Original Article

Automating Brachial Plexus Scan: Wireless Handheld Ultrasound with Deep Learning over Ten Locations

CME Credits

Min-Jie Yang[1,2]*, Hao-Kuang Wang[1,3], Yi-Qi Zhang[1,4]

[1]Department of Neurosurgery, E-Da Hospital, Kaohsiung, Taiwan, [2]Department of Surgery, School of Medicine, National Taiwan University, Taipei, Taiwan, [3]Department of Surgery, Institute of Clinical Medicine, National Cheng Kung University, Tainan, Taiwan, [4]Department of Surgery, School of Medicine, China Medical University, Taichung, Taiwan

Abstract

Background: Topographical ultrasound is gaining traction for brachial plexus visualization due to its value in regional anesthesia. However, existing artificial intelligence models for nerve localization are trained on high-resolution stationary ultrasound images, limiting their applicability to more convenient, low-resolution handheld devices. This study addresses this challenge by proposing a novel image segmentation model suitable for low-resolution images. **Methods:** Thirty adult patients provided informed consent for participation in this study. A high-frequency, portable ultrasound probe was used to acquire B-mode images and video clips at 20 predefined positions. A training dataset of 60,000 images was constructed with expert annotations for landmark localization. A two-stage convolutional neural network architecture was implemented: Stage 1 for image classification and Stage 2 for segmentation with centroid refinement. Four novice physicians underwent testing for comparison. Model performance was evaluated using classification accuracy and segmentation precision metrics. **Results:** Our model achieved high accuracy in classifying brachial plexus ultrasound image positions (99.2% sensitivity and 84.5% specificity) compared to novice physicians (81.1% sensitivity and 59.8% specificity). In addition, the model demonstrated significantly superior performance in landmark segmentation, with lower median distance error (0.19 mm vs. 4.9 mm) and superior shape similarity metrics (average symmetric surface distance, Hausdorff distance and intersection over union) compared to novice physicians. **Conclusion:** We present a novel image segmentation model for brachial plexus ultrasound images. The model achieved high classification accuracy and significantly surpassed novice physicians in landmark segmentation. This performance suggests the potential for the model as an educational tool to aid novice physicians in learning brachial plexus anatomy.

Keywords: Artificial intelligence-based nerve identification, brachial plexus, handheld ultrasound, portable device deployment

Introduction

Recent years have witnessed a growing body of research investigating the application of topographical ultrasound for brachial plexus visualization, driven by its recognized clinical utility in regional anesthesia and peripheral nerve blocks.[1-3] The complex anatomy of the cervical region has spurred the development of artificial intelligence (AI)-assisted models to facilitate localization of relevant structures, particularly for novice practitioners. However, a key limitation of existing models lies in their training on high-resolution images acquired with cart-based ultrasound systems. These models may not generalize well to images obtained using more convenient, portable, wireless handheld devices. The advantages of handheld devices include their compact size, lightweight design, and the ability to store acquired images on personal smartphones or tablets. However, these devices typically have fewer channels in micro-beamforming system, leading to inherently lower image resolution.

Conventional AI-assisted image segmentation tasks often rely on U-net architectures as the machine learning model. However, U-net's inherent feature extraction processes, including max pooling and convolution layers, can lead to significant feature loss. This limitation may render U-net unsuitable for low-resolution images acquired with wireless handheld ultrasound devices. Such images might lose crucial

Received: 29-05-2024 **Revised:** 26-07-2024 **Accepted:** 29-08-2024 **Available Online:** 17-03-2025

Access this article online

Quick Response Code:

Website: https://journals.lww.com/jmut

DOI: 10.4103/jmu.jmu_61_24

Address for correspondence: Dr. Min-Jie Yang, Department of Neurosurgery, E-Da Hospital, No. 36, Ln. 146, Chongli Road, Zuoying, Kaohsiung 813024, Taiwan.
E-mail: b05401128@ntu.edu.tw

How to cite this article: Yang MJ, Wang HK, Zhang YQ. Automating brachial plexus scan: Wireless handheld ultrasound with deep learning over ten locations. J Med Ultrasound 2025;33:316-28.

anatomical details during the downsampling (descending arm) phase of the U-net architecture, ultimately compromising their suitability for training convolutional neural networks (CNNs). In this study, we propose a novel image segmentation model to address this challenge. We aim to demonstrate that this model can achieve performance comparable to, or exceeding, that of novice physicians, thereby aiding them in learning the complexities of the brachial plexus anatomy.

Materials and Methods

Ethical approval and participant recruitment

This study was conducted in accordance with the Declaration of Helsinki and received ethical approval from the institutional ethics committee of E-DA Hospital, Kaohsiung, Taiwan (protocol number: EMRP-113-092) on September 2, 2023.

Thirty adult patients (aged 18–80 years, ASA physical status I or II) undergoing elective surgery at E-DA Hospital between October 2023 and January 2024 were recruited following written informed consent.

Inclusion and exclusion criteria

Patients were included if they met the following criteria: age 18–80 years, ASA physical status I or II, and scheduled for elective surgery. Exclusion criteria included: skin lesions or infection of the neck, known peripheral neuropathy, history of brachial plexus injury, previous neck injury/surgery, and allergy to ultrasound gel.

Data collection

Baseline demographics (patient identification, gender, age, height, and weight) were collected preoperatively in the ward setting. Patients were positioned supine with a slight head turn away from the operator and arms at their sides.

Ultrasound device

This study employed a high-frequency, portable, wireless ultrasound probe (Apache Linear 154, Aco Healthcare Cooperation, Taoyuan City, Taiwan) for image acquisition. The probe emits ultrasonic waves with frequencies ranging from 4 to 15 MHz and offers a maximum scanning depth of 12 cm. Its microwave beamforming system incorporates 64 channels.

Ultrasound imaging protocol

Target positions were identified based on established views for brachial plexus ultrasound, as outlined by Simoni *et al*.[4] [Figure 1]. Both B-mode images and approximately 10-s video clips were captured at a fixed depth of 4 cm. The acquisition process involved tracing the brachial plexus course across a total of 20 predefined positions (10 per side), as specified in Table 1. These positions encompassed the roots/rami (acquired in axial and sagittal oblique views), the C7 spinal root (axial view), the trunk (axial and coronal views), the divisions and cords (captured in both supraclavicular and infraclavicular views), and the axillary nerve (visualized using both axillary and posterior approaches).

All images and videos were prospectively reviewed and confirmed by a qualified neurosurgeon. The same procedure was repeated for the left brachial plexus.

Table 1: Position numbering

Numbering	Detailed description of the position
Position 1	Roots of the brachial plexus (axial view)
Position 2	Roots of the brachial plexus (sagittal oblique view)
Position 3	C7 spinal root (axial view)
Position 4	Trunk (axial view)
Position 5	Trunk (coronal view)
Position 6	Divisions
Position 7	Cords (supraclavicular view)
Position 8	Cords (infraclavicular view)
Position 9	axillary nerve (axillary approach)
Position 10	axillary nerve (posterior approach)

Training image dataset creation and labeling

Image frames were extracted from each video at a rate of 8 frames per second (fps), resulting in approximately 80 images per video. This procedure yielded a dataset of 60,000 images (6000 images per position) with a spatial resolution of 330 × 320 pixels. A subset of 40,000 images was randomly selected for training in the initial image classification stage. In addition, 4000 unqualified images (defined as lacking brachial plexus within the image scope) were collected across 20 positions. For test dataset formation, 400 qualified and 80 unqualified images were randomly selected from the remaining pools of 20,000 qualified and 4000 unqualified images, respectively. The test data were divided into four subsets, each containing 120 questions. Of these, 100 were associated with qualified images, whereas the remaining 20 featured unqualified images.

The second stage focused on pinpointing the specific locations of nerves and other anatomical structures within each image. A random sample of 60 images (30 per side) was drawn from each position, resulting in a total of 600 images for annotation. Two neurosurgical residents performed the initial annotation, with each image subsequently reviewed and confirmed by a senior neurosurgeon. Annotations were performed using Adobe Photoshop 2022 (San Jose, USA). Only images with concordant annotations from both residents and the senior neurosurgeon were included in the final dataset. This validated dataset was then divided for training and testing purposes, with 400 images allocated for training and the remaining 200 for testing. The 200 test images were further subdivided into four test sets, each containing 50 images for image segmentation evaluation. The expert annotations served as the "ground truth," representing the actual locations of the brachial plexus and other structures within each test image. A representative example of anatomical landmark annotation across 10 positions is presented in Figure 1. The overall study workflow is illustrated in Figure 2.

Novice physician recruitment and testing procedures

Recruitment targeted novice physicians in their 1st year of postgraduate training. The testing process spanned 2 days. On the 1st day, four participants attended a 2-h lecture course, including a 1-h hands-on training session on operating the

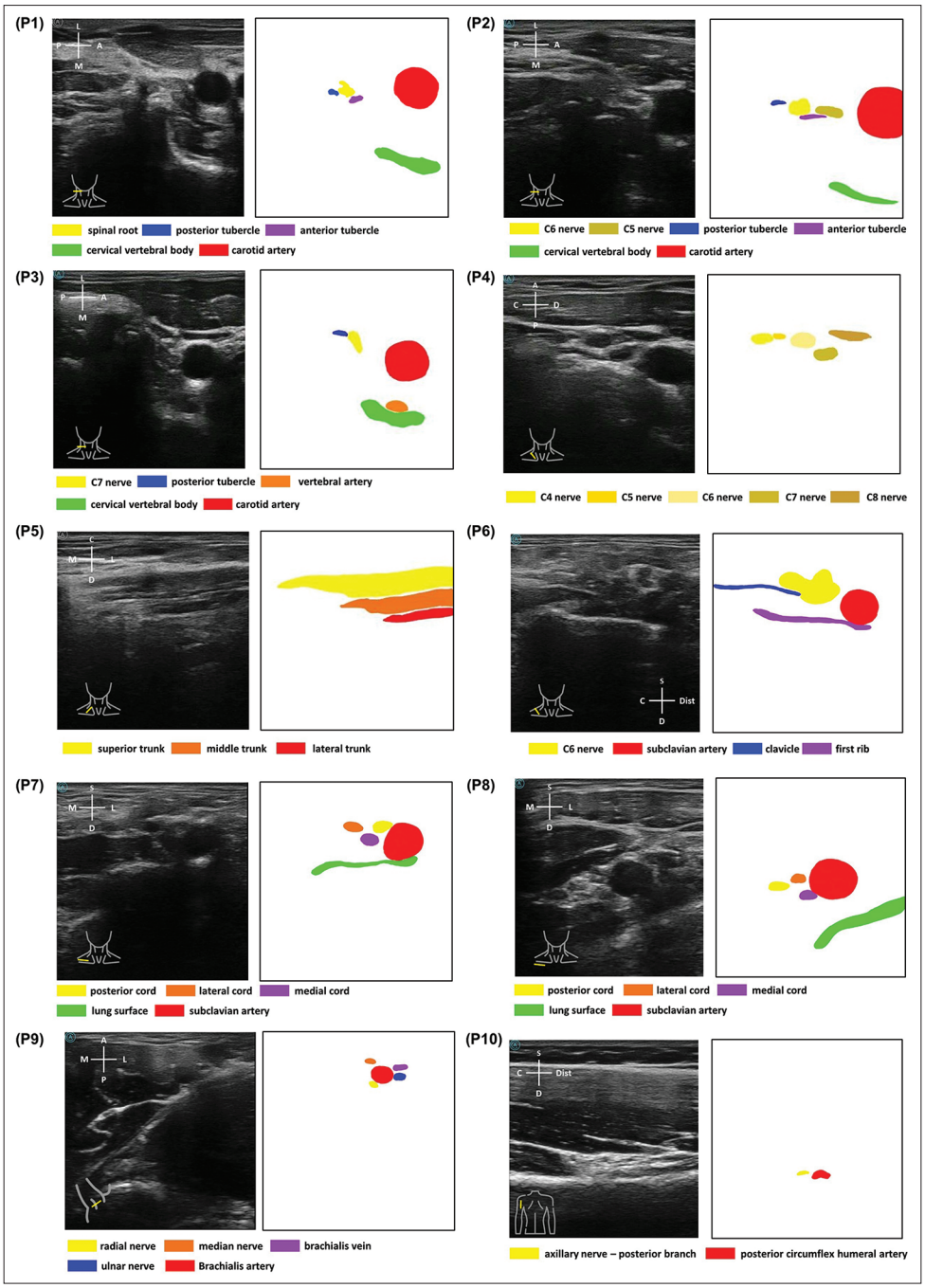

Figure 1: Anatomical landmark annotations and scanning views for brachial plexus ultrasound. Ultrasound imaging of the brachial plexus at ten designated sites (P1–P10). Anatomical structures are highlighted with distinct colors and identified within each image panel

handheld wireless ultrasound device under supervision. On the 2nd day, each physician received two electronic test sheets: One for image classification (120 questions) and another for image segmentation (50 questions). Participants were allocated 90 min to complete the image classification test and 100 min for the image segmentation test, with a 20-min break separating the two assessments. To ensure test independence, each physician received a unique test sheet set. The electronic test sheets were compatible with personal electronic devices such as laptops, tablets, or desktop computers.

Computational equipment

Our machine learning tasks were executed on the Google Colab platform utilizing an A100 GPU. Python 3.10 served as the primary programming language. Data processing and analysis were facilitated by a Microsoft Surface Laptop Go 3 equipped with an Intel® Core™ i5-1235U processor and Intel® Iris® Xe Graphics.

Convolutional neural network architecture

Given the inherent variability in shape, position, and echogenicity of the brachial plexus across different levels,

a CNN emerged as the optimal choice for our AI-based localization task within ultrasound images.

The first stage employed a CNN for image classification, primarily to determine the exact position (or view) represented within each image [Figure 3]. Images failing classification into any position received a "zero" probability across all classes in the prediction and were automatically assigned to a "position 0" classification.

The second stage of the segmentation process consisted of multiple steps. Initially, potential centers of mass (centroids) were identified within the image, followed by the generation of regular geometric masks (circles, rectangles, and ellipses) around these points. These masks underwent evaluation based on criteria including average echogenicity, area, and spatial relationship with other anatomical structures. Subsequent refinement of the mask area and centroid was performed based on these evaluation results. A position-specific CNN model, tailored for each of the ten predefined positions, was then employed to precisely calibrate the centroid of anatomical structures using diverse image features. The final output of the program represented the predicted centroid for these structures, as visualized in Figures 4 and 5. The resulting annotated images are depicted in Figure 6.

Model performance evaluation

The performance of both the trained model and novice physicians was assessed on the testing dataset. Classification accuracy and segmentation precision were employed as evaluation metrics. Reflecting the model's two-stage nature, the evaluation process was divided into two parts:

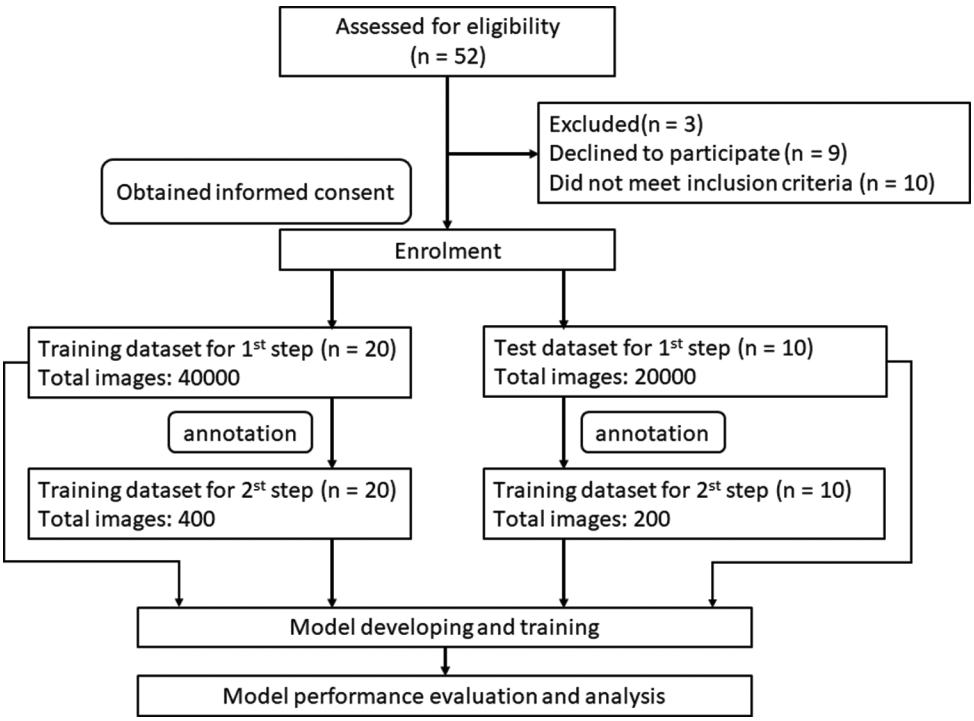

Figure 2: Preparation of the datasets and study workflow

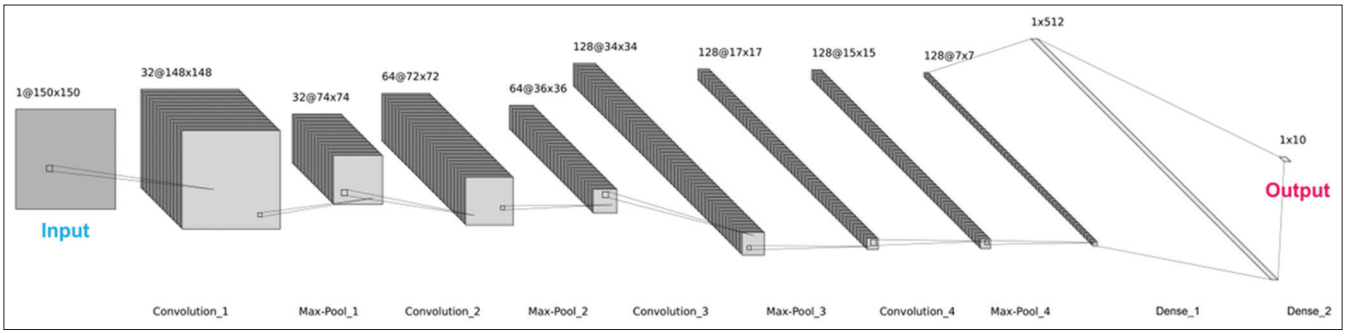

Figure 3: First stage classification model: Convolutional neural network (CNN) architecture. The figure illustrates the architecture of the proposed CNN model used for image classification. The model consists of four convolutional layers, four max-pooling layers, and two fully connected layers (Dense layer). The input to the network is a 150 × 150-pixel image. Convolutional layers extract features from the input image through a series of filters. Max-pooling layers down-sample the feature maps to reduce computational complexity while preserving essential information. The final fully connected layers combine the extracted features to produce the classification output

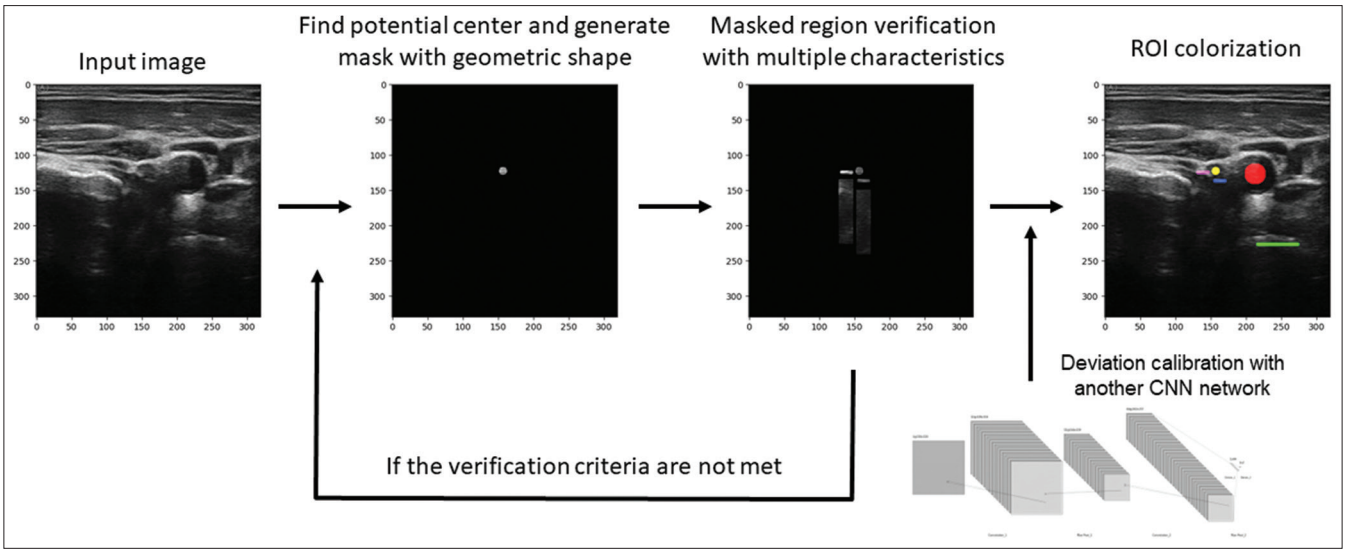

Figure 4: Second stage segmentation model: Region of interest localization. The input ultrasound image (left) is processed to identify potential centers of anatomical structures, around which geometric masks are generated (second from left). These masked regions undergo a verification process based on multiple characteristics (second from right). If the masked region fails the verification criteria, a new candidate point is automatically identified, triggering a subsequent iteration of the process (bottom). If verification is successful, the centroid and area of the mask are refined, and a position-specific convolutional neural network model is employed for precise centroid calibration (right). The final output is the predicted centroid for the target anatomical structures, depicted as colored dots on the original ultrasound image. ROI: Region of interest

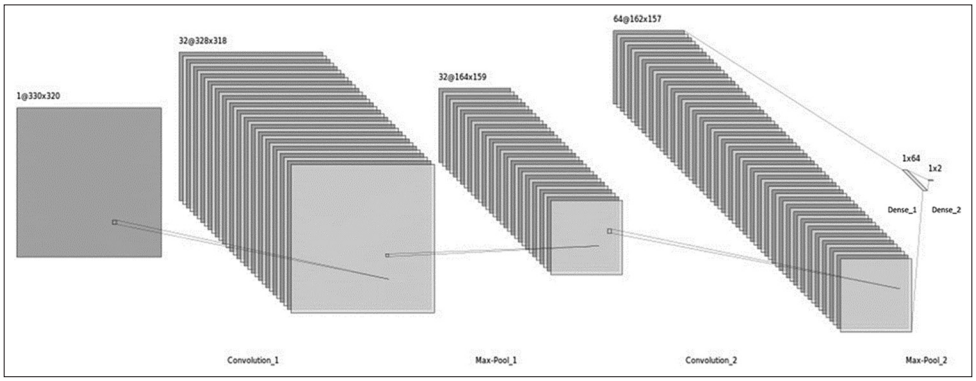

Figure 5: Deviation calibration model: Convolutional neural network (CNN) architecture. The CNN model is designed for fine-tuning the predicted centroid in ultrasound images. The input is a 330 × 320 image, which is passed through two convolutional layers (Conv1, Conv2) with 32 filters each. The filter size is 3 × 3 and stride 1. Max-pooling layers (Max-Pool 1, Max-Pool 2) with a pool size of 2 × 2 and stride 2 follow each convolutional layer. The output of the second max-pooling layer is flattened and passed through two dense layers (Dense 1, Dense 2) with 64 and 2 neurons, respectively. The final output is a two-dimensional vector representing the refined coordinates for the centroid

Classification performance assessment

Sensitivity and specificity were utilized to evaluate the performance of both the classification model and novice physicians during image prediction tasks.

Localization performance assessment

Segmentation precision was evaluated using a combination of distance and shape similarity metrics. The median distance, defined as the distance between the model's predicted centroid of the nerve and the corresponding ground truth annotations, was employed to assess localization accuracy [Figure 7]. In addition, shape similarity was quantified using three established metrics: Average symmetric surface distance (ASSD), Hausdorff Distance (HD), and Intersection over Union (IoU). ASSD measures the average distance between the surfaces of the predicted and ground truth segmentation masks. HD represents the maximum distance between any point on the surface of the predicted segmentation mask and the closest corresponding point on the ground truth mask. Finally, IoU calculates the ratio of the area of overlap between the predicted and ground truth segmentation masks to the total area of the union between these masks. A detailed visual representation of these concepts and their corresponding mathematical formulae can be found in Figure 8.

Statistical analysis

Patient characteristics are presented as mean ± standard deviation or number (%). Results are presented as median (interquartile range [IQR] or range) or percentage (95% confidence interval [CI]). The 95% CIs for proportions were calculated

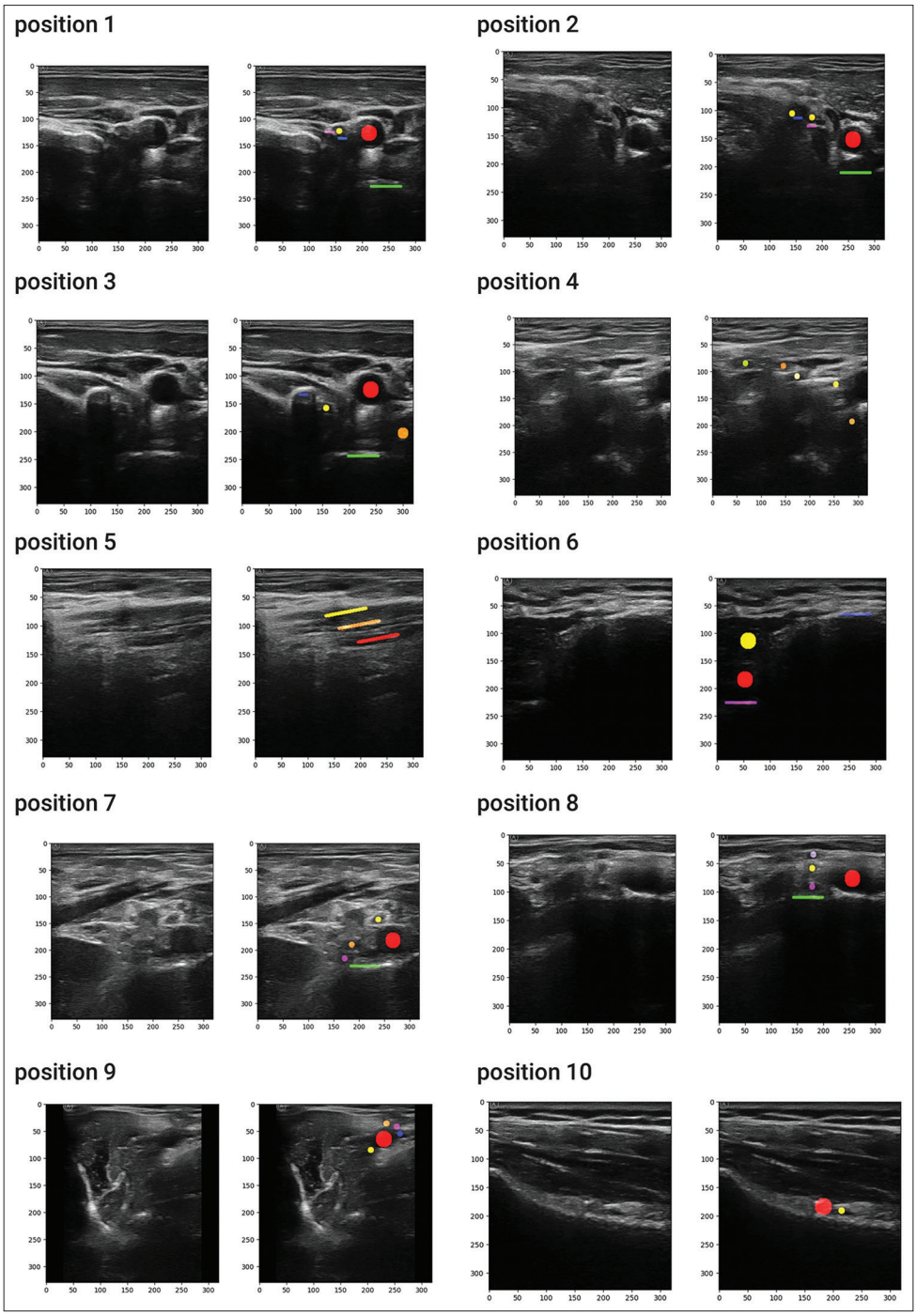

Figure 6: Model-generated annotations for ultrasound image. The figure presents ultrasound images of the brachial plexus at ten predefined positions. Overlaid on each image are color-coded annotations generated by the proposed model, indicating the localization of key anatomical landmarks. Refer to Figure 2 for the color code legend

using the exact binomial method. Statistical analyses were performed using Python 3.12.

Results

Patient demographics and dataset characteristics

A total of 30 patients were recruited for this study. Table 2 summarizes the baseline characteristics of the participants and the image datasets. No statistically significant differences were observed between the training and testing sets in terms of patient demographics. In addition, no image from the same patient was included in both the training and testing datasets to ensure independent evaluation.

Model performance

Classification model

Our trained network achieved high overall classification accuracy, with a sensitivity of 99.2% (95% CI: 98.6%–99.8%)

and specificity of 84.5% (95% CI: 73.8%–91.2%) as shown in Table 3 and Figure 8. These values were significantly superior to those achieved by novice physicians (sensitivity: 81.1%, 95% CI, 72.0%–90.2%; specificity: 59.8%, 95% CI, 52.6%–67.0%). However, a detailed analysis of individual positions revealed no statistically significant difference in specificity between the model and novice physicians at positions 4, 5, and 10. Similarly, position 7 exhibited no significant difference in either sensitivity or specificity. Despite this, the model maintained a trend of higher sensitivity (95.3%, 95% CI: 91.5%–99.1%) and specificity (80.4%, 95% CI, 66.5%–94.3%) at this position compared to novice physicians (sensitivity: 82.4%, 95% CI: 72.4%–92.4%; specificity: 63.6%, 95% CI: 57.9%–69.3%).

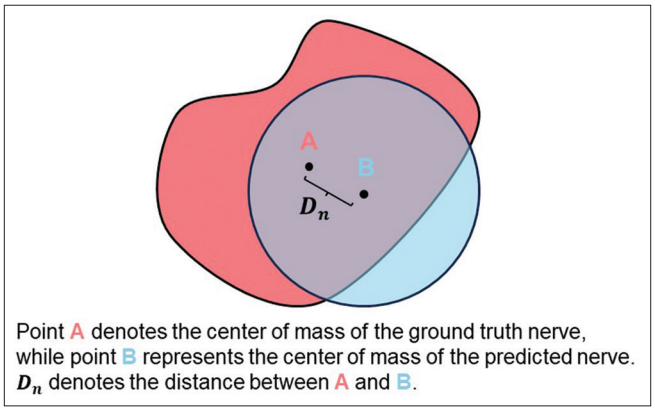

Figure 7: Distance metric: Euclidean distance between predicted and ground truth nerve centroids. The ground truth nerve is represented by the red shape, while the predicted nerve is shown in blue. Point A denotes the centroid of the ground truth nerve, and Point B denotes the centroid of the predicted nerve. The distance between these two points, denoted by (Dn), represents the Euclidean distance between the predicted and ground truth nerve centroids, a metric for quantifying prediction accuracy

Segmentation model

As illustrated in Table 4 and Figure 9, our model achieved significantly superior performance in landmark segmentation compared to novice physicians. The model exhibited a notably lower median distance error (0.19 mm [IQR 0.09–0.89]) compared to novice physicians (4.9 mm [IQR 3.2–8.0]). Shape similarity metrics further supported the model's advantage [Figure 10], with lower ASSD (1.7 mm [IQR 1.5–3.1] vs. 6.5 mm [IQR 5.8–9.1]) and HD (3.9 mm [IQR 2.9–5.7] vs. 15.2 mm [IQR 10.4–24.7]). While the IoU metric showed a less pronounced difference (model: 73.2% [IQR 60.7–79.7] vs. novice physicians: 46.2% [IQR 41.5–55.6], $P = 0.026$), it remained statistically significant. Importantly, this trend of superior model performance persisted across all individual landmark positions, with statistically significant differences observed for each landmark.

Discussion

This study explores the development of a novel image segmentation algorithm capable of achieving comparable performance to existing methods, even when applied to low-resolution ultrasound images. The algorithm leverages ultrasound image features to facilitate the localization of brachial plexus branches and other relevant anatomical landmarks within the cervical region.

Our CNN model achieved statistically superior performance compared to novice physicians in terms of both sensitivity and specificity, on average. However, positions 4, 5, and 10 exhibited no significant difference in specificity between the two groups. This can be attributed to the minimal anatomical variation and relatively straightforward topology in these regions. Conversely, position 7 demonstrated no

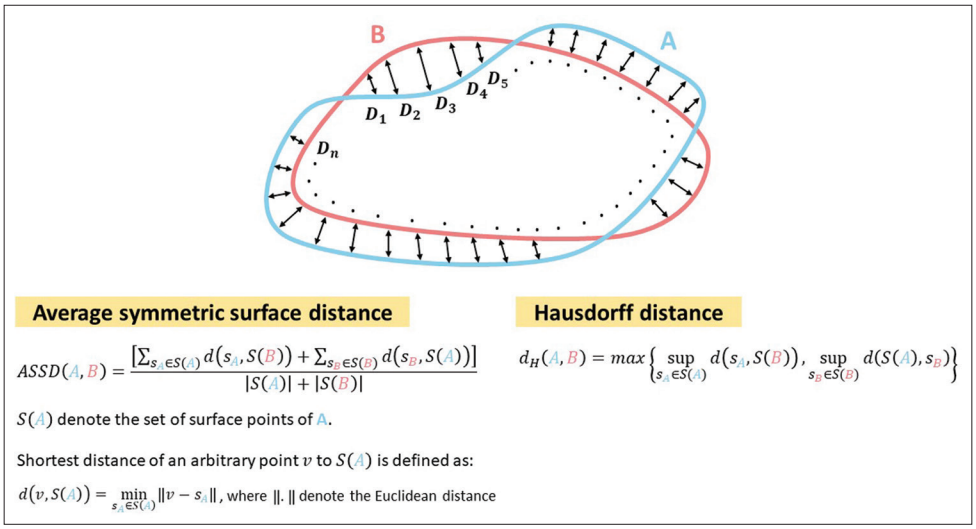

Figure 8: Visualization and mathematical formulation of average symmetric surface distance (ASSD) and Hausdorff distance (HD). The figure illustrates the calculation of ASSD and HD between two shapes, A and B. The ASSD is computed as the mean of the sum of the shortest distances from each point on the surface of one shape to the closest point on the surface of the other shape. The HD is defined as the maximum distance between any point on the surface of one shape to the closest point on the surface of the other shape. The figure visually depicts these metrics using sample points (Dn) and distance vectors (arrows) between the two shapes. ASSD: Average symmetric surface distance; HD: Hausdorff distance

Table 2: Characteristics of patients and dataset

	Training set (*n*=20)	Test set (*n*=10)
Patient characteristic		
Age (years)	57.5±13.5	57.2±16.5
Sex, *n* (%)		
Female	3 (15)	4 (40)
Male	17 (85)	6 (60)
Height (cm)	163.2±6.5	161.9±9.1
Weight (kg)	64.5±10.4	60.7±11.6
BMI (kg/m^2)	24.2±3.5	23.2±4.3
Image set characteristics		
Number of images in total	40,000	20,000
Number of images at each position (1–10) in total	4000	2000
Number of images at each position with target structures	3200	1600
Number of images at each position without target structures	800	400
Number of images with labeling mask	400	200
Number of images with labeling mask at each position (1–10)	40	20

Data are mean±SD, *n* (%). SD: Standard deviation, BMI: Body mass index

Table 3: Comparison of classification performance between model and novice physician

Algorithm used in 1st step (%)	Model	Novice physician average	*P*
Position 1			
Sensitivity	99.4 (99.2–99.6)	84.0 (73.1–94.9)	0.028
Specificity	81.6 (77.9–85.3)	61.1 (52.7–69.5)	0.015
Position 2			
Sensitivity	99.3 (99.2–99.4)	79.3 (70.2–88.4)	<0.001
Specificity	82.4 (75.2–89.6)	46.6 (38.7–54.5)	<0.001
Position 3			
Sensitivity	99.6 (99.3–99.9)	71.5 (62.4–80.6)	<0.001
Specificity	86.7 (83.0–90.4)	59.6 (52.3–66.9)	0.005
Position 4			
Sensitivity	99.8 (99.7–99.9)	81.4 (74.2–88.6)	<0.001
Specificity	80.5 (71.2–89.8)	62.8 (53.1–72.5)	0.066
Position 5			
Sensitivity	99.6 (99.4–99.8)	84.0 (74.7–93.3)	<0.001
Specificity	87.4 (79.4–95.4)	67.1 (54.7–79.5)	0.051
Position 6			
Sensitivity	99.6 (99.5–99.7)	84.9 (78.0–91.8)	<0.001
Specificity	80.7 (60.6–94.7)	53.1 (47.5–58.7)	0.047
Position 7			
Sensitivity	95.3 (91.5–99.1)	82.4 (72.4–92.4)	0.072
Specificity	80.4 (66.5–94.3)	63.6 (57.9–69.3)	0.118
Position 8			
Sensitivity	98.5 (97.6–99.4)	79.9 (69.0–90.8)	0.046
Specificity	85.8 (83.5–88.1)	65.0 (57.0–73.0)	0.023
Position 9			
Sensitivity	99.9 (99.8–100.0)	83.4 (77.4–89.4)	<0.001
Specificity	81.2 (74.3–88.1)	58.1 (51.3–64.9)	0.035
Position 10			
Sensitivity	99.8 (99.7–99.9)	80.0 (69.2–90.8)	0.034
Specificity	81.6 (65.7–97.5)	63.8 (57.4–70.2)	0.175
Overall performance			
Sensitivity	99.2 (98.6–99.8)	81.1 (72.0–90.2)	0.009
Specificity	84.5 (73.8–91.2)	59.8 (52.6–67.0)	0.026

significant difference in either sensitivity or specificity. This finding suggests that the originally captured images may contain objects with a high degree of similarity to the target object (nerves), posing a challenge for both human experts and the model in achieving perfect differentiation.

Prior research on machine learning applications in brachial plexus evaluation has yielded limited data on model sensitivity and specificity. The most recent study by Yang *et al*.[5] reported overall sensitivity and specificity of 97.7% and 84.6%, respectively, which aligns with our findings. While our study identified a significant difference between the model and physician performance, Yang *et al*.[5] did not observe such a disparity. This discrepancy may be attributed to differences in physician experience. Our study involved 1st-year postgraduate trainees, whereas Yang *et al*.[5] included anesthesia residents and attending physicians. The disparity in subspecialty training likely explains this observed difference.

Our model demonstrably surpassed novice physician performance in all segmentation metrics, including median distance and shape similarity metrics (ASSD, HD, and IoU). This outcome aligns with expectations, as human object recognition primarily relies on overall visual patterns rather than nuanced pixel-level differences. The latter is critical for precise edge detection in segmentation tasks. Notably, the model outperformed novice physicians even for positions with less anatomical variation, such as positions 4, 5, and 10. This finding further strengthens the preceding discussion regarding the model's ability to handle low-resolution ultrasound images.

A review of prior studies [Table 5] revealed that our model's IoU metric (73.2%, 95% CI = 60.7–79.7) falls within the reported range (53%–81.2%) and aligns with the acceptable error threshold (75%) established by Liu *et al*.[6] However, our model exhibited lower performance in shape similarity metrics compared to studies utilizing the largest image training datasets.[5] This observation was anticipated, as our model

Table 4: Comparison of segmentation performance between model and novice physician

Algorithm used in 2nd step	Model	Novice physician average	*P*
Position 1 (spinal root) (*n*=36)			
Median distance (mm)	0.09 (0.05–0.12)	2.5 (1.4–3.9)	<0.001
ASSD (mm)	1.3 (1.0–2.5)	5.1 (3.3–6.1)	0.007
HD (mm)	1.8 (1.6–3.2)	6.5 (3.7–11.2)	0.031
IoU (%)	76.1 (62.8–85.0)	53.5 (45.8–61.6)	0.037
IoU distribution, *n* (%)			
>80%	10 (27.7)	4 (11.1)	
50%–80%	20 (55.6)	8 (22.2)	
20%–50%	6 (16.7)	19 (52.8)	
<20%	0	5 (13.9)	
Position 2 (C5,6 nerve) (*n*=72)			
Median distance (mm)	0.16 (0.15–0.22)	4.1 (2.2–5.5)	<0.001
ASSD (mm)	2.0 (1.8–2.8)	8.0 (6.2–8.3)	<0.001
HD (mm)	4.2 (2.6–7.3)	15.5 (8.8–27.9)	0.007
IoU (%)	69.8 (59.2–77.1)	43.2 (36.5–49.2)	0.019
IoU distribution, *n* (%)			
>80%	15 (20.8)	1 (1.4)	
50%–80%	46 (63.9)	15 (20.8)	
20%–50%	11 (15.3)	49 (68.1)	
<20%	0	7 (9.7)	
Position 3 (C7 nerve) (*n*=36)			
Median distance (mm)	0.12 (0.08–0.19)	3.2 (1.3–5.3)	<0.001
ASSD (mm)	1.7 (1.5–3.1)	6.3 (4.8–11.3)	0.009
HD (mm)	2.9 (1.9–4.2)	9.9 (8.0–12.1)	0.007
IoU (%)	74.9 (62.3–82.0)	49.0 (47.3–57.6)	0.037
IoU distribution, *n* (%)			
>80%	6 (16.7)	3 (8.3)	
50%–80%	23 (63.9)	6 (16.7)	
20%–50%	7 (19.4)	23 (63.9)	
<20%	0	4 (11.1)	
Position 4 (C4, 5, 6, 7, 8 nerve) (*n*=180)			
Median distance (mm)	0.15 (0.07–0.37)	6.4 (4.2–9.8)	<0.001
ASSD (mm)	1.9 (1.6–2.8)	6.8 (4.9–11.6)	0.009
HD (mm)	4.6 (2.4–6.7)	17.0 (12.8–21.3)	0.009
IoU (%)	68.2 (56.4–76.6)	45.9 (40.8–54.1)	0.048
IoU distribution, *n* (%)			
>80%	40 (22.2)	15 (8.3)	
50%–80%	110 (61.1)	36 (20.0)	
20%–50%	30 (16.7)	105 (58.4)	
<20%	0	24 (13.3)	
Position 5 (sup., mid., inf., trunk) (*n*=108)			
Median distance (mm)	0.11 (0.10–0.23)	0.48 (0.24–1.5)	0.023
ASSD (mm)	1.1 (0.9–1.8)	3.6 (1.7–4.6)	0.033
HD (mm)	2.3 (1.5–3.9)	8.1 (7.0–13.7)	0.009
IoU (%)	81.7 (66.4–85.5)	60.9 (57.8–68.8)	0.047
IoU distribution, *n* (%)			
>80%	28 (25.9)	12 (11.1)	
50%–80%	63 (58.4)	65 (60.2)	
20%–50%	17 (15.7)	18 (16.7)	
<20%	0	13 (12.0)	
Position 6 (nerve divisions) (*n*=36)			
Median distance (mm)	0.53 (0.24–1.58)	7.2 (4.7–10.2)	0.006
ASSD (mm)	2.3 (2.0–4.4)	9.9 (8.1–12.3)	<0.001
HD (mm)	5.8 (3.4–7.9)	23.7 (19.0–30.9)	<0.001

Contd...

Table 4: Contd...

Algorithm used in 2nd step	Model	Novice physician average	*P*
IoU (%)	63.8 (47.2–72.7)	36.1 (34.8–42.5)	0.006
IoU distribution, *n* (%)			
>80%	6 (16.7)	0	
50%–80%	20 (55.6)	6 (16.7)	
20%–50%	10 (27.7)	24 (66.6)	
<20%	0	6 (16.7)	
Position 7 (lat., med., post. cord) (*n*=108)			
Median distance (mm)	0.29 (0.23–1.0)	10.5 (8.7–12.6)	<0.001
ASSD (mm)	2.1 (1.9–3.9)	7.4 (4.9–8.6)	0.008
HD (mm)	5.5 (5.1–9.4)	22.5 (14.5–31.6)	<0.001
IoU (%)	67.9 (57.8–74.4)	42.9 (41.3–52.5)	0.029
IoU distribution, *n* (%)			
>80%	21 (19.5)	7 (6.5)	
50%–80%	71 (65.7)	20 (18.5)	
20%–50%	16 (14.8)	69 (63.9)	
<20%	0	12 (11.1)	
Position 8 (lat., med., post. cord) (*n*=108)			
Median distance (mm)	0.08 (0.06–0.11)	10.7 (8.0–14.7)	<0.001
ASSD (mm)	1.1 (0.84–2.0)	3.5 (1.4–5.1)	0.012
HD (mm)	2.1 (1.5–2.5)	7.8 (4.3–13.5)	0.007
IoU (%)	82.9 (70.8–85.6)	58.4 (55.1–61.3)	0.049
IoU distribution, *n* (%)			
>80%	30 (27.8)	10 (9.2)	
50%–80%	59 (54.6)	61 (56.5)	
20%–50%	19 (17.6)	33 (30.6)	
<20%	0	4 (3.7)	
Position 9 (median, radial, and ulnar nerve) (*n*=108)			
Median distance (mm)	0.21 (0.13–0.36)	2.0 (0.8–5.4)	0.009
ASSD (mm)	1.9 (1.5–3.0)	9.3 (7.7–12.5)	<0.001
HD (mm)	3.8 (2.6–6.1)	13.3 (6.9–24.7)	0.048
IoU (%)	70.1 (53.3–80.0)	40.4 (37.4–46.1)	0.008
IoU distribution, *n* (%)			
>80%	26 (24.1)	10 (9.3)	
50%–80%	57 (52.8)	20 (18.5)	
20%–50%	25 (23.1)	67 (62.0)	
<20%	0	11 (10.2)	
Position 10 (axillary nerve) (*n*=36)			
Median distance (mm)	0.10 (0.06–0.19)	4.6 (2.3–6.7)	0.006
ASSD (mm)	1.4 (1.1–2.4)	4.6 (3.1–4.9)	0.008
HD (mm)	3.2 (1.9–5.1)	11.5 (8.5–17.3)	0.027
IoU (%)	77.6 (63.8–80.5)	48.7 (46.2–51.2)	<0.001
IoU distribution, *n* (%)			
>80%	9 (25.0)	2 (5.6)	
50%–80%	26 (72.2)	7 (19.4)	
20%–50%	1 (2.8)	23 (63.9)	
<20%	0	4 (11.1)	
Overall performance (*n*=828)			
Median distance (mm)	0.19 (0.09–0.89)	4.9 (3.2–8.0)	0.003
ASSD (mm)	1.7 (1.5–3.1)	6.5 (5.8–9.1)	0.011
HD (mm)	3.9 (2.9–5.7)	15.2 (10.4–24.7)	0.019
IoU (%)	73.2 (60.7–79.7)	46.2 (41.5–55.6)	0.026
IoU distribution, *n* (%)			
>80%	191 (23.1)	64 (7.7)	
50%–80%	495 (59.8)	298 (36.0)	
20%–50%	142 (17.1)	430 (51.9)	
<20%	0	90 (10.9)	

ASSD: Average symmetric surface distance, HD: Hausdorff distance, IoU: Intersection over union

Table 5: Review of prior image segmentation performance for brachial plexus in ultrasound images

Authors, year of publication	Number of images for training	Image resolution	Training model	Scanning device	Model performance (IoU) (%)
C. Liu *et al*., 2018[6]	5271	425 × 575	CNN	X	73.3
J. van Boxtel *et al*., 2021[7]	9541	420 × 580	U-net, M-net	X	72.0; 67
Y. Wang *et al*., 2021[8]	5635	420x580	U-Net	X	64.9
Yang, X. Y. *et al*., 2024[5]	11,392	603 × 840	U-net	Sonosite Edge II	81.2
Yu Wang *et al*., 2024[9]	320	510 × 356 and 595 × 529	U-net	Sonosite S-nerve and BK3000	53.0
Yang Xi *et al*., 2024[10]	502	790 × 1044	CNN	X	74.8

CNN: Convolutional neural network, IoU: Intersection-over-union

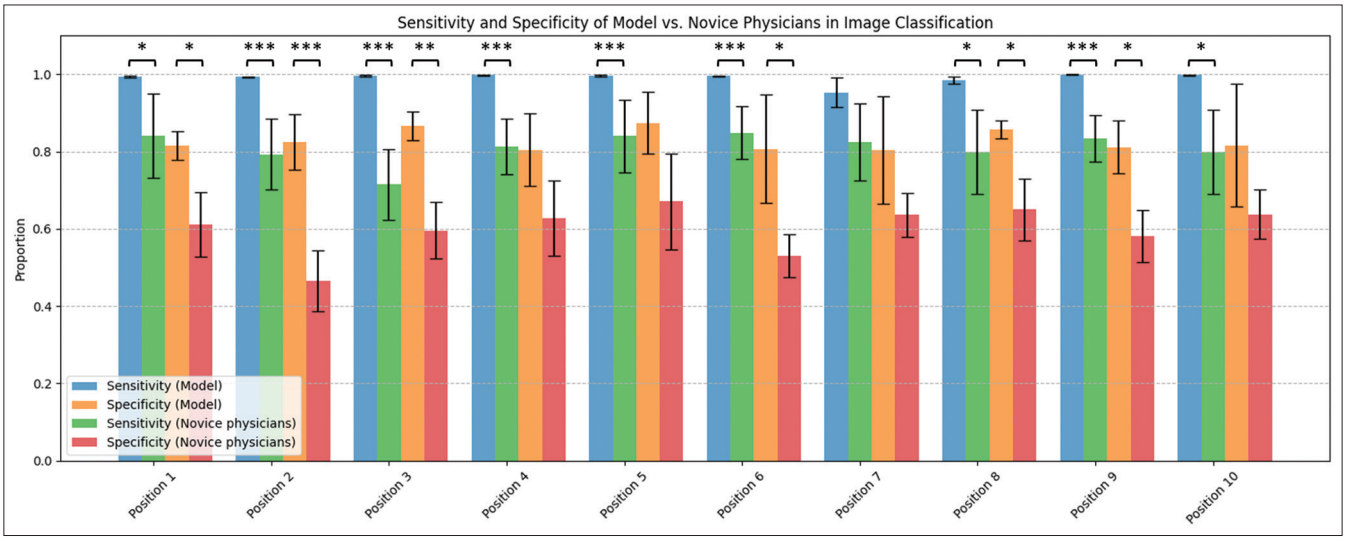

Figure 9: Sensitivity and specificity of model versus novice physicians in image classification. The performance of an artificial intelligence model and novice physicians in image classification was assessed across 10 different anatomical positions. Sensitivity (the ability to correctly identify positive cases) and specificity (the ability to correctly identify negative cases) are shown for both groups. Error bars represent 95% confidence intervals. One asterisk (*) typically indicates $P < 0.05$; Two asterisks (**) indicate $P < 0.01$; Three asterisks (***) indicate $P < 0.001$

prioritized the accurate positioning of the centroid within the segmented regions. In addition, the generated masks in our study employed regular shapes (circles, rectangles, and ellipses) in contrast to the irregular shapes adopted by Yang *et al*.,[5] which aimed to precisely approximate the actual nerve contours. In essence, we prioritized accurate centroid localization over precise nerve contour fidelity. While this approach might limit the model's utility in real-time ultrasound-guided procedures like nerve blocks, it does not hinder its educational value. Novice physicians primarily focus on learning the correct anatomical structure positions for diagnostic purposes, where nuanced nerve contour details are less critical.

Moreover, our models exhibit significant compactness in file size compared to existing models, which often reach hundreds of megabytes. In our case, the first classification component is approximately 40 Mb, and each landmark localization model is around 3 Mb. This compact design is advantageous for portable wireless devices, where storage capacity is a valuable resource. Even though modern mobile devices boast increasingly larger storage capacities, the reduced model size allows for additional space to accommodate other image processing or data transmission programs.

This study has limitations. While the model is designed to facilitate learning for novice physicians, its impact on their classification and segmentation skills was not directly assessed. Integrating the model into the target hardware, a portable wireless ultrasound device, would have been necessary for such an evaluation. However, this study lacked access to the required hardware due to limitations in sponsorship and intellectual property considerations related to the ultrasound device and its software. Consequently, the study focused on evaluating image segmentation performance in a static setting, rather than real-time implementation. Future research efforts will prioritize collaboration with handheld ultrasound manufacturers to enable a comprehensive evaluation that includes the model's impact on physician skill development in a real-world setting.

Second, the single-center design with a limited sample size restricts the generalizability of our findings. External validation on a larger, multicenter cohort is warranted to confirm the model's efficacy in real-world clinical practice.

Third, our dataset size is relatively modest compared to other deep-learning studies. In addition, the annotations exclusively performed by three experts within a single subspecialty might

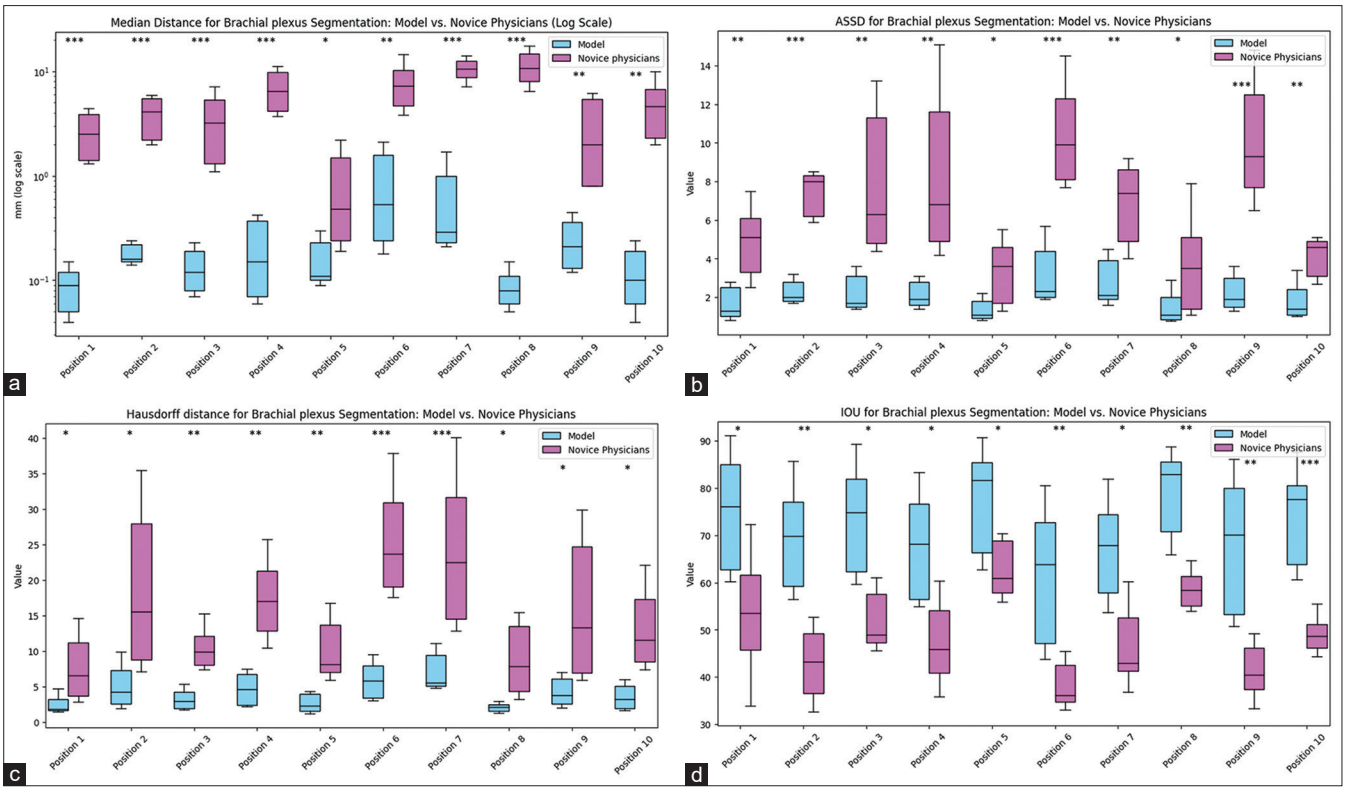

Figure 10: Comparative evaluation of model and novice physician segmentation performance across multiple metrics. The figure illustrates the comparative performance of an artificial intelligence (AI) model and novice physicians in segmenting the brachial plexus across ten different anatomical positions. Four distinct metrics were employed for evaluation: (a) Median distance (Log Scale): Demonstrates the median distance (in mm) between the predicted and ground truth segmentation. Notably, the AI model consistently outperformed novice physicians across all positions, with median distances often <0.5mm. A logarithmic scale was used due to the substantial difference in performance magnitudes between the two groups. (b) Average symmetric surface distance (ASSD): Measures the average distance between the surfaces of the predicted and ground truth segmentations. The AI model consistently exhibited lower ASSD values compared to novice physicians, indicating superior surface conformity. (c) Hausdorff Distance: Quantifies the maximum distance between the surfaces of the predicted and ground truth segmentations. Similar to ASSD, the AI model displayed lower Hausdorff distances than novice physicians, signifying greater overall segmentation accuracy. (d) Intersection over Union (IoU): Assesses the overlap between the predicted and ground truth segmentations. The AI model consistently achieved higher IoU scores compared to novice physicians, reflecting a more precise delineation of the brachial plexus. One asterisk (*) typically indicates $P < 0.05$; Two asterisks (**) indicate $P < 0.01$; Three asterisks (***) indicate $P < 0.001$

introduce bias and limit the model's ability to generalize to anatomical variations. Future studies could benefit from collaboration with multidisciplinary teams to ensure a more comprehensive and objective labeling process.

Conclusion

This study successfully developed a novel image segmentation algorithm for brachial plexus ultrasound images. The model achieved high accuracy in image classification and significantly outperformed novice physicians in landmark segmentation metrics. This performance suggests promise for the model's role in educational settings, particularly for aiding novice physicians in learning brachial plexus anatomy.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Lapegue F, Faruch-Bilfeld M, Demondion X, Apredoaei C, Bayol MA, Artico H, *et al.* Ultrasonography of the brachial plexus, normal appearance and practical applications. Diagn Interv Imaging 2014;95:259-75.
2. Hsu PC, Chang KV, Mezian K, Naňka O, Wu WT, Yang YC, *et al.* Sonographic pearls for imaging the brachial plexus and its pathologies. Diagnostics (Basel) 2020;10:324.
3. Gu S, Zhao Q, Yao J, Zhang L, Xu L, Chen W, *et al.* Diagnostic ability of ultrasonography in brachial plexus root injury at different stages post-trauma. Ultrasound Med Biol 2022;48:1122-30.
4. Simoni P, Ghassemi M, Le VD, Boitsios G. Ultrasound of the normal brachial plexus. J Belg Soc Radiol 2017;101:20.
5. Yang XY, Wang LT, Li GD, Yu ZK, Li DL, Guan QL, *et al.* Artificial intelligence using deep neural network learning for automatic location of the interscalene brachial plexus in ultrasound images. Eur J Anaesthesiol 2022;39:758-65.
6. Liu C, Liu F, Wang L, Ma L, Lu ZM. Segmentation of nerve on ultrasound images using deep adversarial network. Int J Innov Comput Inform Control 2018;14:53-64.
7. van Boxtel J, Vousten VR, PluimJ, Rad NM. Hybrid deep neural network for brachial plexus nerve segmentation in ultrasound images. arXiv Preprint arXiv 2106.00373, 2021. 2021 29th European Signal Processing Conference (EUSIPCO), p. 1246-50. Available from: https://

doi.org/10.48550/arXiv.2106.00373.
8. Wang Y, Geng J, Zhou C, Zhang Y. Segmentation of ultrasound brachial plexus based on U-Net. In 2021 International Conference on Communications, Information System and Computer Engineering (CISCE), IEEE; 2021. p. 482-5. Available from: https://ieeexplore.ieee.org/document/9445925 [Last accessed on 2024 Sep 22].
9. Wang Y, Zhu B, Kong L, Wang J, Gao B, Wang J, *et al.* BPSegSys: A brachial plexus nerve trunk segmentation system using deep learning. Ultrasound Med Biol 2024;50:374-83.
10. Xi Y, Chong H, Zhou Y, Zhu F, Yao Y, Wang G. Convolutional neural network for brachial plexus segmentation at the interscalene level. BMC Anesthesiol 2024;24:17.