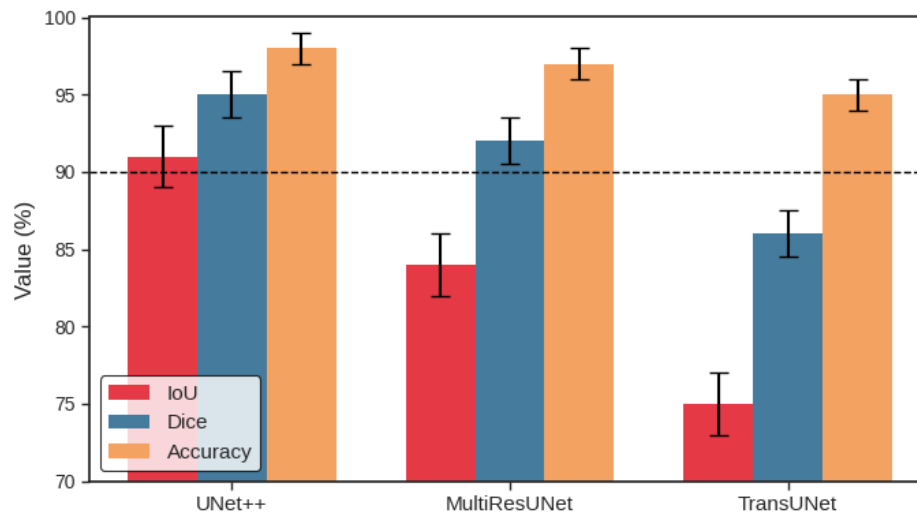


Figure 3. Comparison of Accuracy, Dice and IoU between models (Table 2).

ing the ability not to miss correct cases. Precision indicates the proportion of positive predictions that actually correspond to correct cases, reflecting the model's reliability in its predictions. False negatives at the edges tend to underestimate the arterial diameter, resulting in FMD values lower than the real ones, while false positives may overestimate it. Thus, the choice of a model should consider not only global segmentation metrics, but also the impact on the derived physiological measure.

Methodological limitations: The experimental protocol adopted cross-validation with only two folds ($K=2$). This decision sought to balance computational cost and hyperparameter exploration, but limits the estimate of inter-subject variability. Future work should consider a larger number of folds or external validation on independent databases, as suggested in investigations on vascular segmentation [31, 9, 11], in order to increase the robustness of the findings.

Another relevant aspect is the absence of a direct analysis between the FMD values derived from the predicted masks and those obtained manually by specialists. This comparison would allow quantifying the systematic error between methods, bringing the results closer to the real clinical context. We recognize this gap as a limitation and identify it as a priority path for future work.

Furthermore, although the study adopted consolidated metrics (Dice, IoU and Accuracy), these measures may be insufficient to capture edge discrepancies, critical for the arterial diameter calculation. In this sense, metrics such as Hausdorff Distance, Average Symmetric Surface Distance and boundary F1-score offer more sensitive alternatives for evaluating contour quality. Preliminary tests indicated that morphological post-processing produced only marginal adjustments (0.2% in average Dice), without altering the hierarchy between architectures, which is why we chose to report exclusively the raw predictions in this article.

Table 2. Average segmentation results by architecture.

Model	Accuracy	Dice	IoU
UNet++	0.98	0.9503	0.9052
MultiResUNet	0.97	0.9236	0.8400
TransUNet	0.95	0.8575	0.7511

Implementation improvements: During the analysis of the results, we identified implementation adjustments capable of increasing the consistency of the pipeline without altering the main conclusions: (i) align the metric aggregation, saving the IoU from the same epoch where the Dice reaches its best value; (ii) apply light gradient clipping ($\|g\|_2 \leq 1.0$), a technique that limits the maximum value of gradients during training to avoid numerical explosions; (iii) smooth some data augmentations in MultiResUNet, reducing validation oscillations; (iv) standardize seed, deterministic cuDNN mode and early stopping/scheduler logs, ensuring reproducibility; and (v) maintain τ values specific to each architecture ($\tau = 0.45$ for UNet++; $\tau = 0.50$ for MultiResUNet; $\tau = 0.48$ for TransUNet), adjusted only in validation. These refinements can be incorporated in future replications, favoring stability and fairness in the comparison between models.

Finally, ablation analyses, the adoption of more sensitive contour metrics, and the direct comparison between predicted FMD values and those measured by specialists constitute promising directions for future work. Furthermore, the use of ensemble strategies (which combine multiple models to generate a more robust prediction) can be explored to improve the stability and generalization of the results.

5. Conclusion

This study demonstrated the comparative performance of three deep learning architectures in brachial lumen segmentation in B-mode ultrasound, providing practical support for sim-

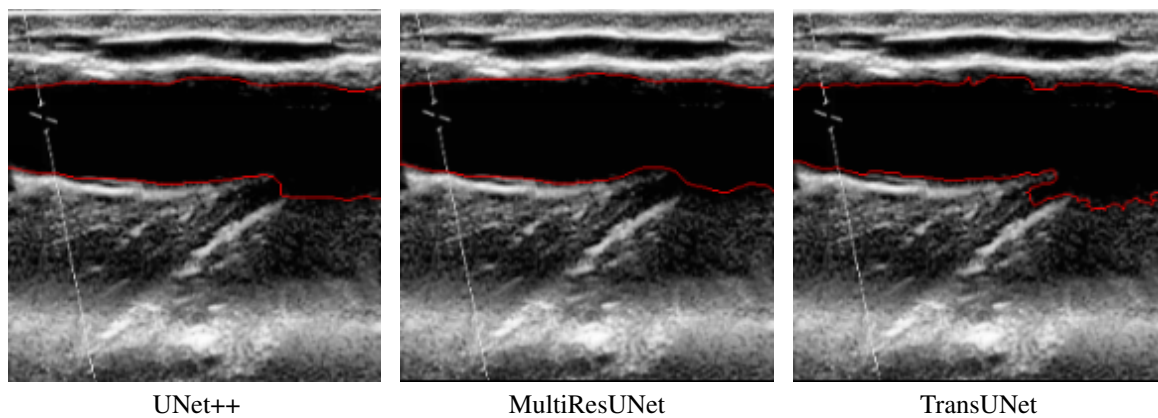


Figure 4. Qualitative segmentation examples: predicted contours (red) for each architecture.

ilar clinical scenarios. The analysis, conducted on 935 image–mask pairs with preprocessing via CLAHE and cross-validation, indicated that UNet++ presented the best average performance. The systematic comparison with MultiResUNet and TransUNet reinforces the suitability of UNet++ for low-contrast contexts, highlighting the relevance of multi-scale skip connections in edge stabilization.

In addition to providing quantitative and qualitative evidence, the results highlight the importance of relating segmentation metrics to the direct clinical impact on FMD calculation. The physiological interpretation depends on the accuracy of the contours, and subtle edge errors can significantly alter the derived dilation measure, influencing the assessment of endothelial function.

As immediate future work, we plan to: (i) perform ablations involving pre and post-processing steps (ROI, CLAHE and morphological operations); (ii) incorporate more sensitive contour metrics, such as Hausdorff Distance and boundary F1-score, to evaluate edge discrepancies; (iii) systematically vary the binarization thresholds (τ) between architectures; (iv) apply ensemble strategies in order to investigate the complementarity between models; and (v) conduct external validation on new ultrasound sets, consolidating the generalization capacity of the methods.

In summary, the work reinforces the potential of deep segmentation models as an auxiliary tool in FMD examination, bringing computational results closer to clinical practice and opening the way for future advances in vascular ultrasound analysis.

Author Contributions

Breno Luiz Nunes da Silva was responsible for the complete implementation of the segmentation pipeline (including image preprocessing, dataset preparation and organization, application of the UNet++, MultiResUNet, and TransUNet architectures, execution of experiments, hyperparameter tuning, and both quantitative and qualitative evaluation of the results), critical analysis of the findings, figure preparation, and technical writing of the manuscript; Alexandre Gonçalves Silva

contributed to the methodological design of the study, overall supervision of the research, implementation of the automatic image registration module, definition of ROI extraction strategies, scientific review, guidance of computational aspects, ensuring consistency and integration across methodological steps, and interpretation of the results; Lucas Ziza Nastari was responsible for constructing and reviewing the reference masks (ground truth), manual labeling and dataset curation, definition of validation strategies, and complementary analyses of the results; Rangel Arthur contributed to the construction and organization of the dataset, experimental planning and training protocol definition, methodological review, and validation of preprocessing and data structuring steps.

References

- [1] World Health Organization. *Cardiovascular Diseases (CVDs) — Fact Sheet*. 2021. [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)). Accessed: 2025-08-30.
- [2] KASS, M.; WITKIN, A.; TERZOPOULOS, D. Snakes: Active contour models. *International Journal of Computer Vision*, Springer, v. 1, n. 4, p. 321–331, 1988.
- [3] ZHOU, S.; CHELLAPPA, R.; MOGHADDAM, B. Tracking and segmenting multiple objects simultaneously using particle filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 27, n. 3, p. 441–456, 2006.
- [4] ZUIDERVELD, K. Contrast limited adaptive histogram equalization. In: *Graphics Gems IV*. [S.l.]: Academic Press, 1994. p. 474–485.
- [5] RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. *Proceedings of MICCAI*, Springer, p. 234–241, 2015.
- [6] HUANG, Q. et al. Deep learning in medical ultrasound image analysis: A review. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, IEEE, v. 70, n. 1, p. 55–72, 2023.

- [7] MARTINEZ, L.; PATEL, R.; WONG, K. Advances in vascular ultrasound imaging and segmentation: Emerging tools and clinical applications. *Frontiers in Cardiovascular Medicine*, Frontiers, v. 10, p. 1123456, 2023.
- [8] SONG, Y. et al. Automated carotid plaque detection and segmentation in ultrasound using u-net++. *Computer Methods and Programs in Biomedicine*, Elsevier, v. 214, p. 106549, 2022.
- [9] ZHOU, L. et al. Automatic brachial artery segmentation in flow-mediated dilation ultrasound using deep learning and post-processing strategies. *Ultrasound in Medicine & Biology*, Elsevier, v. 49, n. 2, p. 345–356, 2023.
- [10] AZZOPARDI, G.; CAMILLERI, K. et al. Vascular ultrasound image segmentation using convolutional neural networks: A comparative study. *Biomedical Signal Processing and Control*, Elsevier, v. 68, p. 102731, 2021.
- [11] ZHANG, J. et al. Spatial attention multiscale network for carotid artery segmentation in ultrasound. *Medical Image Analysis*, Elsevier, v. 80, p. 102480, 2022.
- [12] LI, W. et al. Hybrid cnn-transformer architecture for peripheral artery ultrasound segmentation. *Artificial Intelligence in Medicine*, Elsevier, v. 152, p. 102435, 2024.
- [13] NGUYEN, T.; ZHAO, M.; ROBERTS, P. Hybrid 3d deep learning architectures for ultrasound vascular segmentation. In: SPRINGER. *Proceedings of MICCAI*. [S.l.], 2023. p. 245–256.
- [14] VASWANI, A. et al. Attention is all you need. *Advances in Neural Information Processing Systems (NeurIPS)*, v. 30, p. 6000–6010, 2017.
- [15] BUSLAEV, A. et al. Alumentations: Fast and flexible image augmentations. *Information*, MDPI, v. 11, n. 2, p. 125, 2020.
- [16] BRADSKI, G. The opencv library. *Dr. Dobb's Journal of Software Tools*, 2000.
- [17] VINCENT, L. Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms. *IEEE Transactions on Image Processing*, IEEE, v. 2, n. 2, p. 176–201, 1993.
- [18] SILVA, A. G.; LOTUFO, R. de A. Efficient computation of new extinction values from extended component tree. *Pattern Recognition Letters*, v. 32, n. 1, p. 79–90, 2011. ISSN 0167-8655.
- [19] VACHIER, C.; MEYER, F. Extinction value: A new measurement of persistence. In: *Proceedings of IEEE Workshop on Nonlinear Signal and Image Processing*. [S.l.: s.n.], 1995. p. 254–257.
- [20] ZHOU, Z. et al. Unet++: A nested u-net architecture for medical image segmentation. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. [S.l.]: Springer, 2018. (Lecture Notes in Computer Science, v. 11045), p. 3–11.
- [21] IBTEHAZ, N.; RAHMAN, M. S. Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation. *Neural Networks*, Elsevier, v. 121, p. 74–87, 2020.
- [22] CHEN, J. et al. Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. *Medical Image Analysis*, Elsevier, v. 92, p. 103280, 2024.
- [23] CHEN, J. et al. Transunet: Transformers make strong encoders for medical image segmentation. In: *arXiv preprint arXiv:2102.04306*. [S.l.: s.n.], 2021.
- [24] ABRAHAM, N.; KHAN, N. M. A novel focal tversky loss function with improved attention u-net for lesion segmentation. *Proceedings of ISBI*, IEEE, p. 683–687, 2019.
- [25] PEDREGOSA, F. et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.
- [26] LOSHCHILOV, I.; HUTTER, F. Decoupled weight decay regularization. *International Conference on Learning Representations (ICLR)*, 2019.
- [27] MICIKEVICIUS, P. et al. Mixed precision training. *International Conference on Learning Representations (ICLR)*, 2018.
- [28] PASZKE, A. et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems (NeurIPS)*, v. 32, p. 8024–8035, 2019.
- [29] SOILLE, P. *Morphological Image Analysis: Principles and Applications*. [S.l.]: Springer, 2003.
- [30] TAHA, A. A.; HANBURY, A. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging*, BioMed Central, v. 15, n. 1, p. 29, 2015.
- [31] SONG, Y. et al. Automatic carotid artery segmentation in ultrasound images using u-net++. *Computer Methods and Programs in Biomedicine*, Elsevier, v. 221, p. 106899, 2022.