Using deep learning for ultrasound images to diagnose carpal tunnel syndrome with high accuracy

Shinohara, Issei ; Inui, Atsuyuki ; Mifune, Yutaka ; Nishimoto, Hanako ; Yamaura, Kohei ; Mukohara, Shintaro ; Yoshikawa, Tomoya ; Kato,…

(Citation)
Ultrasound in Medicine & Biology,48(10):2052-2059

(Issue Date)
2022-10-01

(Resource Type)
journal article

(Version)
Accepted Manuscript

Using deep learning for ultrasound images to diagnose carpal tunnel syndrome with high accuracy

Issei Shinohara, Atsuyuki Inui*, Yutaka Mifune, Hanako Nishimoto, Kohei Yamaura, Shintaro Mukohara, Tomoya Yoshikawa, Tatsuo Kato, Takahiro Furukawa, Yuichi Hoshino, Takehiko Matsushita, and Ryosuke Kuroda

Department of Orthopaedic Surgery, Kobe University Graduate School of Medicine, 5-2, Kusunoki-cho7, Chuo-ku, Kobe-shi, Hyogo 650–0017, Japan

*Correspondence to: Dr. Atsuyuki Inui

Department of Orthopaedic Surgery, Graduate School of Medicine, Kobe University, 5-2, Kusunoki-cho7, Chuo-ku, Kobe-shi, Hyogo 650–0017, Japan

Tel: +81-78-382-5111

Fax: +81-78-351-6944

E-mail: ainui@med.kobe-u.ac.jp

Abstract

Recently, deep learning (DL) algorithms have been adapted for the diagnosis of medical images. The purpose of this study is to detect image features by DL without measuring median nerve cross-sectional area (CSA) in ultrasonography (US) images of carpal tunnel syndrome (CTS) and calculate the diagnostic accuracy from the obtained confusion matrix. US images of 50 hands without CTS and 50 hands diagnosed with CTS were used in this study. The short axis image of the median nerve was visualized and 5000 images of both groups were prepared. Forty hands in each group were used as training data for the DL algorithm while the remaining were used as test data. Transfer learning was performed using three pre-trained models. The confusion matrix and receiver operating characteristic curves were used to evaluate the diagnostic accuracy. Furthermore, regions where DL was determined to be important were visualized. The highest score had an accuracy of 0.96, precision of 0.99, and recall of 0.94. Visualization of the important features showed that the DL models focused on the epineurium of the median nerve and the surrounding soft tissue. The proposed technique enables the accurate prediction of CTS without measuring the CSA.

Key words: artificial intelligence, carpal tunnel syndrome, confusion matrix, deep learning, electrophysiological studies, pre-trained models, ultrasonography, visualization

Introduction

Carpal tunnel syndrome (CTS) is a peripheral neuropathy caused by compression of the median nerve and has been studied extensively (Olney 2001). CTS is caused by various factors, including edema around the wrist joint, tendon inflammation, and manual activity (Padua, et al. 2016). Symptoms of CTS include intermittent paresthesia and dysesthesia, which tend to worsen at night (Genova, et al. 2020). In severe cases, weakness in the muscles innervated by the median nerve may occur, leading to weakness of thumb abduction (Nataraj, et al. 2014). To diagnose CTS, an accurate history is necessary to evaluate the presence of motor and sensory disturbances (Padua, et al. 2016), and can be confirmed using diagnostic modalities such as electrophysiological studies (EPS, i.e., nerve conduction studies; NCS) or ultrasonography (US) (Demino and Fowler 2020). According to the literature published by the American Association of Electrodiagnostic Medicine (AAEM), the sensitivity of NCS for CTS ranges from 63% to 85%, with a specificity of over 97% (Jablecki, et al. 2002). Although EPS reveals the level of the lesion with high diagnostic accuracy, it does not provide local information about the nerve or the etiology of the disease (Tai, et al. 2012). In this regard, US images can be used to evaluate the location of lesions and nerve conditions in real time and have been used as a diagnostic device for CTS in recent years (Inui, et al. 2016). The measurement of the median nerve cross-sectional area (CSA) at the entrance of the carpal tunnel is useful in diagnosing CTS (Tai, et al. 2012). Although US imaging is noninvasive and has been widely accepted for nerve evaluation in recent years, nerve identification itself requires a level of skill that novice surgeons may lack.

In recent years, research on using deep learning (DL) to assist in making a diagnosis for medical data has gained attention (Weston, et al. 2019). Especially in the field of medical imaging, DL using convolutional neural networks (CNNs) is used extensively to automatically learn image features (Shin, et al. 2016). This study focuses on the accuracy of DL and decision basis visualization techniques for US images of CTS.

The purpose of this study is to detect the features of US images of CTS through DL without measuring CSA and calculate the diagnostic accuracy from the obtained confusion matrix. This study hypothesized that AI would accurately predict CTS without measuring CSA by learning the characteristic echo patterns of nerves and surrounding tissues in US images.

Materials and Methods

The ethics committee of the Kobe University Graduate School of Medicine approved this study (No. B21009), and informed consent was obtained from all the patients involved. This is a retrospective case series (consecutive) study and all participants provided written informed consent.

The hands of 50 healthy volunteers (22 men and 28 women), without any CTS by clinical symptoms or EPS (control group), and the hands of 50 patients (19 men, 31 women) diagnosed with CTS by EPS between 2019 and 2021 (CTS group) were used in this study. The severity of CTS was defined based on previous reports and the results of EPS as follows (Bulut, et al. 2014, Kanatani, et al. 2013); stage 1 (normal distal motor latency [DML] and normal sensory nerve conduction velocity [SNCV]), stage 2 (DML $\geq$ 4.5 ms and normal SNCV), stage 3 (DML $\geq$ 4.5 ms and SNCV< 40 m/s), stage 4 (DML $\geq$ 4.5 ms and the absence of sensory response), stage 5 (absence of DML and SNCV). Patients with severity of stage 2 or higher were included in the CTS group. Healthy volunteers and those with EPS results of stage 1 were considered as the control group. Patients with a history of wrist surgery, including carpal tunnel release, were excluded. The sample size was determined by power analysis based on data from a previous study using G*Power 3.1 (Tai, et al. 2012)·(Inui, et al. 2016). A prior sample size calculation showed that a difference in CSA of 3 mm^2 was detectable in the two groups with a sample size of 70 participants (35 participants in each group) using a t-test (effect size = 0.6, α = 0.05, power = 0.8). The diagnosis of CTS was performed by a certified hand surgeon

(A.I), and US imaging was performed by a certified surgeon (A.I) with 11 years of musculoskeletal US imaging experience.

For US imaging, the short-axis image of the median nerve was delineated at the inlet of the carpal tunnel using an 18MHz linear probe (Canon APLIO300, TUS-A300, Toshiba, Canon Medical Systems, Tochigi, Japan). The gain, dynamic range, and frame rate were kept constant throughout all measurements and for all participants. The US movie was recorded by sliding the probe within 20 mm of the distal wrist crease (Figure 1). From the obtained US movies, 100 images per hand were captured, and 5000 images were prepared for both groups, and a 15 × 15 mm area, including the median nerve, was cropped and used for DL. The Deeplearning Toolbox in MATLAB (Mathworks, Massachusetts, Natick, US) was used for DL, where forty hands in each group were used as training data, and the remaining were used as test data. (Figure 2). During preprocessing, data augmentation was performed to increase the variation in the original dataset. The *ImageAugmentor* tool in MATLAB was used to augment training and validation images by applying horizontal flipping, rotation (-10° to 10°), scaling (× 0.8 to × 1.2), horizontal translation, vertical translation, and random shearing. Transfer learning was performed using three pre-trained models (SqueezeNet, MobileNet_v2, and EfficientNet). These models selected are widely used for medical image data and have been improved by reducing computation time and memory size. The models have different convolutional layers, namely: 18, 53, and 82 layers in SqueezeNet, MobileNet_v2, and EfficientNet, respectively. The confusion matrix was obtained using the test dataset of each training model. Furthermore, the AI detected the features from the original US images without measuring the CSA in this study. The image features which the DL models focused their attention to were visualized as heatmap and overlaid to the original images. In this study, local interpretable model agnostic explanations (LIME) and occlusion sensitivity were used to visualize the important features detected by the network (Zhou, et al. 2016).

Statistical analysis

Using the test data, the accuracy of DL with three different learning models was evaluated. The accuracy (percentage of correct answers for all data), precision (percentage of artificial intelligence (AI) correctly judging CTS group), recall (percentage of data correctly judged by AI as CTS group, same as sensitivity), specificity (percentage of data correctly judged by AI as control group), and F-measure (the harmonic mean of the accuracy and recall), which are widely used in the field of machine learning, were calculated based on the confusion matrix. In addition, the area under the ROC curve was calculated by plotting the true positive rate and false positive rate on the coordinate axes. The 95 % confidence intervals (CIs) for sensitivity, specificity, and area under the curve (AUC) were calculated using the bootstrap method (Matsuo, et al. 2020).

For comparison of manual CSA measurements between the two groups, Mann Whitney U test was performed using IBM SPSS Statistics v.21 (IBM, Armonk, NY, USA). For comparison of ROC curves, DeLong's test was performed using R software package. A statistically significant difference was defined as $p<0.05$.

Results

The flow diagram of the participants is shown in Figure 3. Of the 112 participants, including healthy volunteers who did not have symptoms (eight cases), 12 were excluded for surgical or other reasons. All participants, except healthy volunteers, underwent EPS and were assigned to two groups according to the severity. The breakdown of the severity was as follows; stage 1: 42 cases, stage 2: 12 cases, stage 3: 26 cases, stage 4: eight cases, stage 5: four cases. Finally, 50 hands including healthy volunteers were defined as the control group, and another 50 hands with severity of stage 2 or higher were defined as the CTS group. The mean age of the control group was 45.0 ± 7.3 years old (range: 35-55 years old), and the mean age of the CTS group was 69.5 ± 13.2 years old (range: 37-88 years old). The mean CSA (at the inlet of the carpal tunnel) measured manually was 8.2 ± 2.4 mm^2 (range: 7- 10mm^2) in the control group and 14.7 ± 5.3 mm^2 (range: 9- 22mm^2) in the CTS group, which was significantly larger in the CTS group ($p<0.01$).

The diagnostic accuracy of each model was evaluated based on the confusion matrix obtained from the test data. The prediction accuracy of each learning model is shown in Table 1. For the prediction of CTS by the DL model, the best score of accuracy was 0.96 in EfficientNet, for precision was 0.99 in EfficientNet, for recall was 0.94 in SqueezeNet, and for specificity was 0.99 in EfficientNet. The AUC, which is a plot of the true positive rate and false positive rate on the coordinate axis, was 0.978 (95 % CI; 0.975-0.980) for SqueezeNet, 0.964 (95 % CI; 0,962-0.967) for MobileNet_v2, and 0.998 (95 % CI; 0.995-0,999) for EfficientNet (Figure 4). There was no statistical difference of AUC between the three DL models. Occlusion sensitivity and image LIME visualized the important features detected by AI using an overlaid heat map. The results show that the learning models predict the presence or absence of CTS by focusing on the inner and surrounding tissues of the median nerve (Figure 5).

Discussion

In this study, the accuracy of diagnosis using DL for US images of CTS was evaluated. Although CTS diagnosis is typically performed based on CSA measurement by US imaging, image recognition by DL can predict CTS based on the echo pattern of the median nerve and surrounding tissue without CSA measurement. The highest score was an accuracy of 0.96, sensitivity of 0.94, and specificity of 1.00, indicating that the diagnosis could be made with higher accuracy even without CSA measurement. Furthermore, all models were able to predict CTS with a higher accuracy than the reported diagnostic accuracy of CSA measurements (sensitivity 0.87 and specificity 0.83) (Tai, et al. 2012).

In recent years, there has been an increasing number of reports on DL for such US images of CTS (Table 2). It has been reported that CSA measurements of the median nerve using a state-of-the-art CNN (Mask R-CNN) were in high agreement with CSA measurements of sonographers (Smerilli, et al. 2022); (Cosmo, et al. 2021). Wu et al. and Wang et al. reported a technique for automatic tracking of the median nerve from dynamic US (Wu, et al. 2021); (Wang, et al. 2020). In this study, we focused on how AI recognizes the characteristics of CTS, such as surrounding soft tissue changes and US intensity changes in the median nerve.

In this study, three pre-trained models with different convolutional layers were used. Recent advances in deep learning have focused not only on improving accuracy but also on reducing the weight of the model (Forrest N. Iandola, et al. 2016). A reduced network architecture results in shortened training time and reduced memory size. In the processing of medical images, it is important to make decisions with high accuracy in a short period of time from a huge amount of data. SqueezeNet uses model compression techniques to reduce the size of the convolutional layer (Forrest N. Iandola, et al. 2016) and is often applied in medical imaging for chest X-ray diagnosis (Ucar and Korkmaz 2020). MobileNet_v2 simplifies DL, improves efficiency, and reduces memory footprint by introducing inverted residuals with linear bottlenecks (Mark Sandler, et al. 2019). MobileNet_v2 is efficient in image classification and object detection and has been applied to the lung CT (Gang, et al. 2021). EfficientNet is capable of processing 6.1 times faster with 8.4 times smaller capacity than the previous learning models such as ResNet, which were widely used before 2019 (Tan and LE 2019). Although with more convolutional layers, a more detailed evaluation is possible in the learning model used in this study, good accuracy was obtained with SqueezeNet, which has the fewest convolutional layers. The accuracy of AI-based recognition and measurement of the CSA of the median nerve may be reduced by anatomical variations associated with CTS

(Smerilli, et al. 2022). In this study, AI learned the features of the images taking into account the anatomical variations, therefore, the accuracy may have been comparable to previous reports.

In medical research files, it is important to know the basis for the AI decision. Various explanatory methods are currently being studied for this, including gradient-weighted class activation mapping (Grad-CAM), occlusion sensitivity, and LIME (Zhou, et al. 2016). Grad-CAM is a technique used to visualize important pixels by weighing the gradient against the prediction (Selvaraju, et al. 2017). Grad-CAM is useful for image classification, while occlusion sensitivity and LIME are useful for detailed lesion observation (Panwar, et al. 2020); (Aminu, et al. 2021). However, Grad-CAM has a lower resolution than occlusion sensitivity and LIME. Therefore, in this study, occlusion sensitivity and LIME were used to visualize the basis of AI decisions. The occlusion sensitivity can effectively visualize multifocal glass opacities and consolidations and can evaluate important image features in detail (Aminu, et al. 2021). LIME is a method of extracting important regions by creating superpixels, and has recently become popular in the field of medical imaging (Ahsan, et al. 2021). Visualizing the basis for AI's decision could provide the necessary information for diagnosis. In the visualization of the region of interest by occlusion sensitivity and LIME in this study, the contrast in echogenicity between the epineurium and its internal tissues and the echogenicity of perineural tissues were captured as features. This may reflect the presence of a pseudo-neuroma due to compression by the transcarpal ligament and inflammation of the surrounding flexor tendon synovium in the CTS. As a result, it was possible to diagnose CTS with high accuracy without measuring CSA. The application of DL-based imaging to clinical practice can lead to a more accurate and convenient diagnosis of CTS. Furthermore, because DL models learn image features using an uninhibited and unbiased neural model compared to humans, DL feature visualizations may enable physicians to detect previously overlooked and unquantified features.

This study has some limitations. First, although power analysis was performed, the number of cases in which the analysis images were based was not large. Further studies are required to corroborate the results of the present study. Second, the present results were obtained with only a single US instrument, and no comparison of diagnostic accuracy with other instruments was made. US imaging is excellent for the diagnosis of soft tissues and we hope that this system can be extended to provide a more convenient and accurate diagnosis of CTS. Third, the AI in this study was trained through supervised learning based on images with established diagnoses. Therefore, comparisons of the accuracy of CTS predictions by AI were performed based on historical data. Prospective comparisons between AI diagnosis and CSA measurement should be performed in the future. Finally, US images are taken by one evaluator and reproducibility with other evaluators is not considered. Further research is expected to support this study.

Conclusion

In recent years, AI-based medical imaging diagnosis has attracted attention. In this study, DL was used to predict the presence of CTS from US images of the median nerve. As an evaluation, the prediction accuracy of three learning models with different convolutional layers was examined. Results showed that all three learning models were able to predict with high accuracy, with the highest model having an accuracy of 0.96, sensitivity of 0.94, and specificity of 1.00. This study has the potential to be extended and applied clinically in the future, based on further studies of more cases, to provide a more convenient and accurate diagnosis of CTS.

Acknowledgments: This work was supported by a grant from the Japanese Society for Surgery of the Hand (2020)

References

Ahsan MM, Nazim R, Siddique Z, Huebner P. Detection of COVID-19 patients from CT scan and chest X-ray data using modified MobileNetV2 and lime. Healthcare (Basel). 2021;9:1099. https://doi.org/10.3390/healthcare9091099.

Aminu M, Ahmad NA, Mohd Noor MH. Covid-19 detection via deep neural network and occlusion sensitivity maps. Alex Eng J. 2021;60:4829–55. https://doi.org/10.1016/j.aej.2021.03.052.

Bulut HT, Yildirim A, Ekmekci B, Gunbey HP. The diagnostic and grading value of diffusion tensor imaging in patients with carpal tunnel syndrome. Acad Radiol. 2014;21:767–73. https://doi.org/10.1016/j.acra.2014.02.009.

Cosmo MD, Chiara Fiorentino M, Villani FP, Sartini G, Smerilli G, Filippucci E, Frontoni E, Moccia S. Learning-Based Median Nerve Segmentation From Ultrasound Images For Carpal Tunnel Syndrome Evaluation. Annu Int Conf IEEE Eng Med Biol Soc 2021; 2021:3025-28.

Demino C, Fowler JR. Comparison of borderline ultrasound and nerve conduction studies for carpal tunnel syndrome. Hand (NY) 2020. 1558944720964963.

Forrest N. Iandola, Song Han, Matthew W. Moskewicz, Khalid Ashraf, William J. Dally, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. arXiv:1602.07360 2016.

Gang L, Haixuan Z, Linning E, Ling Z, Yu L, Juming Z. Recognition of honeycomb lung in CT images based on improved MobileNet model. Med Phys 2021; 48:4304-15.

Genova A, Dix O, Saefan A, Thakur M, Hassan A. Carpal tunnel syndrome: a review of literature. Cureus. 2020;12:e7333. https://doi.org/10.7759/cureus.7333.

Inui A, Nishimoto H, Mifune Y, Kokubu T, Sakata R, Kurosaka M. Ultrasound measurement of median nerve cross-sectional area at the inlet and outlet of carpal tunnel after carpal tunnel release compared to electrodiagnostic findings. Arch Orthop Trauma Surg. 2016;136:1325–30. https://doi.org/10.1007/s00402-016-2514-9.

Jablecki CK, Andary MT, Floeter MK, Miller RG, Quartly CA, Vennix MJ, Wilson JR. Practice parameter: electrodiagnostic studies in carpal tunnel syndrome. Report of the American Association of electrodiagnostic Medicine, American Academy of Neurology, and the American Academy of Physical Medicine and Rehabilitation. Neurology. 2002;58:1589–92. https://doi.org/10.1212/wnl.58.11.1589.

Kanatani T, Fujioka H, Kurosaka M, Nagura I, Sumi M. Delayed electrophysiological recovery after carpal tunnel release for advanced carpal tunnel syndrome: a two-year follow-up study. J Clin Neurophysiol. 2013;30:95–7. https://doi.org/10.1097/WNP.0b013e31827ed839.

Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, Chen L-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. arXiv: 1801.04381 2019

Matsuo H, Nishio M, Kanda T, Kojita Y, Kono AK, Hori M, Teshima M, Otsuki N, Nibu KI, Murakami T. Diagnostic accuracy of deep-learning with anomaly detection for a small amount of imbalanced data: discriminating malignant parotid tumors in MRI. Sci. Rep.:19388. Sci Rep. 2020;10:19388. https://doi.org/10.1038/s41598-020-76389-4.

Nataraj R, Evans PJ, Seitz WH, Jr., Li ZM. Effects of carpal tunnel syndrome on reach-to-pinch performance. PLOS ONE. 2014;9:e92063. https://doi.org/10.1371/journal.pone.0092063.

Olney RK. Carpal tunnel syndrome: complex issues with a "simple" condition. Neurology. 2001;56:1431–2. https://doi.org/10.1212/wnl.56.11.1431.

Padua L, Coraci D, Erra C, Pazzaglia C, Paolasso I, Loreti C, Caliandro P, Hobson-Webb LD. Carpal tunnel syndrome: clinical features, diagnosis, and management. The Lancet Neurology 2016; 15:1273-84. https://doi.org/10.1016/S1474-4422(16)30231-9.

Panwar H, Gupta PK, Siddiqui MK, Morales-Menendez R, Bhardwaj P, Singh V. A deep learning and grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-Scan images. Chaos Solitons Fract. 2020;140:110190. https://doi.org/10.1016/j.chaos.2020.110190.

Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Visual explanations from deep networks via gradient-based localization. Ieee I conference Comp Vis; 2017, p. 618–26.

Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I Yao J, Mollura D, Summers RM. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging. 2016;35:1285–98. https://doi.org/10.1109/TMI.2016.2528162.

Smerilli G, Cipolletta E, Sartini G, Moscioni E, Di Cosmo M, Fiorentino MC, Moccia S, Frontoni E, Grassi W, Filippucci E. 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 2022; 24:38

Tai TW, Wu CY, Su FC, Chern TC, Jou IM. Ultrasonography for diagnosing carpal tunnel syndrome: a meta-analysis of diagnostic test accuracy. Ultrasound Med Biol. 2012;38:1121–8. https://doi.org/10.1016/j.ultrasmedbio.2012.02.026.

Tan M, Le QV. EfficientNet: rethinking model scaling for convolutional neural networks. ICML. 2019;2019.

Ucar F, Korkmaz D. COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. Med Hypotheses 2020; 140:109761

Wang YW, Chang RF, Horng YS, Chen CJ. MNT-DeepSL: Median nerve tracking from carpal tunnel ultrasound images with deep similarity learning and analysis on continuous wrist motions. Comput Med Imaging Graph 2020; 80:101687.

Weston AD, Korfiatis P, Kline TL, Philbrick KA, Kostandy P, Sakinis T Sugimoto M, Takahashi N, Erickson BJ. Automated abdominal segmentation of CT scans for body composition analysis using deep learning. Radiology. 2019;290:669–79. https://doi.org/10.1148/radiol.2018181432.

Wu CH, Syu WT, Lin MT, Yeh CL, Boudier-Reveret M, Hsiao MY, Kuo PL. Automated Segmentation of Median Nerve in Dynamic Sonography Using Deep Learning: Evaluation of Model Performance. Diagnostics (Basel) 2021; 11.

Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. IEEE Conference; 2016.

Figure Legends

Figure 1. (a) The ultrasonography (US) transducer positioned at the inlet of the carpal tunnel. (b) Short-axis image of the median nerve (red arrows) at the inlet of the carpal tunnel

Figure 2. Randomly extracted images using Matlab's Deeplearning Toolbox (Mathworks) (Blue; Control, Red; carpal tunnel syndrome (CTS)).

Figure 3. Flow of participants

Figure 4. Area under the curve (AUC) based on the receiver operating characteristic (ROC) curve was high for all learning models.

Figure 5. Visualization of region of interest using occlusion sensitivity and image local interpretable model agnostic explanations (LIME). Learning models focus on the neural interior and perineural tissue.

Table Legends

Table 1; Diagnostic accuracy of each learning model

Network	Accuracy	Precision	Recall (Sensitivity)	Specificity	F measure
SqueezeNet	0.943 (0.941–0.945)	0.957 (0.954–0.959)	0.943 (0.941–0.946)	0.937 (0.935–0.939)	0.950 (0.948–0.951)
MobileNet_v2	0.941 (0.940–0.943)	0.961 (0.960–0.963)	0.937 (0.935–0.939)	0.953 (0.951–0.955)	0.949 (0.947–0.951)
EfficientNet	0.959 (0.958–0.961)	0.998 (0.997–0.998)	0.935 (0.933–0.937)	0.997 (0.995–0.998)	0.965 (0.964–0.966)

(95 % confidence interval)

Table2; Summary of recent studies on US CTS identification

Reference	Method	N	Results and evaluations
Smerlli et al. (2022)	Localize and segment the median nerve section (Mask R-CNN)	246 images	Precision; 0.86, Recall; 0.88 Mean average precision; 0.88 Dice similarity coefficient; 0.86
Cosmo et al. (2021)	Localize and segment the median nerve section (Mask R-CNN)	151 images	Dice similarity coefficient; 0.93
Wu et al. (2021)	Segment the median nerve in dynamic US (Deeplabv3+, U-Net FPN, Mask R-CNN)	52 dynamic US images	Intersection over union Average 0.83 for Deeplabv3+ and Mask R-CNN
Wang et al. (2020)	Median nerve tracking using a DL model. (MNT-DeepSL)	100 cases, 84 with CTS	Accuracy; 0.9
Our study	Visualization (SqueezeNet, MobileNet_v2, EfficientNet)	100 cases, 10,000 images	Best score Accuracy; 0.96, Precision; 0.99 Recall; 0.94, F measure; 0.97

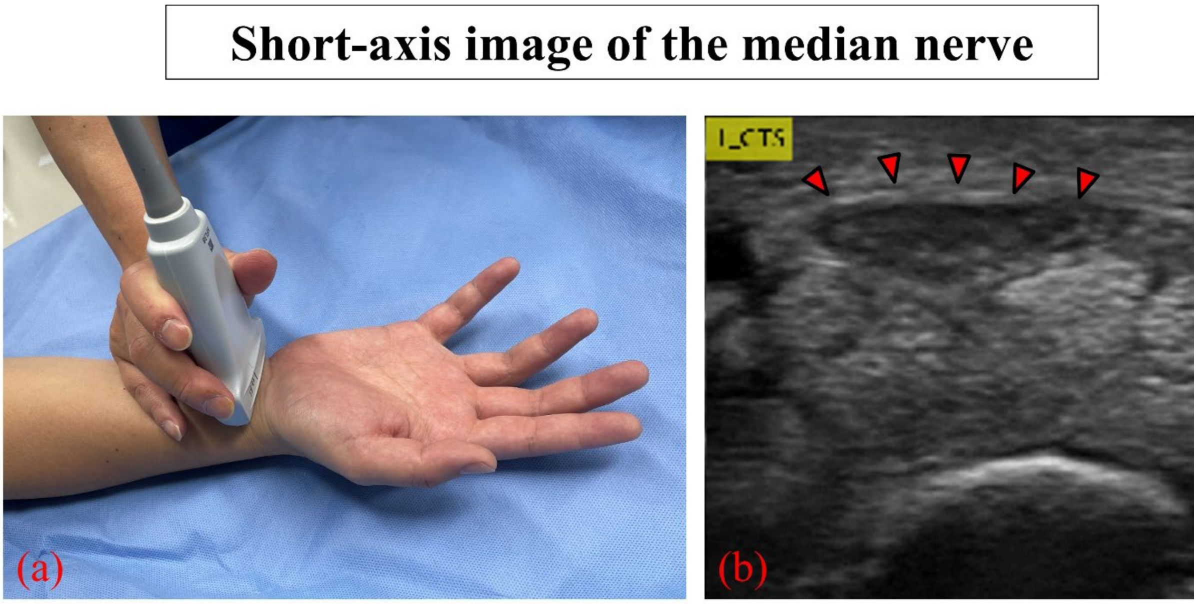

Figure 1

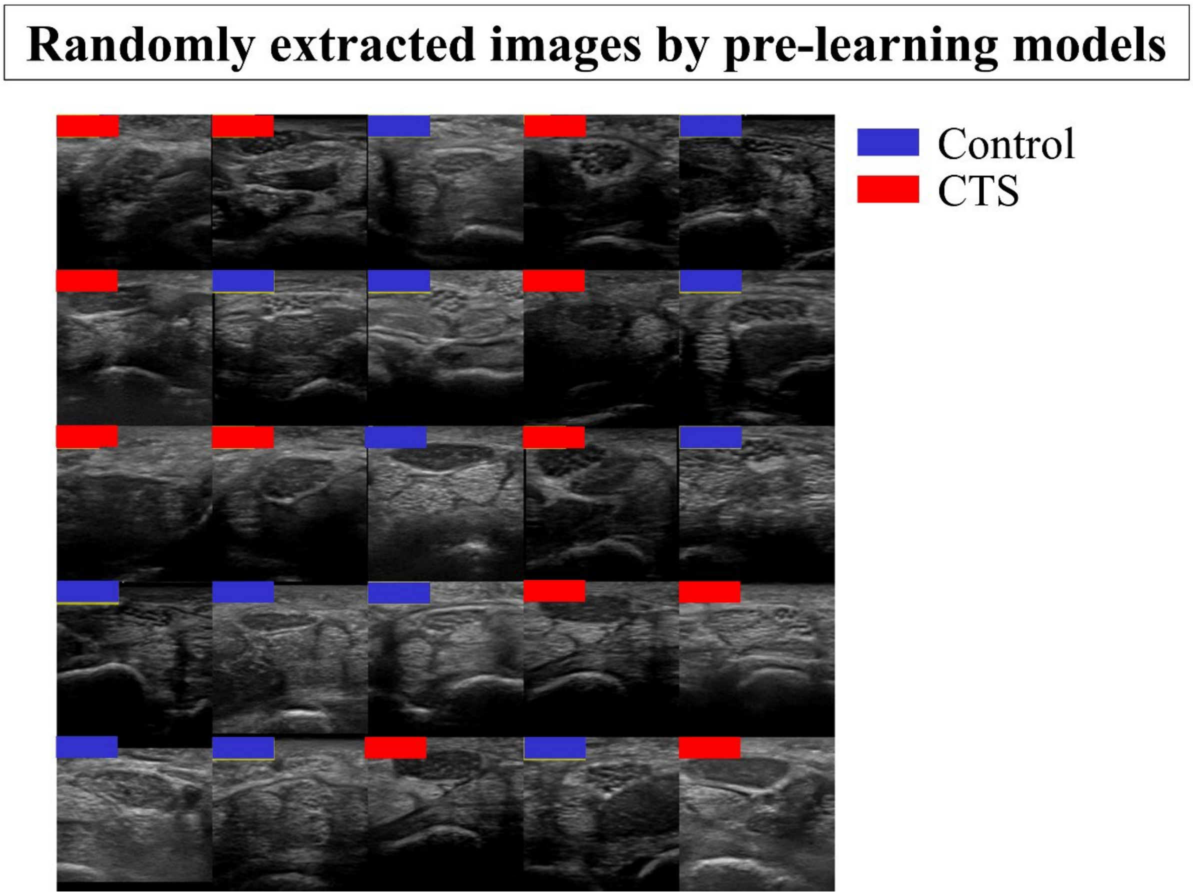
Randomly extracted images by pre-learning models
Control
CTS

Figure 2

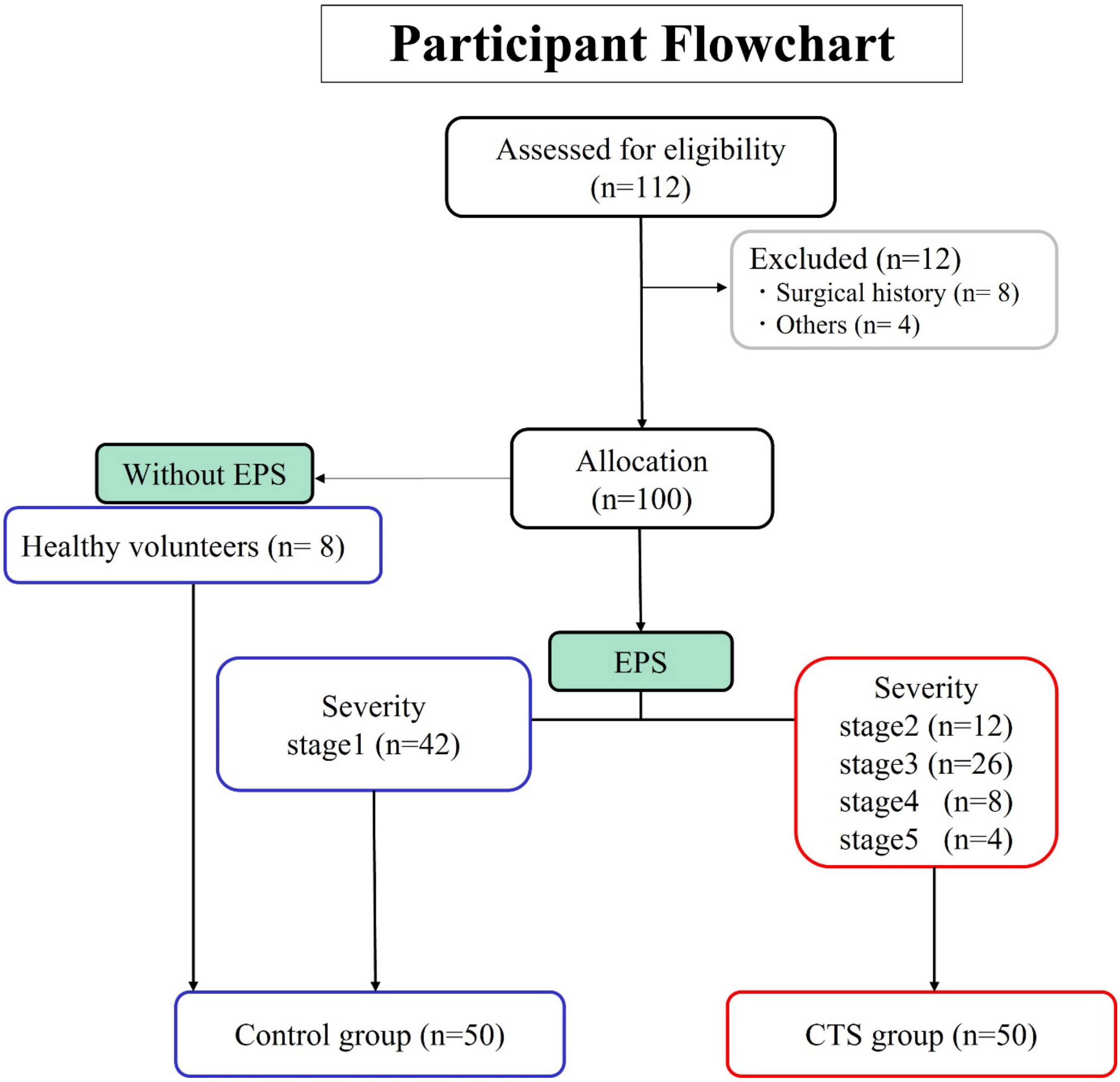
Participant Flowchart
Assessed for eligibility (n=112)
Excluded (n=12)
・Surgical history (n= 8)
・Others (n= 4)
Allocation (n=100)
Without EPS
Healthy volunteers (n= 8)
EPS
Severity stage1 (n=42)
Severity
stage2 (n=12)
stage3 (n=26)
stage4 (n=8)
stage5 (n=4)
Control group (n=50)
CTS group (n=50)

Figure 3

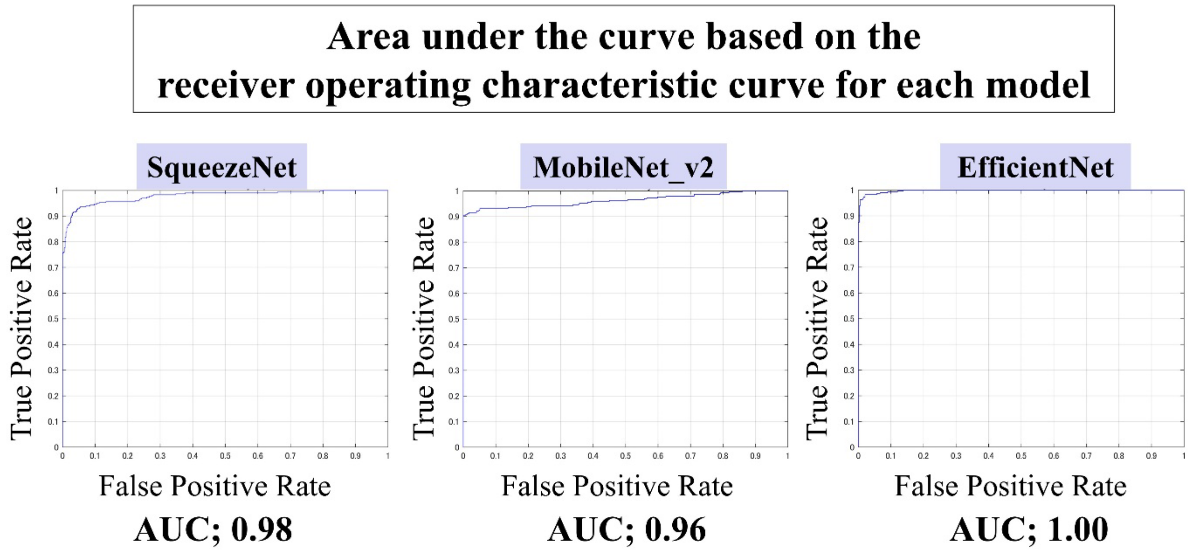
Area under the curve based on the
receiver operating characteristic curve for each model
SqueezeNet
MobileNet_v2
EfficientNet
True Positive Rate
False Positive Rate
AUC; 0.98
AUC; 0.96
AUC; 1.00

Figure 4

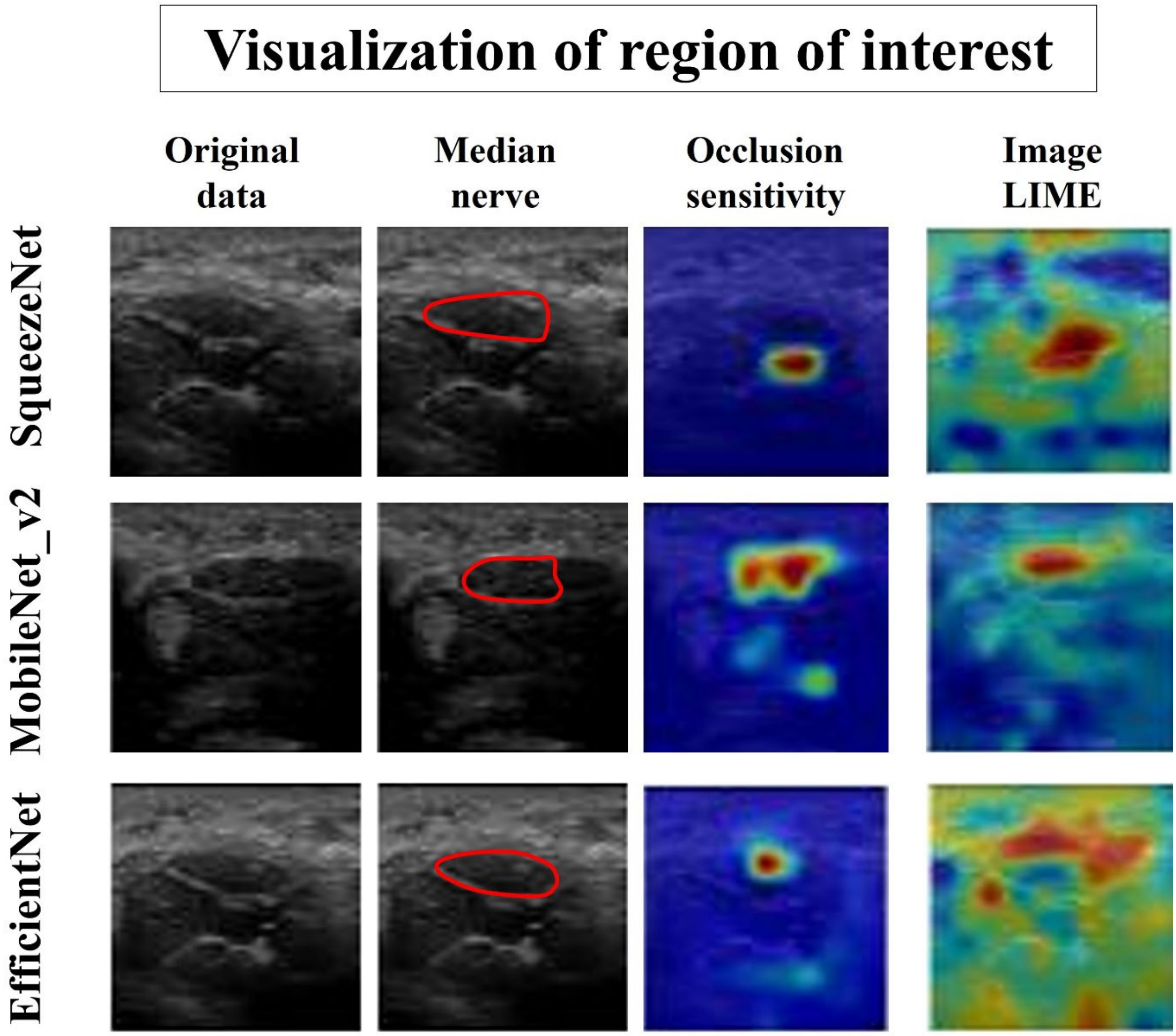
Visualization of region of interest
Original data
Median nerve
Occlusion sensitivity
Image LIME
SqueezeNet
MobileNet_v2
EfficientNet

Figure 5