



OPEN Masked autoencoder pretraining for peripheral nerve segmentation in ultrasound images

Matthew Webster^{1,2,6}, Ko Eun Kim^{3,6}, Jaewon Lim⁴, Joonnyong Lee^{2,6}✉ & Beom Suk Kim^{4,5,6}✉

Despite recent advances in medical image segmentation, deep learning applications in musculoskeletal ultrasound remain limited by small labeled datasets and a narrow clinical focus. Most existing work centers on the median nerve, leaving other clinically relevant peripheral nerves understudied. This study explores masked autoencoder (MAE) pretraining as a scalable approach to overcome data scarcity in peripheral nerve segmentation. A dataset of 10,603 ultrasound images across eight nerves was constructed, spanning over 1,500 patients with normal nerve conduction studies. A SegFormer-based encoder was pretrained on unlabeled ultrasound images using MAE and subsequently fine-tuned on nerve-specific segmentation tasks. Pretraining was applied per cross-validation fold to prevent data leakage. Compared to randomly initialized baselines, MAE-pretrained models achieved consistent performance gains across all nerves. Improvements were most pronounced for low-resource targets such as the sural (0.144 → 0.399 Dice) and lateral femoral cutaneous (0.280 → 0.550 Dice) nerves. Nerves with more abundant training data, like the median (0.890 → 0.897 Dice) and ulnar (0.862 → 0.869 Dice), showed only marginal improvements. These findings indicate that MAE pretraining particularly strengthens feature extraction for small, difficult-to-label structures in ultrasound segmentation. This underscores the potential of self-supervised learning to ease annotation demands and support the development of more comprehensive clinical AI tools.

Keywords Ultrasound, Peripheral nerves, Segmentation, Deep learning

Ultrasonography has emerged as a highly valuable diagnostic tool for evaluating peripheral nerves. Through ultrasound examination, clinicians can observe the course, morphology, and echogenicity of peripheral nerves and determine their spatial relationships with surrounding anatomical structures. Furthermore, ultrasound enhances procedural safety by providing real-time guidance during interventional procedures^{1,2}. Ultrasound is often preferred as a first-line imaging modality for peripheral nerve assessment because it is inexpensive, widely accessible, and provides real-time, high-resolution visualization of superficial nerves at the point of care. Unlike static modalities, ultrasound enables dynamic evaluation (e.g., nerve gliding and positional changes) and readily supports image-guided procedures. While MRI offers complementary information—particularly for deep lesions or proximal pathology—it is generally more resource-intensive and less suited for rapid bedside assessment. Therefore, accurate ultrasound-based nerve delineation is clinically valuable for standardizing measurements and facilitating scalable computer-aided assessment. In particular, standardized delineation is essential for reproducible quantification (e.g., cross-sectional area) and for reducing observer-dependent variability in routine ultrasound practice. However, several limitations have been identified in the ultrasound evaluation of peripheral nerves. First, the examination process is labor-intensive, as the segmentation of peripheral nerves and the measurement of parameters require manual operations, resulting in considerable time and effort expenditure. Second, accurate recognition of peripheral nerves on ultrasound requires a relatively steep learning curve. The appearance of peripheral nerves can vary significantly depending on the angle, position, and pressure of ultrasound probe application, and nerves often resemble tendons or blood vessels.

¹Chung-Ang University Industry-Academic Cooperation Foundation, Seoul, South Korea. ²Mellowing Factory Co. Ltd., Seoul, South Korea. ³Department of Dermatology, Korea University Guro Hospital, Seoul, Republic of Korea. ⁴Chung-Ang University College of Medicine, Seoul, Republic of Korea. ⁵Department of Physical and Rehabilitation Medicine, Chung-Ang University Gwangmyeong Hospital, Gwangmyeong-si, Gyeonggi-do, Republic of Korea. ⁶Matthew Webster and Ko Eun Kim contributed equally to this work as first authors. Beom Suk Kim and Joonnyong Lee contributed equally to this work as corresponding authors. ✉email: joon@mellowingfactory.com; grit@cauhs.or.kr

Given these challenges, there has been an unmet need for automated systems that can segment peripheral nerves and measure parameters on ultrasound images. Because nerve recognition relies heavily on clinical experience, expertise, and intuition, peripheral nerve ultrasound has been identified as a promising application area for artificial intelligence^{3–5}.

Deep learning has led to substantial progress in medical image segmentation, particularly through the use of convolutional encoder-decoder architectures such as U-Net⁶ and its variants^{7–9}. These models have demonstrated strong performance across a variety of imaging modalities, including ultrasound, by learning hierarchical features that enable accurate delineation of anatomical structures even in the presence of noise, low contrast, and imaging artifacts¹⁰. Segmentation in ultrasound has mostly focused on large structures or major clinical targets, with relatively little attention given to peripheral nerves. Generally, acquiring labeled ultrasound data targeting peripheral nerves is both time-consuming and resource-intensive, requiring extensive training and expertise. Previous studies primarily targeted the median (and occasionally femoral) nerve using small datasets^{11–13}, whereas segmentation beyond the median nerve—including the common peroneal, lateral femoral cutaneous, sciatic, tibial, radial, ulnar, and sural nerves—has been less systematically investigated. Small-caliber sensory nerves, such as the sural and lateral femoral cutaneous nerves, are clinically relevant targets (e.g., metabolic peripheral neuropathies and meralgia paresthetica), yet they are technically challenging to visualize on ultrasound due to their small caliber, variable course, and low contrast. Establishing robust segmentation for these underrepresented nerves is essential for extending ultrasound-based assessment beyond a single, relatively straightforward target nerve.

One of the most effective strategies for addressing limited labeled data is through model pretraining¹⁴. Traditional transfer learning approaches, such as initializing from networks trained on natural image datasets like ImageNet, can improve convergence and performance. However, the domain shift between natural and medical images often limits the effectiveness of such methods¹⁵. Self-supervised learning (SSL) has emerged as a promising alternative, allowing models to learn useful representations directly from unlabeled medical images. Among SSL methods, masked autoencoder (MAE) pretraining has shown particular promise. MAEs train models to reconstruct missing patches in an image, encouraging the network to learn semantic and spatial structure without supervision¹⁶. In ultrasound, recent studies have demonstrated that MAE-pretrained models can outperform supervised pretraining baselines in both classification and segmentation tasks, particularly when downstream labeled data is scarce¹⁷.

Despite these advances, two key gaps remain for peripheral nerve ultrasound segmentation. First, most existing studies are constrained to a single nerve—typically the median nerve—leaving it uncertain whether learned representations generalize to anatomically diverse and clinically important nerves in both the upper and lower extremities. Second, the annotation burden is especially prohibitive for small-caliber nerves with variable courses and low contrast, where even expert delineation can be inconsistent. Accordingly, a systematic evaluation of MAE pretraining for segmentation across multiple peripheral nerve types is still lacking. Addressing these gaps may facilitate more consistent measurement and annotation workflows in routine practice, which is a prerequisite for broader clinical translation.

Therefore, this study aims to evaluate whether MAE pretraining improves SegFormer-based ultrasound segmentation across a broad set of clinically important peripheral nerves in both upper and lower extremities. Using a large-scale dataset spanning eight nerves, we compare MAE-pretrained models with randomly initialized baselines under patient-level five-fold cross-validation, with fold-specific pretraining to prevent data leakage. To the best of our knowledge, this is among the first studies to systematically assess MAE pretraining for ultrasound peripheral nerve segmentation across multiple nerve types beyond the median nerve. While MAE as a self-supervised pretraining paradigm is not new, its leakage-aware evaluation for ultrasound segmentation across multiple peripheral nerve types—particularly low-resource, small-caliber nerves—has remained limited. Experiments were conducted exclusively on nerves with normal nerve conduction study findings.

Methods

Background

Deep learning has driven advances in semantic segmentation by enabling models to learn rich hierarchical features from images. SegFormer is a recent adaptation of this paradigm that combines a lightweight decoder with a hierarchical Transformer encoder to produce U-Net-style outputs⁸. Its encoder applies overlapping patch merging to generate multi-scale feature maps, and an efficient self-attention mechanism captures long-range spatial dependencies without requiring explicit positional embeddings. The resulting architecture maintains high segmentation accuracy while reducing model complexity, and initial evaluations in medical imaging have shown that SegFormer can match or exceed the performance of standard U-Net variants on tasks such as cardiac and endoscopic image segmentation¹⁸.

Self-supervised learning offers a strategy to leverage large volumes of unlabeled data when annotated datasets are scarce. Masked autoencoders exemplify this approach by randomly removing a large fraction of image patches and training the model to reconstruct the missing regions. Because the encoder only processes the visible patches and the decoder is lightweight, MAE training scales efficiently to high-resolution inputs. Studies have demonstrated that high masking rates (e.g., 75%) encourage the model to learn global and semantic image structure, which in turn leads to strong transfer learning performance¹⁶. Furthermore, MAEs exhibit robustness to variations in patch size, although excessively large patches can degrade reconstruction quality and downstream accuracy by obscuring finer details.

After self-supervised pretraining, fine-tuning adapts the entire model to a specific downstream task by continuing gradient-based training on labeled data. While fine-tuning has more commonly focused on classification, recent work has extended it to segmentation in medical imaging¹⁹. For example, MAE-pretrained vision transformers have been fine-tuned for both 2D and 3D medical image segmentation, yielding substantial

improvements in delineating anatomical structures from modalities such as chest X-ray, CT, MRI and ultrasound^{20,21}. This two-stage process, consisting of self-supervised pretraining followed by task-specific fine-tuning, provides a practical route to harness large unlabeled datasets and improve performance on clinically important segmentation tasks where labeled data are limited.

Datasets

Ethical standards

This study was approved by the Institutional Review Boards of Korea University Guro Hospital (IRB No. 2020GR0389) and Chung-Ang University Gwangmyeong Hospital (IRB No. 2511-284-179). Patient consent was waived because the study used de-identified ultrasound images, and both IRBs determined that obtaining individual consent was not required.

Study subjects

Electronic medical records and picture archiving and communication system databases were systematically reviewed to identify individuals who received both electrodiagnostic studies and ultrasound evaluation of upper and lower extremity peripheral nerves between March 2019 and February 2024. From an initial cohort of 1,782 patients, a subset of 1,362 participants (763 males, 599 females; mean age 52 ± 7 years [SD]; age range 23–78 years) was selected based on normal nerve conduction test results for subsequent analysis. The total patient and image counts for each nerve are reported in Table 1.

Data collection procedures

All ultrasound examinations were performed using a standardized protocol during the initial assessment period. A single physiatrist with over fifteen years of experience in neuromuscular ultrasonography conducted all scanning procedures and image annotations to ensure consistency. Image annotation was performed using the built-in nerve tracing functionality within the ultrasound operating application of the miniSONO system.

During the examination, participants were positioned either supine or prone, depending on the peripheral nerve being evaluated. To ensure accurate nerve identification, the examiner employed dynamic scanning techniques, systematically moving the transducer along the nerve pathway in both proximal and distal directions. Peripheral nerve visualization was achieved using a miniSONO device (ALPINION Medical Systems, Korea).

Because peripheral nerve ultrasonography is highly operator-dependent and accurate delineation often relies on real-time dynamic scanning context (e.g., probe angulation/pressure adjustments and continuous visualization along the nerve course), annotations were created contemporaneously with image acquisition by the same expert operator. Specifically, after selecting and saving the representative still frame during the examination, the operator immediately delineated the nerve boundary using the in-app drawing tool, and the corresponding contour/ROI was stored together with the image to construct paired image–annotation sets. This workflow was adopted to minimize pairing errors and reduce ambiguity when interpreting isolated static images after the examination. As all acquisitions and annotations were performed by a single experienced specialist, no additional second-reader adjudication, consensus review, or separate quality-control rounds were applied.

Data processing

All ultrasound images underwent a standardized preprocessing pipeline to ensure consistency across both the pretraining and fine-tuning stages. During training, each image was cropped or zero-padded to a fixed resolution of 512 × 384 pixels after data augmentation. At test time, the original full-resolution image was used to preserve as much information as possible, with additional padding applied when necessary to maintain compatibility with SegFormer’s encoder-decoder structure and its associated downsampling and upsampling operations.

To enhance model generalizability and robustness, a consistent set of data augmentations was applied during both pretraining and fine-tuning. These included random horizontal flipping with a 50% probability, random resizing within ± 25% of the original scale, random cropping to the target input size, random rotation within ± 10 degrees, and random shearing within ± 10 degrees. These augmentations were selected to reflect the real-world

Nerve	No. Images	No. Patients	Avg. No. Training Images per Fold	Avg. No. Testing Images per Fold
Median	3,811	413	3048.8	762.2
Ulnar	4,729	545	3783.2	945.8
Common Peroneal	412	143	329.6	82.4
Sciatic	99	35	79.2	19.8
Tibial	219	80	175.2	43.8
Radial	208	55	166.4	41.6
Lateral Femoral Cutaneous	150	64	120.0	30.0
Sural	76	27	60.8	15.2
Unlabeled*	1,311	368	No fold, pretraining only.	No fold, pretraining only.

Table 1. Number of patients and images available for each peripheral nerve in the dataset studied. *Unlabeled images are only used for the pretraining stage and not during finetuning. Data folds are maintained across both pretraining and finetuning.

variability encountered in ultrasound acquisition, including changes in transducer angle, patient positioning, and tissue deformation.

Masked autoencoder pretraining with SegFormer

A MAE pretraining strategy using the SegFormer architecture as the backbone model is adopted for this study. SegFormer is well-suited for this purpose due to its strong performance on segmentation tasks, including those in medical imaging. Although most prior studies on masked autoencoders use Vision Transformer (ViT) models as the encoder^{22,23}, ViT-based autoencoders were found to perform poorly on the ultrasound data. In preliminary experiments, these models produced suboptimal reconstructions during pretraining and showed weak performance when fine-tuned for downstream segmentation. Transformer-based architectures often require lots of data, and due to the limited size of our dataset, ViT models were found to be ineffectual in generalization. In contrast, SegFormer yielded both improved reconstruction quality and stronger segmentation accuracy across all nerve datasets.

A limitation of using SegFormer for masked autoencoding is that the encoder cannot easily exclude masked tokens from its computation. In typical MAE training with ViT, masked tokens are removed before encoding, which substantially reduces memory consumption and accelerates training. SegFormer does not offer a straightforward mechanism to omit masked spatial regions, leading to higher memory requirements and slower training. Despite this drawback, the improved quality of the learned features justified the added computational cost. Further, as previously mentioned, the Segformer architecture has specific design choices that make generally more computationally efficient than standard ViT models. Importantly, even in standard MAE workflows, inference time remains unaffected by masked token removal, as the full image is processed during downstream tasks regardless of masking.

MAE pretraining is applied separately for each cross-validation fold. Specifically, five cross-validation folds are defined for each of the eight nerves in the dataset. These folds are grouped such that each of the five groups contains one fold from each nerve. This grouping allows for pretraining five distinct models, one per group, ensuring that the pretraining data remains strictly disjoint from the data used for evaluation in the corresponding downstream tasks. This approach is necessary to prevent data leakage during fine-tuning. Additionally, the pretraining set is augmented with 1,311 unlabeled ultrasound images (368 individuals) not associated with any downstream nerve labels or patients. The unlabeled data does include pathological cases but no such abnormal cases are included during the finetuning phase.

During fine-tuning, each pretrained model is adapted to segment a single nerve. This means that for each nerve, five models are trained using the five corresponding pretrained backbones. Further, there is a data imbalance of the nerve types during pretraining, but not during finetuning. Training separate segmentation models per nerve enables evaluation of how well pretraining on a broad, unlabeled dataset generalizes to specific downstream targets. It also highlights which nerves benefit most from shared pretraining across different anatomical structures. The entire pipeline from MAE pretraining to fine-tuning on the segmentation task is illustrated in Fig. 1.

Experiments

Implementation and data splits

PyTorch 2.7.1 was used on an Nvidia A6000 GPU, and all models were built with the segmentation-models-pytorch library²⁴. All segmentation experiments employed five-fold patient-level cross-validation. Each fold was fixed across both baseline and MAE-pretrained runs to avoid leakage.

Outcome variables

In this study, the performance was verified using two common segmentation metrics with the literature: Dice score and intersection over union (IoU).

$$(Dice) = \frac{2 \cdot |X \cap Y|}{|X| + |Y|} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (1)$$

$$(IoU) = \frac{|X \cap Y|}{|X \cup Y|} = \frac{TP}{TP + FP + FN} \quad (2)$$

where X and Y represent the predicted and ground truth segmentation masks, respectively. Further, TP is the true positive rate, FP is the false positive rate, and FN is the false negative rate.

Baseline segmentation

As a reference, SegFormer MiT-B3 models use random initialization for each nerve. MiT-B3 was chosen after initial tests showed it offered the best trade-off between accuracy and cost. Each model was trained for 150 epochs with Adam (learning rate starts at 1×10^{-4} with a decay of $\times 0.1$ every 50 epochs and weight decay is set at 1×10^{-7}). Additionally a batch size of 32 is used during training, and optimized with an averaged Dice and cross-entropy loss. Dice score was the primary metric. Further we also compare out results to the U-Net, U-Net++, and DeepLabV3 architectures using the same training setup. The computational costs and parameter counts can be seen in Table 2.

MAE pretraining

A SegFormer MiT-B3 model was pretrained as a masked autoencoder with a patch size of 32×32 and a 65% masking ratio (these hyperparameters were found via simple preliminary testing runs and not thorough

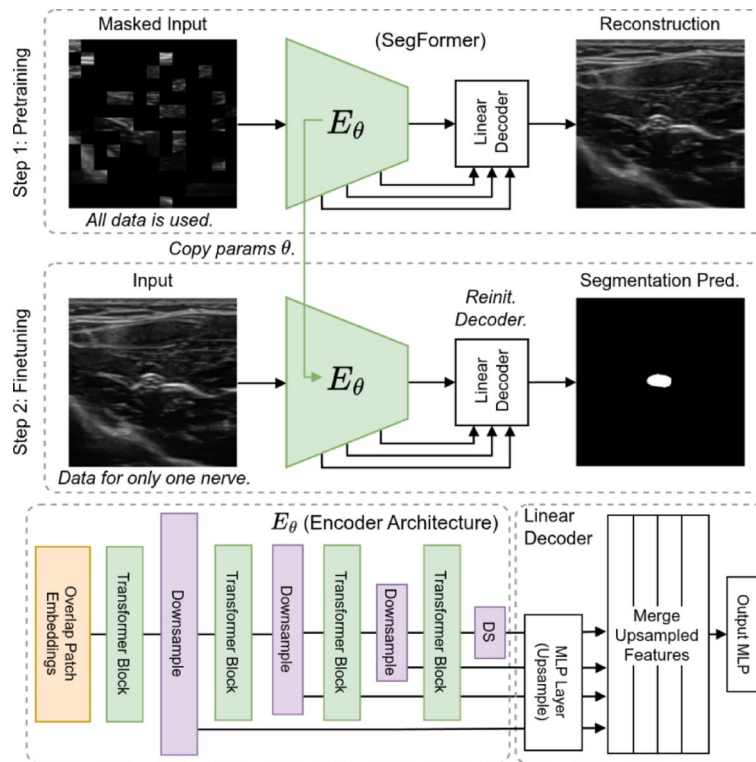


Fig. 1. Visualization of the masked autoencoder pretraining and fine-tuning pipeline for peripheral nerve segmentation. Weights θ , are copied from the encoder E after masked autoencoder pretraining as a starting place for fine-tuning on the nerve segmentation task.

	U-Net	U-Net++	DeepLabV3	Segformer
Parameters	24 M	26 M	26 M	45 M
MACs	31G	73G	109G	36G

Table 2. Parameter counts (in millions) and computational cost in MACs (in giga) for each model architecture tested.

ablation). Figure 2 shows a mask sample and prediction sample made by one of the trained MAE models. The model learned to reconstruct masked regions via an L1 loss against the original ultrasound image. Training ran for 400 epochs using the Adam optimizer (learning rate starts at 2×10^{-4} with a decay of $\times 0.1$ every 100 epochs and weight decay is set to 1×10^{-6}). As before, a batch size of 32 is used.

Fine-tuning segmentation

For each nerve, the MAE-pretrained encoder finetuned with a reinitialized SegFormer decoder to prevent convergence failures. Fine-tuning used the same hyperparameters and loss combination as the baseline (150 epochs, Adam with $\text{lr} = 1 \times 10^{-4}$, decay $\times 0.1$ every 50 epochs, weight decay $= 1 \times 10^{-7}$, batch size 32, Dice + cross-entropy). Keeping all settings constant ensures a fair comparison of baseline models and MAE-pretrained models.

Results

Table 3 shows that MAE pretraining consistently improves segmentation accuracy across all eight peripheral nerves compared to the non-pretrained baseline. Additionally we have included results from the U-Net, Unet++, and DeepLabV3 architectures. As an additional point of reference we have also included the IoU results in Table 4. The most dramatic gains are observed for the lateral femoral cutaneous nerve, nearly doubling in Dice score from 0.28 ± 0.08 to 0.55 ± 0.091 , and the sural nerve increasing from 0.144 ± 0.043 to 0.399 ± 0.197 . Other substantial improvements occur for the radial nerve ($0.429 \pm 0.071 \rightarrow 0.648 \pm 0.041$), sciatic nerve ($0.513 \pm 0.053 \rightarrow 0.725 \pm 0.051$), and tibial nerve ($0.563 \pm 0.038 \rightarrow 0.668 \pm 0.049$). The common peroneal nerve has a less substantial increase from 0.737 ± 0.026 to 0.771 ± 0.032 and did not achieve a p-value of less than 0 over the cross-fold validation (Table 5). Prediction samples for each nerve are shown in Fig. 3.

For the median and ulnar nerves, MAE pretraining yields slight gains over the strong baseline performance. The Dice score for the median nerve improves from 0.89 ± 0.006 to approximately 0.90 ± 0.004 , and the ulnar

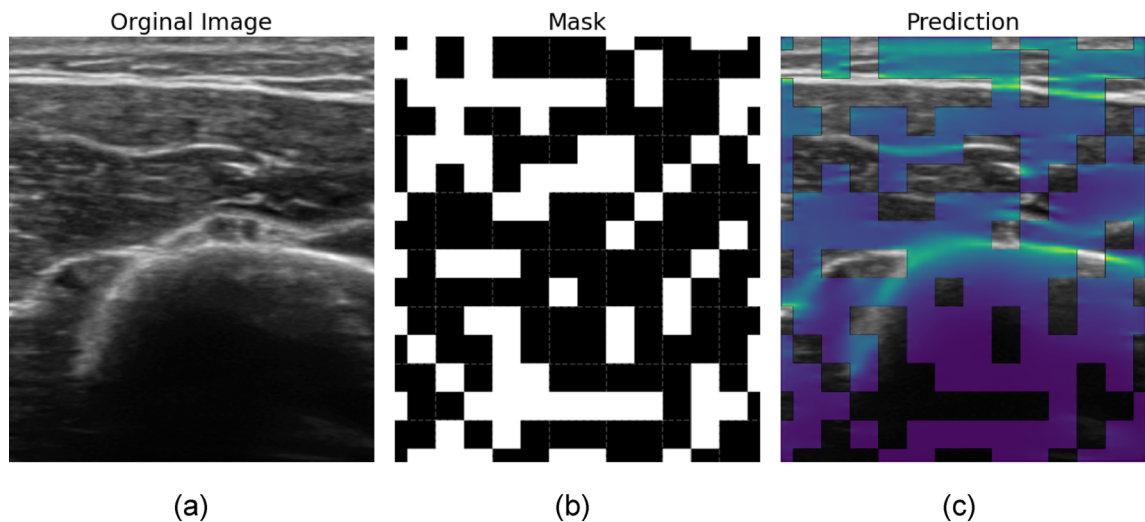


Fig. 2. (a) shows the original image before masking, (b) shows the mask applied to the image (white represents the area removed), and (c) shows the predictions with the unmasked model input, (b) overlaid. Note that the padding used during training and inference is not represented here and causes the patches to appear uneven at the edges.

Nerve	Baseline U-Net Dice Score ↑	Baseline U-net++ Dice Score ↑	Baseline DeepLabV3 Dice Score ↑	Baseline Segformer Dice Score ↑	Fine-tuned Segformer Dice Score ↑
Median	0.894 ± 0.004	0.896 ± 0.007	0.890 ± 0.007	0.890 ± 0.006	0.897 ± 0.004
Ulnar	0.866 ± 0.004	0.865 ± 0.004	0.868 ± 0.005	0.862 ± 0.007	0.869 ± 0.005
Common Peroneal	0.731 ± 0.013	0.728 ± 0.025	0.708 ± 0.033	0.737 ± 0.026	0.771* ± 0.032
Sciatic	0.636 ± 0.102	0.625 ± 0.068	0.587 ± 0.064	0.513 ± 0.053	0.725* ± 0.051
Tibial	0.629 ± 0.052	0.630 ± 0.076	0.613 ± 0.060	0.563 ± 0.038	0.668* ± 0.049
Radial	0.286 ± 0.203	0.497 ± 0.088	0.526 ± 0.047	0.429 ± 0.071	0.648* ± 0.041
Lateral Femoral Cutaneous	0.195 ± 0.182	0.372 ± 0.063	0.388 ± 0.073	0.280 ± 0.080	0.550* ± 0.091
Sural	0.32 ± 0.193	0.290 ± 0.248	0.290 ± 0.115	0.144 ± 0.043	0.399* ± 0.197

Table 3. Test result dice scores for each nerve using MAE-pretrained models, compared to a non-pretrained baseline. Test results are averaged over five cross-validation folds. * indicates a p-value < 0.05 compared to Segformer baseline using a relative t-test.

Nerve	Baseline U-Net IoU Score ↑	Baseline U-net++ IoU Score ↑	Baseline DeepLabV3 IoU Score ↑	Baseline Segformer IoU Score ↑	Fine-tuned Segformer IoU Score ↑
Median	0.824 ± 0.002	0.827 ± 0.005	0.818 ± 0.006	0.818 ± 0.006	0.827 ± 0.005
Ulnar	0.792 ± 0.004	0.791 ± 0.006	0.791 ± 0.005	0.785 ± 0.007	0.793* ± 0.006
Common Peroneal	0.619 ± 0.014	0.622 ± 0.025	0.594 ± 0.029	0.618 ± 0.027	0.659 ± 0.033
Sciatic	0.510 ± 0.103	0.501 ± 0.067	0.471 ± 0.056	0.380 ± 0.049	0.608* ± 0.054
Tibial	0.522 ± 0.053	0.520 ± 0.079	0.495 ± 0.057	0.446 ± 0.032	0.563* ± 0.049
Radial	0.205 ± 0.147	0.395 ± 0.091	0.416 ± 0.043	0.323 ± 0.069	0.544* ± 0.044
Lateral Femoral Cutaneous	0.235 ± 0.147	0.267 ± 0.048	0.279 ± 0.056	0.191 ± 0.065	0.435* ± 0.083
Sural	0.137 ± 0.132	0.210 ± 0.188	0.191 ± 0.086	0.085 ± 0.027	0.306* ± 0.171

Table 4. Test result IoU scores for each nerve using MAE-pretrained models, compared to a non-pretrained baseline. Test results are averaged over five cross-validation folds. * indicates a p-value < 0.05 compared to Segformer baseline using a relative t-test.

	Median	Ulnar	Common Peroneal	Sciatic	Tibial	Radial	Lateral Femoral Cutaneous	Sural
Improvement in Dice (%)	1.1%	1.2%	4.1%	41.2%	19.6%	51.2%	96.4%	185.7%

Table 5. Improvement in Dice of the proposed MAE approach over the Segformer baseline model.

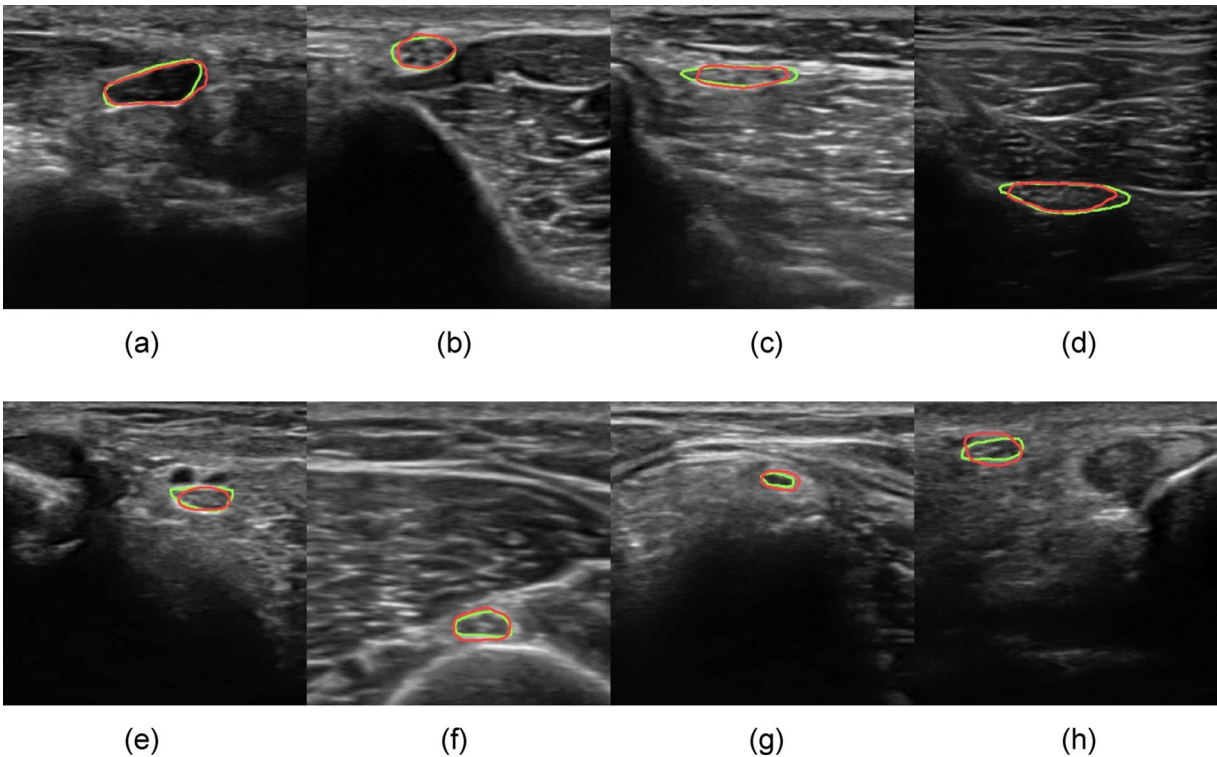


Fig. 3. Randomly sampled segmentation predictions close to the median Dice score for each nerve (less than 0.01 difference to the median). The predictions are made by the MAE-pretrained SegFormer model for each of the eight nerves types. From left to right, top to bottom, the nerves shown are the (a) median (0.93 Dice), (b) ulnar (0.93 Dice), (c) common peroneal (0.87 Dice), (d) sciatic (0.88 Dice), (e) tibial (0.86 Dice), (f) radial (0.84 Dice), (g) lateral femoral cutaneous (0.72 Dice), and (h) sural (0.74 Dice). Predicted segmentations are shown in red and ground truth labels are shown in green.

nerve increases from 0.86 ± 0.007 to approximately 0.87 ± 0.005 , however the ulnar nerve did not achieve a p-value less than 0.05. These results indicate that self-supervised pretraining is most beneficial for nerves where the baseline model struggles, while it still provides modest gains for already well performing nerves (Table 5).

Discussion

This study investigated the use of MAE pretraining to improve segmentation of peripheral nerves in ultrasound images. This method was applied to a dataset spanning eight clinically important nerves of the upper and lower extremities, with segmentation performance evaluated under a five-fold cross-validation framework. MAE-pretrained SegFormer models consistently outperformed non-pretrained baselines, especially in cases where baseline performance was weak due to limited labeled data.

Several recent studies have attempted to apply artificial intelligence to peripheral nerve ultrasound imaging; however, these studies have been limited by small sample sizes and a narrow scope of target nerves. Most studies have focused solely on the median nerve, lacked standardized ultrasound examination protocols, and rarely incorporated nerve conduction study data. Most importantly, there has been an insufficient number of studies validating the accuracy of artificial intelligence-based segmentation for each of the clinically important peripheral nerves of the upper and lower extremities^{25–33}. In the upper extremity, the ulnar nerve, which causes cubital tunnel syndrome (the second most common compressive neuropathy after carpal tunnel syndrome), and the radial nerve, which causes Saturday night palsy (wrist drop after arm support), are clinically significant nerves. In the lower extremity, the sciatic nerve, which is the largest nerve in the lower limb, and its two major branches, the peroneal nerve and tibial nerve, are also of clinical interest. Additionally, the lateral femoral cutaneous nerve, which causes meralgia paresthetica, and the sural nerve, a sensory nerve commonly affected in metabolic peripheral neuropathies, should not be overlooked in terms of their clinical importance^{34–36}. In contrast, this study includes a broad set of nerves, encompassing both commonly and less commonly studied types. These

include the median, ulnar, radial, tibial, sciatic, sural, common peroneal, and lateral femoral cutaneous nerves. The dataset employed in this study contains over 10,000 images from more than 1,300 patients, representing one of the largest and most diverse peripheral nerve ultrasound segmentation datasets to date.

The results in Table 3 demonstrate that MAE pretraining yields consistent and, in many cases, substantial improvements in segmentation accuracy across multiple peripheral nerves. Nerves with low baseline performance, most notably the lateral femoral cutaneous (LFC) and sural nerves, experienced the greatest gains, improving by 0.27 and 0.26 points in Dice score, respectively. The radial and sciatic nerves also saw large increases of 0.22 and 0.21 points, while the common peroneal and tibial nerves improved by 0.03 and 0.11. Despite these improvements, the sural, radial, and LFC nerves still exhibited relatively low Dice scores after MAE pretraining. This is likely due to their smaller diameters, more complex anatomical paths, and greater inter-individual variability compared to other nerves. When the ultrasound beam is oriented perpendicularly to the nerve's trajectory, these structures appear clearly and are easier to segment. However, small, winding nerves are technically challenging to capture in such ideal orientations, making even expert-generated annotations less accurate and reliable. In contrast to the underrepresented nerves, the median and ulnar nerves showed only modest gains of approximately 0.01 each and only the median nerve had a statistically significant improvement in Dice score. This limited improvement likely reflects the fact that these nerves already had relatively large labeled datasets, enabling high baseline performance that left little room for additional benefit from self-supervised pretraining.

Technically, this study differs from most prior MAE applications by using the SegFormer backbone rather than ViT. While this choice led to higher memory usage it also enabled more robust feature learning suited to ultrasound image characteristics. This architectural decision contributed meaningfully to the improved performance across nerve types. Clinically, the pretraining approach renders segmentation of smaller or more challenging nerves more viable. For example, the sural nerve improved from 0.144 to 0.399 Dice score, and the lateral femoral cutaneous nerve from 0.280 to 0.550; these levels may be sufficient to assist in annotation workflows, whereas their baseline-method performance below 0.30 would typically be considered too low for any practical use. These results demonstrate that MAE pretraining is a particularly compelling option when attempting to segment peripheral nerves with limited available annotations.

From a clinical workflow perspective, reliable nerve delineation is a prerequisite for standardized quantitative ultrasound assessment, enabling consistent region-based measurements (e.g., cross-sectional area) and longitudinal comparisons across visits. Even when pixel-level segmentation is not used as a fully automated endpoint, providing a reasonable initial contour can reduce the time required for manual tracing and improve inter-operator consistency through expert review and correction (human-in-the-loop). This is particularly relevant for small-caliber nerves with variable courses and low contrast, where boundary tracing is labor-intensive and prone to variability. In this sense, the practical value of MAE pretraining in our study is best interpreted as strengthening reproducible measurement and efficient annotation workflows, rather than implying immediate standalone automation.

Nonetheless, the segmentation performance of all but the ulnar and median nerves still falls below generally acceptable performance thresholds for deployment in computer aided diagnostic (CAD) workflows^{37,38}, and further work, such as acquiring more labeled images or refining network architectures, will be needed to raise its performance. Further, the median and ulnar nerves, which already performed strongly without pretraining, remain robust but did not cross any new operational thresholds. This outcome highlights the continued importance of collecting additional labeled data, refining architectures, or incorporating complementary techniques, such as uncertainty modeling³⁹, for such difficult targets.

This study has several important limitations. First, pretraining SegFormer as a masked autoencoder is computationally intensive. Due to the inability to exclude masked tokens during encoding, memory requirements are significantly higher than in ViT-based MAEs. This constraint limits model scalability and may be a barrier to adoption in resource-constrained settings. Second, although over 1,300 additional unlabeled ultrasound images were incorporated totaling to 10,603 images, the overall dataset size remains relatively small compared to large-scale vision datasets. This limitation likely constrains the maximal benefit that self-supervised learning can confer. Third, all ultrasound images were acquired using a single scanner model under standardized conditions. While this design ensured consistency, it may limit the generalizability of the results to other ultrasound systems or clinical environments. Future work should explicitly test robustness to variations in equipment, operator, and acquisition protocol. Fourth limitation is the potential for domain shift in ultrasound. Although all images were acquired and annotated by a single experienced operator, which improves internal consistency, performance may differ when applied to images from other operators, scanning styles, or ultrasound vendors due to variations in probe handling, imaging presets, and vendor-specific processing. Future studies should include multi-operator, multi-vendor datasets and external validation, and may consider additional approaches such as targeted fine-tuning or domain adaptation to enhance generalizability. Finally, the dataset includes only nerves from patients with normal nerve conduction studies. This restriction limits the clinical relevance of these models for diagnostic tasks and restricts its utility to annotation or measurement workflows in normal-appearing nerves. Evaluating performance on abnormal nerves remains an essential next step.

Future studies should include peripheral nerves with pathological findings to evaluate whether the models generalize to abnormal nerve presentations. In addition, expanding the dataset to include more patients and more nerve types, particularly in underrepresented anatomical regions, would improve robustness and external validity. Exploring more advanced backbone architectures, additional pretraining approaches within the literature, or more in-depth analysis of loss design could further enhance segmentation accuracy. Finally, MAE-pretrained models may prove useful in semi-supervised learning or human-in-the-loop annotation frameworks where partially labeled data are available.

Conclusion

This study presents a comprehensive investigation of MAE pretraining for peripheral nerve segmentation in ultrasound images. By leveraging a large, diverse dataset across eight clinically important nerves, this study demonstrates that MAE pretraining substantially improves segmentation accuracy, especially for small or underrepresented nerves such as the sural (0.144 → 0.399 Dice) and lateral femoral cutaneous (0.280 → 0.550 Dice) nerves. While this results on the underrepresented nerves do not enable fully autonomous segmentation, these results may still be useful for human-in-the-loop annotation tasks. The approach explored in this study enables effective transfer learning in settings with limited annotations and offers a practical path forward for enhancing peripheral nerve ultrasound analysis. While challenges remain, particularly in model generalizability and performance on abnormal nerves, the results support the continued development and clinical integration of self-supervised learning techniques in musculoskeletal ultrasound imaging.

Data availability

The datasets generated and/or analyzed during the current study are not publicly available due to patient privacy restrictions, but are available from the corresponding author on reasonable request.

Code availability

All code related to our experiments can be found publicly and downloaded at the following link on Zenodo: <https://doi.org/10.5281/zenodo.20550128>.

Received: 4 September 2025; Accepted: 5 June 2026

Published online: 10 June 2026

References

- Kim, B. S., Kim, K., Day, J., Seilern Und Aspang, J. & Kim, J. Ultrasound-guided popliteal nerve block with short-acting lidocaine in the surgical treatment of ingrown toenails. *Int. J. Environ. Res. Public Health*. **18**, 5059. <https://doi.org/10.3390/ijerph18105059> (2021).
- Kim, B. S., Yu, M., Kim, S., Yoon, J. S. & Baek, S. Scale-attentional U-Net for the segmentation of the median nerve in ultrasound images. *Ultrasonography* **41**, 416–427. <https://doi.org/10.14366/uscg.21214> (2022).
- Webster, M. B., Kim, K. E., Na, Y. J., Lee, J. & Kim, B. S. Deep learning-based segmentation of the ulnar nerve in ultrasound images. *Medicina* **62**, 113. <https://doi.org/10.3390/medicina62010113> (2026).
- Hobson-Webb, L. D. Emerging technologies in neuromuscular ultrasound. *Muscle Nerve*. **61**, 719–725. <https://doi.org/10.1002/mus.26819> (2020).
- Wang, J. C. et al. Application of deep learning algorithms in automatic sonographic localization and segmentation of the median nerve: a systematic review and meta-analysis. *Artif. Intell. Med.* **137**, 102496. <https://doi.org/10.1016/j.artmed.2023.102496> (2023).
- Ronneberger, O., Fischer, P., Brox, T. & U-Net, Convolutional networks for biomedical image segmentation. In *Med Image Comput Comput Assist Interv – MICCAI 2015. Lecture Notes Comput Sci* Vol. 9351 234–241 (Springer, 2015). https://doi.org/10.1007/978-3-319-24574-4_28.
- Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., Liang, J. & UNet++: A nested U-Net architecture for medical image segmentation. *Deep Learn Med Image Anal Multimodal Learn Clin Decis Support. Lecture Notes Comput Sci. Cham: Springer*; **11045**, 3–11. (2018). https://doi.org/10.1007/978-3-030-00889-5_1
- Xie, E. et al. SegFormer: simple and efficient design for semantic segmentation with transformers. *Adv. Neural Inf. Process. Syst.* **35**, 924–937. <https://doi.org/10.48550/arXiv.2105.15203> (2021).
- Chen, L. C., Papandreou, G., Schroff, F. & Adam, H. Rethinking atrous convolution for semantic image segmentation. *Preprint arXiv*. <https://doi.org/10.48550/arXiv.1706.05587> (2017). 1706.05587.
- Shin, Y., Yang, J., Lee, Y. H. & Kim, S. Artificial intelligence in musculoskeletal ultrasound imaging. *Ultrasonography* **40**(1), 30–44. <https://doi.org/10.14366/uscg.20080> (2021).
- Smerilli, G. et al. Development of a convolutional neural network for the identification and the measurement of the median nerve on ultrasound images acquired at carpal tunnel level. *Arthritis Res. Ther.* **24**(1), 38. <https://doi.org/10.1186/s13075-022-02729-6> (2022).
- Horng, M. H., Yang, C. W., Sun, Y. N. & Yang, T. H. DeepNerve: a new convolutional neural network for the localization and segmentation of the median nerve in ultrasound image sequences. *Ultrasound Med. Biol.* **46**(9), 2439–52. <https://doi.org/10.1016/j.ultrasmedbio.2020.03.017> (2020).
- Huang, C. et al. Applying deep learning in recognizing the femoral nerve block region on ultrasound images. *Ann. Transl. Med.* **7**(18), 453. <https://doi.org/10.21037/atm.2019.08.61> (2019).
- Zhao, Z., Alzubaidi, L., Zhang, J., Duan, Y. & Gu, Y. A comparison review of transfer learning and self-supervised learning: Definitions, applications, advantages and limitations. *Expert Syst. Appl.* **242**, 122807. <https://doi.org/10.1016/j.eswa.2023.122807> (2024).
- Hosseinzadeh Taher, M. R., Haghighi, F., Feng, R., Gotway, M. B. & Liang, J. A systematic benchmarking analysis of transfer learning for medical image analysis. *Domain Adapt Represent Transf Afford Healthc AI Resour Divers Glob Health.* ;**12968**:3–13. (2021). https://doi.org/10.1007/978-3-030-87722-4_1
- He, K. et al. Masked autoencoders are scalable vision learners. In: *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* 15979–88. (2022). <https://doi.org/10.1109/CVPR52688.2022.01553>
- Rahman, A. & Patel, V. M. UltraMAE: multi-modal masked autoencoder for ultrasound pre-training. *Proc. Mach. Learn. Res. (MIDL)*. **250**: 1196–1206. (2024). <https://proceedings.mlr.press/v250/rahman24a.html>
- Sourget, T., Hasany, S. N., Mériaudeau, F. & Petitjean, C. Can SegFormer be a true competitor to U-Net for medical image segmentation? *Med. Image Underst. Anal.* **1413**, 85–95. https://doi.org/10.1007/978-3-031-48593-0_8 (2023).
- Kalapos, A. & Gyires-Tóth, B. Self-supervised pretraining for 2D medical image segmentation. *Comput Vis – ECCV 2022 Workshops. Lect Notes Comput Sci.* **13807**, 514–529 (Springer, 2023). https://doi.org/10.1007/978-3-031-25082-8_31
- Kang, Q., Gao, J., Li, K. & Lao, Q. Deblurring masked autoencoder is better recipe for ultrasound image recognition. *Med Image Comput Comput Assist Interv. Lect Notes Comput Sci.* **14220**, 372–383 (Springer, 2023). https://doi.org/10.1007/978-3-031-43907-0_34
- Zhou, L. et al. Self pre-training with masked autoencoders for medical image classification and segmentation. In: *Proc. IEEE Int. Symp. Biomed. Imaging*. 1–6. (2023). <https://doi.org/10.1109/ISBI53787.2023.10230477>
- Xiang, Y., Acharya, R., Le, Q., Tan, J. H. & Chng, C. L. Thyroid nodule segmentation in ultrasound images using transformer models with masked autoencoder pre-training. *Front. Artif. Intell.* **8**, 1618426. <https://doi.org/10.3389/frai.2025.1618426> (2025).

23. Dosovitskiy, A. et al. An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale. *Preprint* <https://doi.org/10.48550/arXiv.2010.11929> (2020). arXiv:2010.11929.
24. Iakubovskii, P. Segmentation Models Pytorch [software]. GitHub; (2019). Available from: https://github.com/qubvel/segmentation_models.pytorch
25. Di Cosmo, M. et al. A deep learning approach to median nerve evaluation in ultrasound images of carpal tunnel inlet. *Med. Biol. Eng. Comput.* **60**(11), 3255–64. <https://doi.org/10.1007/s11517-022-02662-5> (2022).
26. Bowness, J. S. et al. Variability between human experts and artificial intelligence in identification of anatomical structures by ultrasound in regional anaesthesia: a framework for evaluation of assistive artificial intelligence. *Br. J. Anaesth.* **132**(5), 1063–72. <https://doi.org/10.1016/j.bja.2023.09.023> (2024).
27. Di Cosmo, M. et al. A deep learning approach to median nerve evaluation in ultrasound images of carpal tunnel inlet. *Med. Biol. Eng. Comput.* **60**(11), 3255–64. <https://doi.org/10.1007/s11517-022-02662-5> (2022).
28. Gungor, I. et al. A real-time anatomy identification via tool based on artificial intelligence for ultrasound-guided peripheral nerve block procedures: an accuracy study. *J. Anesth.* **35**(4), 591–4. <https://doi.org/10.1007/s00540-021-02947-3> (2021).
29. Kawanishi, K. et al. Automatic identification of ultrasound images of the tibial nerve in different ankle positions using deep learning. *Sensors* **23**(10), 4855. <https://doi.org/10.3390/s23104855> (2023).
30. Moser, F., Müller, S., Lie, T., Lango, T. & Hoff, M. Automated segmentation of the median nerve in patients with carpal tunnel syndrome. *Sci. Rep.* **14**(1), 16757. <https://doi.org/10.1038/s41598-024-65840-5> (2024).
31. Shinohara, I. et al. Diagnosis of cubital tunnel syndrome using deep learning on ultrasonographic images. *Diagnostics* **12**(3), 632. <https://doi.org/10.3390/diagnostics12030632> (2022).
32. Wang, J. C. et al. Application of deep learning algorithms in automatic sonographic localization and segmentation of the median nerve: a systematic review and meta-analysis. *Artif. Intell. Med.* **137**, 102496. <https://doi.org/10.1016/j.artmed.2023.102496> (2023).
33. Yeh, C. L., Wu, C. H., Hsiao, M. Y. & Kuo, P. L. Real-time automated segmentation of median nerve in dynamic ultrasonography using deep learning. *Ultrasound Med. Biol.* **49** (5), 1129–1136. <https://doi.org/10.1016/j.ultrasmedbio.2023.01.002> (2023).
34. Kowalska, B. & Sudol-Szopińska, I. Ultrasound assessment of selected peripheral nerves pathologies. Part II: entrapment neuropathies of the lower limb. *J. Ultrason.* <https://doi.org/10.15557/jou.2012.0033> (2012).
35. Jacobson, J. A., Wilson, T. J. & Yang, L. J. Sonography of common peripheral nerve disorders with clinical correlation. *J. Ultrasound Med.* **35** (4), 683–693. <https://doi.org/10.7863/ultra.15.05061> (2016).
36. Wu, C. H. et al. Sonographic tracking of the upper limb peripheral nerves: a pictorial essay and video demonstration. *Am. J. Phys. Med. Rehabil.* **94** (9), 740–747. <https://doi.org/10.1097/PHM.0000000000000344> (2015).
37. Gao, Y. et al. Medical image segmentation: a comprehensive review of deep learning-based methods. *Tomography* **11** (5), 52. <https://doi.org/10.3390/tomography11050052> (2025).
38. Müller, D., Soto-Rey, I. & Kramer, F. Towards a guideline for evaluation metrics in medical image segmentation. *BMC Res. Notes*. **15** (1), 210. <https://doi.org/10.1186/s13104-022-06096-y> (2022).
39. He, W., Jiang, Z., Xiao, T., Xu, Z. & Li, Y. A survey on uncertainty quantification methods for deep learning. *Preprint arXiv*. <https://doi.org/10.48550/arXiv.2302.13425> (2025). 2302.13425.

Author contributions

M.W. performed the formal analysis, investigation, methodology development, software implementation (data-loading, model training, and validation code), visualization (all figures), and contributed to drafting and revising the manuscript. K.K. contributed to the investigation, methodology, and conceptualization, and was involved in project administration, supervision, and drafting and revising the manuscript. J.W.L. contributed to conceptualization, and software implementation (data-loading and processing). J.N.L. contributed to conceptualization, data curation, project administration, and supervision, and was also involved in drafting and revising the manuscript. B.K. contributed to conceptualization, data curation, funding acquisition, project administration, and supervision, and was also involved in drafting and revising the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (RS-2024-00352120). This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R111A1A01074880).

Declarations

Competing interests

The authors declare no competing interests.

IRB approval

This study was approved by the Institutional Review Boards of Korea University Guro Hospital (IRB No. 2020GR0389) and Chung-Ang University Gwangmyeong Hospital (IRB No. 2511-284-179).

Additional information

Correspondence and requests for materials should be addressed to J.L. or B.S.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026