






## Article

# Diagnosis of Cubital Tunnel Syndrome Using Deep Learning on Ultrasonographic Images

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, Kobe 650-0017, Japan; 203m878m@stu.kobe-u.ac.jp (I.S.); mf\_ship@yahoo.co.jp (Y.M.); hanako-nishi@live.jp (H.N.); koheidesuyo@yahoo.co.jp (K.Y.); no.8-shintaro@hotmail.co.jp (S.M.); tomo\_yoshi\_0926@yahoo.co.jp (T.Y.); tkato@med.kobe-u.ac.jp (T.K.); taka1023@med.kobe-u.ac.jp (T.F.); u1ho4no@med.kobe-u.ac.jp (Y.H.); matsushi@med.kobe-u.ac.jp (T.M.); kurodar@med.kobe-u.ac.jp (R.K.)

\* Correspondence: ainui@med.kobe-u.ac.jp; Tel.: +81-78-382-5111; Fax: +81-78-351-6944

**Abstract:** Although electromyography is the routine diagnostic method for cubital tunnel syndrome (CuTS), imaging diagnosis by measuring cross-sectional area (CSA) with ultrasonography (US) has also been attempted in recent years. In this study, deep learning (DL), an artificial intelligence (AI) method, was used on US images, and its diagnostic performance for detecting CuTS was investigated. Elbow images of 30 healthy volunteers and 30 patients diagnosed with CuTS were used. Three thousand US images were prepared per each group to visualize the short axis of the ulnar nerve. Transfer learning was performed on 5000 randomly selected training images using three pre-trained models, and the remaining images were used for testing. The model was evaluated by analyzing a confusion matrix and the area under the receiver operating characteristic curve. Occlusion sensitivity and locally interpretable model-agnostic explanations were used to visualize the features deemed important by the AI. The highest score had an accuracy of 0.90, a precision of 0.86, a recall of 1.00, and an F-measure of 0.92. Visualization results show that the DL models focused on the epineurium of the ulnar nerve and the surrounding soft tissue. The proposed technique enables the accurate prediction of CuTS without the need to measure CSA.

**Keywords:** artificial intelligence; cubital tunnel syndrome; deep learning; ulnar nerve; ultrasonography



**Citation:** Shinohara, I.; Inui, A.; Mifune, Y.; Nishimoto, H.; Yamaura, K.; Mukohara, S.; Yoshikawa, T.; Kato, T.; Furukawa, T.; Hoshino, Y.; et al. Diagnosis of Cubital Tunnel Syndrome Using Deep Learning on Ultrasonographic Images. *Diagnostics* **2022**, *12*, 632. <https://doi.org/10.3390/diagnostics12030632>

Academic Editor: Ayman El-Baz

Received: 7 February 2022

Accepted: 2 March 2022

Published: 4 March 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Cubital tunnel syndrome (CuTS) is an entrapment neuropathy caused by the compression and traction of the ulnar nerve at the elbow joint [1]. It is the second most common compressive neuropathy of the upper extremities, with a reported incidence of 25 cases per 100,000 person-years in males and 19 cases per 100,000 person-years in females [2]. Typical CuTS symptoms include numbness, tingling, and dysesthesia in the ring/small finger and dorsum of the hand [3]. As the disease progresses, atrophy of the hypothenar and first dorsal interosseous muscles occurs, which may interfere with daily activities [1]. CuTS is likelier to be at an advanced stage at the time of intervention than carpal tunnel syndrome (CTS) [4]. Therefore, early diagnosis and treatment are important [1]. The first step in diagnosis is obtaining the patient's medical history, observing the intrinsic musculature of the hand, and evaluating hand sensation [4]. Froment's and Tinel's signs at the cubital tunnel are widely used as physical indicators [5]. When CuTS is suspected based on medical history and physical examination, diagnostic techniques, such as magnetic resonance imaging (MRI), electromyography (EMG), and ultrasonography (US), are used [4]. The American Academy of Neuromuscular and Electrodiagnostic Medicine reports the following two measures as the basis for the diagnosis of CuTS: conduction velocity across the elbow is less than 50 m/s, and conduction velocity from above the elbow to below

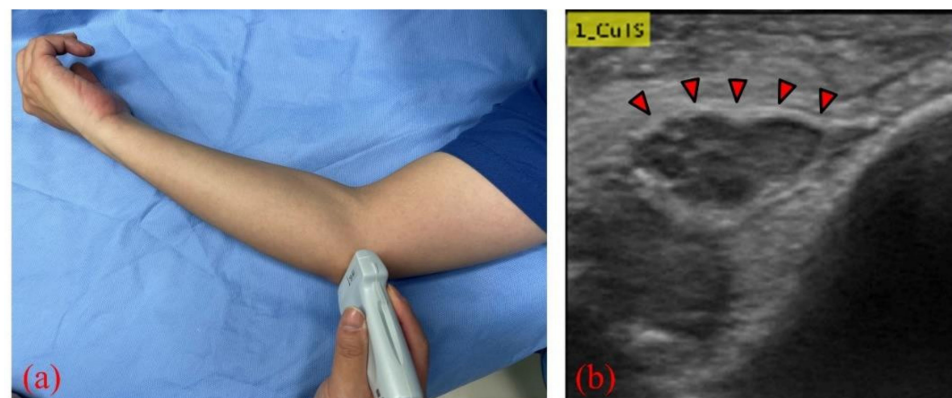
the elbow is slower than that from below the elbow to the wrist by 10 m/s or more [6]. Although EMG testing can help determine the severity of CuTS, MRI can confirm nerve enlargement and compression and is reported to have a 25% higher diagnostic sensitivity than EMG [7]. However, it is impossible to determine the severity of the disease or the prognosis for recovery using MRI [7]. Although both tests are important for diagnosis and staging, they are time consuming and expensive. Furthermore, there are some patients with apparent clinical symptoms of CuTS without abnormal EMG findings, and the diagnostic accuracy (sensitivity 78%) is currently lower than that of the median nerve [8]. US is a minimally invasive and low-cost diagnostic technique that can assess the size of the ulnar nerve. Additionally, US imaging allows for dynamic evaluation, whereas MRI is static. In recent years, high-resolution US has become increasingly effective in the diagnosis of entrapment neuropathies (e.g., CuTS and CTS) [8]. US-based diagnostic studies have shown an enlarged nerve cross-sectional area (CSA) in patients with CuTS symptoms, even when the EMG is normal [8]. The measurement of CSA by US imaging is expected to provide an early diagnosis of CuTS. Chang et al. reported that when  $CSA > 10 \text{ mm}^2$ , the diagnostic accuracy is high, with a sensitivity of 0.85 and a specificity of 0.91 [9,10]. Although there is room for further study on CSA thresholds, US imaging is becoming the first choice for the imaging of peripheral lesions of limbs [11]. On the other hand, US imaging has limitations, such as anisotropy and a lack of technical proficiency [12]. These factors invariably cause errors in CSA measurement, which, in turn, diminish diagnostic accuracy. Therefore, this study focuses on the application of DL to US images.

In recent years, there has been remarkable progress in the field of computer science, due to the increase in processing speed, which has become increasingly popular in medical image processing [13]. In the field of medical imaging, deep learning (DL) using convolutional neural networks (CNNs) is widely used to automatically learn image features [14]. These methods are now being used for various applications, such as the diagnosis of pneumonia and fractures in X-rays [15,16]. DL diagnostic methods have also been applied to US imaging, and its diagnostic accuracy for thyroid nodules is 83% [17]. Additionally, DL visualization methods can intuitively identify important areas to further examine for disease prediction [18]. In the field of orthopedics, DL-based diagnostic imaging techniques have been attracting attention in recent years and the numbers of reports are gradually increasing. Throughout the orthopedic literature, DL has been used for diagnostic purposes in the study of hip and knee joints based on medical imaging [19]. In the diagnosis of osteoarthritis (OA) of the hip joint using X-rays, the CNNs model achieves an accuracy of 92.8%, which is comparable to that of an attending physician with 10 years of experience in diagnosing hip OA [20]. For the knee joint, Ashinsky et al. used the DL algorithm for T2-weighted maps of the medial femoral epicondyle to predict progression to clinically symptomatic OA; the DL algorithm was able to predict the development of OA with 75% accuracy [21]. As shown in these results, DL is reported to have promising potential in the field of intelligent medical imaging practice, especially for OA of the hip and knee joints. On the other hand, there has been no study using DL for entrapment neuropathies, which is becoming the main diagnostic tool using US images. If AI technology can predict diseases as accurately as conventional diagnostic methods, its application in clinical practice may lead to early diagnosis. Therefore, the purpose of this study is to calculate diagnostic accuracy from a confusion matrix using DL on US CuTS images. The hypothesis of this study is twofold: the application of DL to US imaging allows CuTS diagnosis without the need to measure CSA, and it does so at an accuracy at least as high as conventional MRI and EMG methods. Furthermore, we considered the possibility of obtaining useful information for the diagnosis of CuTS by visualizing areas deemed important. To the best of our knowledge, this study is the first to investigate the use of DL on US images for CuTS diagnosis.

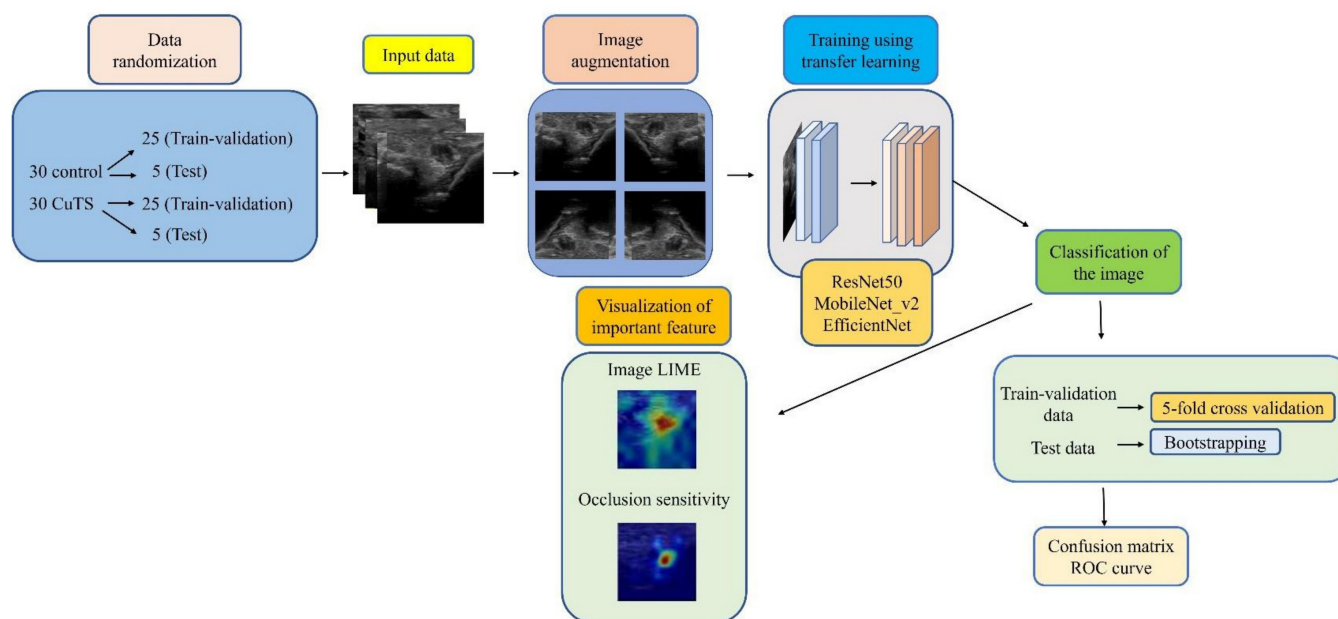
## 2. Materials and Methods

The ethics committee in our hospital approved this study (No. B21009), and informed consent was obtained from all participants.

We collected elbow US images from 60 participants—30 healthy volunteers comprising 16 males and 14 females (control group), and 30 patients comprising 18 males and 12 females who were diagnosed with CuTS via EMG (patient group). The criteria for diagnosis by EMG were based on data from previous reports [22]: absolute motor nerve conduction velocity (mNCV) of less than 50 m/s across the elbow and a decrease in velocity of more than 10 m/s above and below the elbow. EMG of the patient group showed a decrease in mNCV ( $\leq 50$  m/s) across the elbow and a decrease in velocity of more than 10 m/s above and below the elbow in all subjects. A decrease in the amplitude of the compound muscle action potential was observed in 13 patients. Patients with a history of surgery on the elbow, including surgery for CuTS, were excluded. The mean ages of the control and patient groups were  $55.6 \pm 11.5$  years (range: 28–75) and  $63.8 \pm 13.2$  years (range: 43–75), respectively. The sample size was determined by performing a statistical power analysis, based on data from a previous study, using G\*Power 3.1 software [10]. A prior sample size calculation performed at the start of this study showed that a difference in CSA of  $3 \text{ mm}^2$  was detectable in two groups, with a sample size of 26 participants (13 in each group), using a t-test (effect size = 1.2,  $\alpha$  error = 0.05, power = 0.9). US imaging was performed by one certified hand surgeon (A.I.) with 11 years of musculoskeletal ultrasound imaging experience. US imaging showed a short-axis image of the ulnar nerve at the level of the medial epicondyle using an 18M linear probe (Figure 1: Canon APLIO300, TUS-A300, Canon Medical Systems, Tochigi, Japan). The US settings of gain, dynamic range, and frame rate were standardized for all measurements. A US movie was captured by sliding the probe within a 30 mm range around the medial epicondyle. From the obtained US movies, 100 images were captured per elbow, and 3000 images were prepared from each group. An area of  $25 \times 25$  mm, that included the ulnar nerve, was cropped per image prior to DL processing. For DL, the DeepLearning Toolbox in MATLAB (Mathworks, Natick, MA, USA) [23] was used. The architecture of the framework is shown in Figure 2.



**Figure 1.** (a) US probe placed on the medial epicondyle to visualize the ulnar nerve; (b) short-axis image of the ulnar nerve (red arrows) at the level of the medial epicondyle.



**Figure 2.** Flowchart of the proposed framework.

### 2.1. Data Selection

Of the 30 participants in each group, data from 25 participants were randomly selected to validate training by 5-fold cross validation. The images from the remaining five participants were used as test data to evaluate the model performance.

### 2.2. Data Augmentation

For 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 ( $-10^\circ$  to  $10^\circ$ ), scaling ( $\times 0.8$  to  $\times 1.2$ ), horizontal translation, vertical translation, and random shearing.

### 2.3. Training Process

A total of 2500 images were randomly selected from each group and used as training data (Figure 3); the remainder were used as testing data. Transfer learning was performed using three pre-trained models (i.e., Residual Net (ResNet)-50, MobileNet\_v2, and EfficientNet). These models used different numbers of convolutional layers: 50 in ResNet-50, 53 in MobileNet\_v2, and 82 in EfficientNet. The block diagram of each learning model is shown in Figures 4–6. Patient data were randomly divided into two groups (training and testing). During training, five-fold cross validation was performed [24]. In this procedure, training data were randomly divided into five subsets: one was used for validation, and the remaining four were used for training. This process was repeated 5 times until each subset was used exactly once for validation. The hyperparameters of the training models were determined by using the Experimental Manager application in the DeepLearning Toolbox and the Parallel Computing Toolbox. The parameters are summarized in Table 1. To avoid overfitting, training and validation images were augmented by employing the random transformation process defined earlier. The model also validated the network, every 50 iterations, by predicting the responses of the validation data and calculating the loss and accuracy.

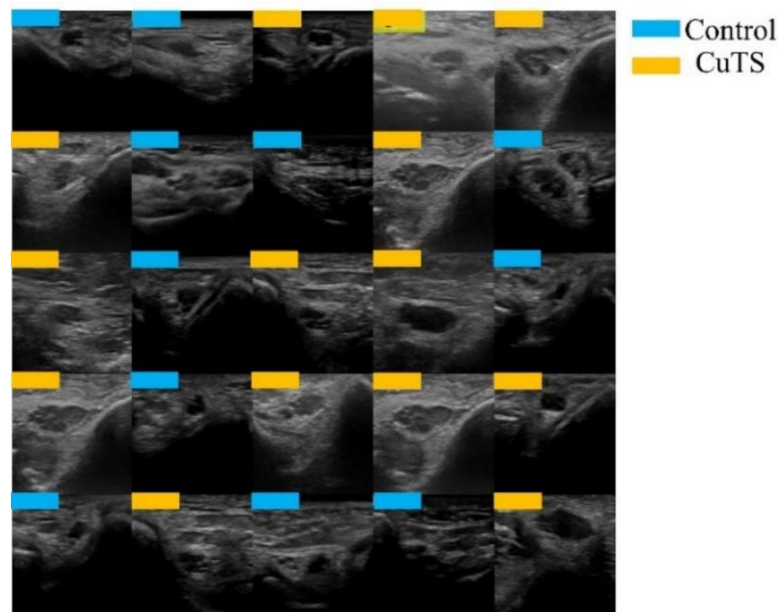


Figure 3. Images were randomly extracted by AI to be used as training data (light blue for control, orange for CuTS patients).

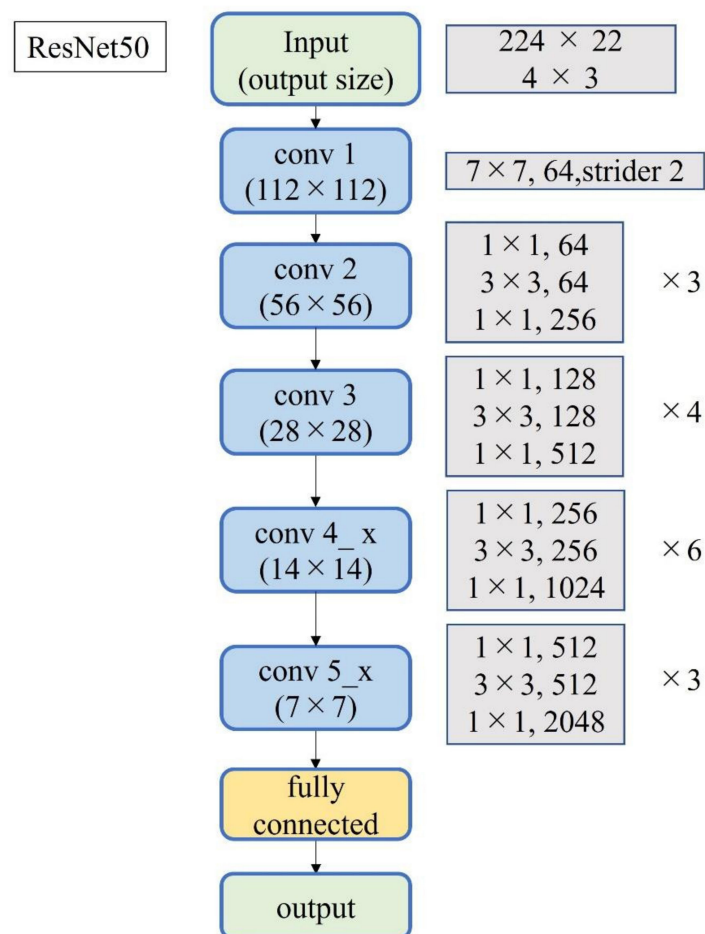
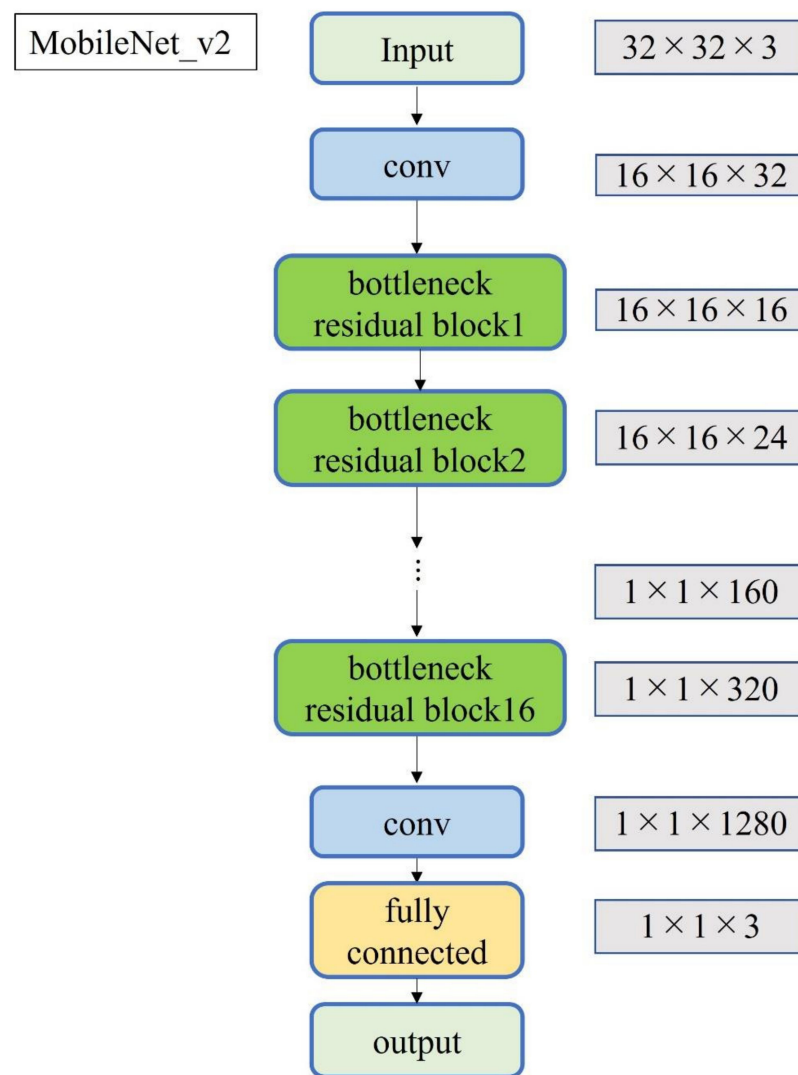


Figure 4. Block diagram of ResNet-50.



**Figure 5.** Block diagram of MobileNet\_v2.

**Table 1.** Best parameters of the training model.

	ResNet-50	MobileNet_v2	EfficientNet
Optimizer	Adam *	Adam	Adam
MiniBatchsize	20	10	10
Epochs	1000	700	1000
Learning rate	0.0001	0.0001	0.0001

\* Adam; adaptive moment estimation.

The accuracy of each DL model was evaluated using a confusion matrix, which is a table containing four combinations of predicted and actual values regarding the presence or absence of disease. The matrix values include true positives, false positives, true negatives, and false negatives (Figure 7).

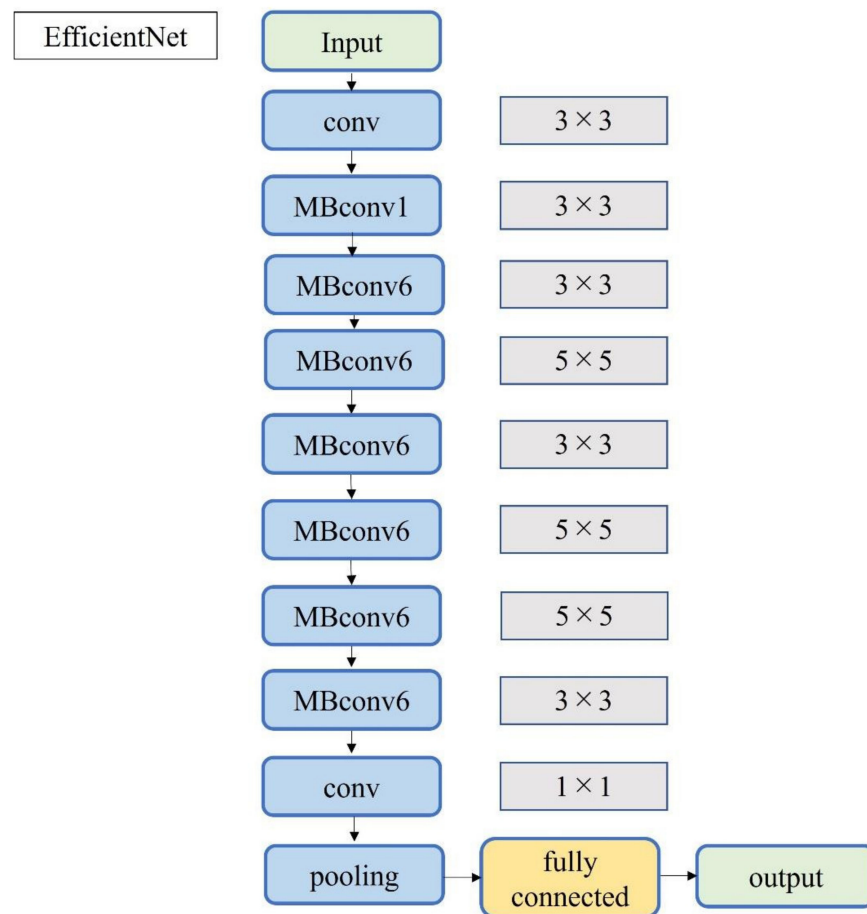


Figure 6. Block diagram of EfficientNet.

a)

		True class	
		CuTS	Control
Predicted class	CuTS	True Positive (TP) (TP)	False Positive (FP)
	Control	False Negative (FN)	True Negative (TN)

b)

Accuracy=  $(TP+TN) / (TP+FP+FN+TN)$   
 Precision=  $TP / (TP+FP)$   
 Recall (sensitivity) =  $TP / (TP+FN)$   
 Specificity=  $TN / FP+TN$   
 F.Measure=  $2 \times (Precision \times Recall) / (Precision + Recall)$

Figure 7. (a) A confusion matrix is a table of four combinations based on predicted and actual values and the presence or absence of disease; (b) diagnostic accuracy from the learning model is calculated from the confusion matrix created using testing data.

#### 2.4. Visualization of Important Features

Local interpretable model-agnostic explanations (LIMEs) and occlusion sensitivities were used to visualize image features deemed important by the DL model [18].

2.5. Measures

The accuracy of the three learning models was evaluated using testing data. The evaluation terms were defined as follows: accuracy (percentage of correct answers for all data), precision (percentage of the patient group correctly judged), recall (percentage of data correctly judged as the patient group, same as sensitivity), specificity (percentage of data correctly judged as the control group), and F-measure (the harmonic mean of accuracy and recall). Additionally, the accuracy of the classifier was evaluated based on the area under the receiver operating characteristic (ROC) curve (AUC). The 95% confidence interval (CI) for each endpoint was calculated using the bootstrap method [25], where 500 replacement resampling procedures were performed to evaluate the model performance on test data. To verify the absence of overfitting training data, all data were validated by five-fold cross validation [24]. For statistical analysis, a one-way ANOVA was performed using IBM SPSS Statistics v.21 (IBM, Armonk, NY, USA).

3. Results

The AUCs for CuTS prediction using the test data with ResNet-50, MobileNet\_v2, and EfficientNet were 0.987 (0.984–0.989), 0.990 (0.987–0.991), and 0.989 (0.987–0.991), respectively (Figure 8). The high AUC indicates the high accuracy of the classifier. For each learning model, the prediction accuracy based on the confusion matrix obtained from the test data is shown in Table 2. The confusion matrix for each learning model is shown in Figure 9. The results of cross validation show no statistically significant differences among the learning models or between the scores of validation and testing, suggesting that overfitting did not occur.

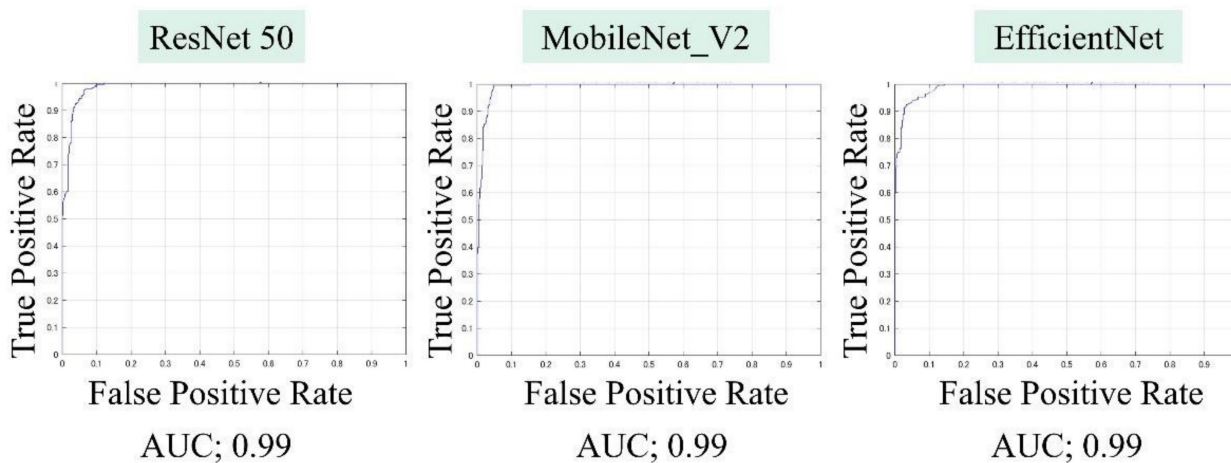


Figure 8. Area under the curve (AUC), based on the receiver operating characteristic (ROC) curve, was high for all learning models.

		ResNet 50		MobileNet_v2		EfficientNet	
		True class		True class		True class	
		CuTS	Control	CuTS	Control	CuTS	Control
Predicted class	CuTS	57.6 %	9.6 %	57.7 %	9.4 %	55.1 %	12.0 %
	Control	0 %	32.9 %	0.1 %	32.7 %	0 %	32.9 %

Figure 9. Confusion matrix of each learning model.

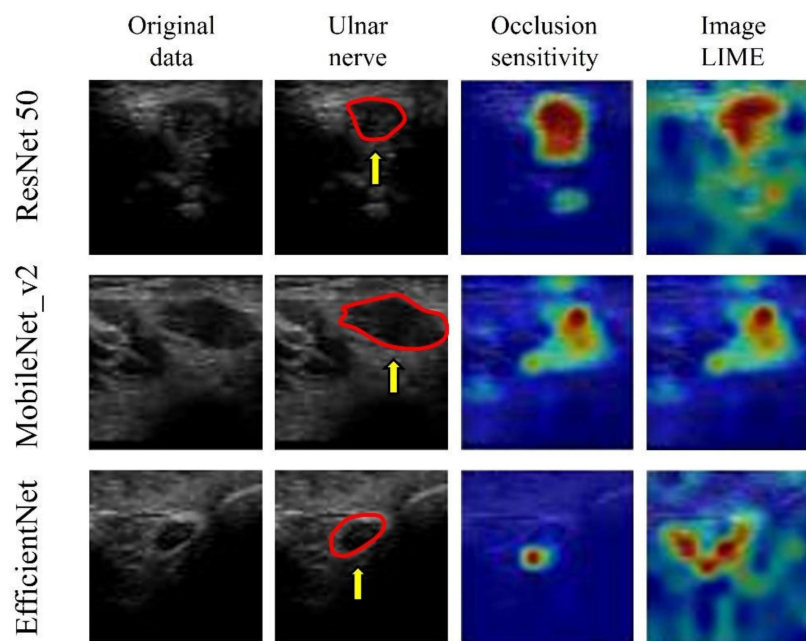


**Table 2.** Diagnostic accuracy from the learning model was calculated from the confusion matrix created from testing data. The best accuracy score was 0.90 in ResNet-50 and MobileNet\_v2; precision was 0.86 in ResNet-50 and MobileNet\_v2; recall was 1.0 in all models; and the F-measure was 0.92 in ResNet-50 and MobileNet\_v2.

Network	Accuracy	Precision	Recall (Sensitivity)	Specificity	F-Measure
ResNet-50	0.904 (0.903–0.907)	0.859 (0.856–0.861)	1.00	0.774 (0.772–0.776)	0.924 (0.922–0.925)
MobileNet_v2	0.904 (0.902–0.906)	0.859 (0.857–0.862)	0.998 (0.997–0.998)	0.776 (0.774–0.778)	0.923 (0.921–0.925)
EfficientNet	0.880 (0.878–0.882)	0.821 (0.819–0.824)	1.00	0.732 (0.731–0.734)	0.902 (0.900–0.904)

95% confidence interval.

The visualization of important features, using image LIMEs and occlusion sensitivity, verified that all three models predicted the presence of CuTS by focusing on hyperechoic changes in the ulnar nerve epithelium and those in the ulnar nerve interior and surrounding tissue (Figure 10).



**Figure 10.** Visualization of the region of interest using occlusion sensitivity and image LIMEs. Learning models focus on neural interior and perineural tissues. The red circle is a cross section of the ulnar nerve in the original image. AI focused on hyperechoic changes in the ulnar nerve epithelium and hypoechoic changes in the ulnar nerve interior and surrounding tissue.

#### 4. Discussion

Applying DL to US CuTS images resulted in highly accurate lesion predictions. The best scores were 0.90 for accuracy (ResNet-50, MobileNet\_v2), 0.86 for precision (ResNet-50, MobileNet\_v2), 1.00 for recall (all models), and 0.92 for F-measure (ResNet-50, MobileNet\_v2), indicating that diagnoses may be determined with high accuracy, even without CSA measurement. Therefore, the proposed technique can be useful in screening for CuTS. As mentioned, this study is the first to investigate the use of DL on US images for CuTS diagnosis.

In recent years, US imaging has been widely used for the diagnosis of musculoskeletal disorders [26,27]. In particular, the use of US imaging for the diagnosis of entrapment neuropathy is rapidly gaining popularity. Nerve enlargement has been observed centrally

above the nerve strangulation site, and the measurement of CSA is helpful in diagnoses [28]. The measurement of the median nerve CSA can be used to diagnose CTS with high accuracy (sensitivity 83%; specificity 86%) [29]. Furthermore, the measurement of the ulnar nerve CSA via US is also useful in the diagnosis of CuTS [9,10]. The ulnar nerve CSA in healthy volunteers rarely exceeds 10 mm<sup>2</sup> at various sites; thus, it is used as the cut-off value for CuTS [10]. Although US imaging is useful in diagnosing CuTS, measuring CSA by identifying and drawing the entrance is challenging. Therefore, we focused on the advantages of DL, which automatically learns image features and has been applied extensively to the analysis of medical data [15]. The application of DL to diagnostic imaging has been actively performed since 2015 [30]. Bar et al. evaluated the diagnostic accuracy of DL by classifying chest radiographs into healthy and pathological (e.g., cardiomegaly, right pleural effusion, and multiple others). A sensitivity of 0.89 and a specificity of 0.79 were observed in their study, indicating that DL using non-medical image databases can be used for general medical image recognition tasks [30]. Diagnostic imaging technology using DL has also been applied to orthopedics [31,32]. Amelia et al. applied DL to images of proximal femoral fractures and reported a prediction accuracy of 0.94 [31]. Additionally, Tomita et al. applied DL to computer tomography (CT) images and reported a diagnostic accuracy of 0.89 for osteoporotic vertebral fractures [33]. DL has also been used with US images to identify the vasculature in the axillary approach of the brachial plexus [34]. The results show that CNNs have a recall of 0.81 and a precision of 0.88 in detecting blood vessels [34]. Although imaging diagnoses using DL have been gradually reported in the orthopedic research field, there are no reports regarding the use of DL in diagnosing entrapment neuropathy. Research on DL for US images is ongoing in the field of breast surgery. The mortality rate of breast cancer is high, and DL has been studied heavily in its US diagnosis. Hence, diagnostic accuracy improvements have been at the forefront of research. Jabeen et al. obtained the best CNN model accuracy to date, of 99.1%, by optimizing five subprocesses: data augmentation, modified DarkNet-53 modeling, transfer learning, feature selection, and feature fusion and classification [35]. Feature fusion augments the features of small lesions by combining feature maps of shallow and deep convolutional layers and is commonly applied to mammography [36]. Meraj et al. used a quantization-assisted U-Net with data augmentation to accurately extract lesions from ultrasound images [37]. Although the lesions are sufficiently segmented by U-Net, the results show that the use of quantization for some segmented lesions leads to increased accuracy ( $\geq 96\%$ ) [37].

Early detection of entrapment neuropathy is also desirable as the progression of symptoms can interfere with daily life. In this field, the diagnosis of entrapment neuropathy by CNNs has not been reported. Since this is the first application of DL to US images of CuTS, this study focuses on comparing the diagnostic accuracy, using three popular pretrained DL models, and visualizing the basis for decision. In the present study, the possibility of predicting CuTS without measuring CSA was investigated by applying CNNs, which captured and evaluated the image features of CuTS diagnoses. The results show that the DL models used in this study predicted CuTS with high accuracy. As indicated by the confusion matrix results, no DL model mistakenly judged CuTS in the control group. US is minimally invasive, and the study shows the possibility of using the DL models for CuTS screening. Compared with previous reports that measured CSA (sensitivity 0.85, specificity 0.91) [9], each learning model in this study was particularly sensitive in predicting CuTS and was effective for screening. The extension and clinical application of this system are expected to facilitate the early detection of CuTS.

Notably, three DL models that are popularly used for medical image classification were selected for this study. ResNet50 solves the accuracy loss of DL by reconfiguring it to learn residual functions [38] to reduce the error rate and achieve high accuracy. ResNet50 has been used with MRI images of ACL injuries [39] and other US images [40]. MobileNet\_v2 simplifies DL by introducing inversion residuals with linear bottlenecks which improve efficiency and reduce the memory footprint [41]. MobileNet\_v2 is efficient in image classification and object detection and it can achieve the same, or higher, accuracy as the same

parameter model with high speed. It has been applied to lung CT [42]. EfficientNet is one of the most powerful CNN architectures reported in recent years [43]. It utilizes a complex scaling method to increase the depth, width, and resolution of the network, and provides state-of-the-art capabilities with fewer computational resources than other models [43]. EfficientNet can process image data 6.1 times faster than ResNet, using 8.4 times fewer memory resources [44]. Although memory cost and processing speed are vital to DL-based diagnostic imaging, the prediction accuracy of ResNet and MobileNet\_v2 was higher than that of EfficientNet in this study. EfficientNet emphasizes its deep convolutional layers and strong learning capability; this may lead to the problem of overlearning, as it retains too much information about each pixel in an image. The development of CNNs has been remarkable, and new models are being developed every year, but it may be important in the future to optimize an appropriate learning model for each diagnostic modality.

Although accuracy is essential in AI-based diagnostic imaging technology, the basis for model decision making must be understandable to humans. In particular, DL was used in this study for diseases lacking a US diagnosis basis, and emphasis was placed on visualizing DL decision making rather than on depicting small lesions. Models falling within the purview of explainable AI include gradient-weighted class activation mapping (Grad-CAM), occlusion sensitivity, and LIME [18]. Grad-CAM visualizes important pixels by weighting the gradient against the predicted value, which is useful for image classification [45]. Grad-CAM was not used in this study owing to its low spatial resolution when assessing the presence of a disease. In the present study, CuTS prediction was based on a detailed comparison with a control group using occlusion sensitivity and LIME. Occlusion sensitivity can effectively visualize multifocal glass opacities and consolidations, allowing detailed visualizations of important image features [23], whereas LIME extracts important regions by creating superpixels and has attracted significant attention in the field of medical imaging [46]. These visualization tools may provide the necessary information for diagnosis by highlighting the basis for predictions made by the DL model. In this study, the DL model focused on hyperechoic changes in the ulnar nerve epithelium, and the hypoechoic changes in the ulnar nerve interior and surrounding tissue, for feature visualization. The model successfully recognized the nerve swelling and edema of the surrounding tissue from US images. Thus, it is possible to diagnose CuTS with high accuracy without measuring CSA. Furthermore, because DL models learn image features using an uninhibited and unbiased neural model compared with humans, DL feature visualizations may enable physicians to detect previously overlooked and unquantified features. Early CuTS diagnosis is important as delays can lead to muscle weakness and atrophy. The clinical application of DL for US imaging may lead to early CuTS diagnosis, and this system is expected to expand in the future.

This study has the following limitations. First, the results were obtained exclusively with US images, and the generalization of the performance of the models on images from other diagnostic devices was not investigated. Second, no accuracy comparison with other diagnostic devices was performed. Moreover, although many images were reviewed, the number of cases considered was not large. Further research is required to corroborate the results of this study. Finally, in this study, the same US device was used in every phase, and no investigation was conducted regarding changes in the environment, such as the addition of noise. The same image may not be obtained even when looking at the same object because the settings of the US device and the examiner's skill may vary. Although multiple patients were examined experimentally using the same US device from different angles by tilting, sliding, and rotating the probe under the same gain and focus conditions, the accuracy of the measurements should be verified by adding data obtained from different instruments to improve generalizability.

## 5. Conclusions

In this study, US images of CuTS regions of interest were evaluated using AI-based medical imaging technology. US images using the DL model predicted CuTS with high

accuracy, with the best accuracy score of 0.90 for ResNet-50 and MobileNet\_v2, the best precision score of 0.86 for ResNet-50, the best sensitivity score of 1.00 for all models, and the best F-measure score of 0.92 for ResNet-50 and MobileNet\_v2. Furthermore, all models successfully predicted CuTS by focusing on the neural interior and perineural tissue of the ulnar nerve. The proposed method can be improved in the future to provide a more convenient and accurate CuTS diagnosis using US imaging, which is minimally invasive, low cost, and excellent for soft tissue examinations. We expect DL-based systems to make the diagnosis of CuTS easier and more accurate.

**Author Contributions:** Conceptualization, I.S. and A.I.; methodology, Y.M.; software, H.N., T.K., and T.F.; validation, K.Y., S.M., T.Y., and Y.H.; formal analysis, I.S.; investigation, I.S.; resources, A.I.; data curation, I.S.; writing—original draft preparation, I.S.; writing—review and editing, A.I. and Y.M.; visualization, A.I. and T.M.; supervision, Y.M.; project administration, R.K.; funding acquisition, R.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of Kobe University Graduate School of Medicine (No. B21009, 21 April 2021).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patients to publish this paper.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to confidentiality concerns.

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

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Staples, J.R.; Calfee, R. Cubital tunnel syndrome: Current concepts. *J. Am. Acad. Orthop. Surg.* **2017**, *25*, e215–e224. [[CrossRef](#)] [[PubMed](#)]
2. Latinovic, R.; Gulliford, M.C.; Hughes, R.A. Incidence of common compressive neuropathies in primary care. *J. Neurol. Neurosurg. Psychiatry* **2006**, *77*, 263–265. [[CrossRef](#)] [[PubMed](#)]
3. Nakashian, M.N.; Ireland, D.; Kane, P.M. Cubital tunnel syndrome: Current concepts. *Curr. Rev. Musculoskelet. Med.* **2020**, *13*, 520–524. [[CrossRef](#)] [[PubMed](#)]
4. Naran, S.; Imbriglia, J.E.; Bilonick, R.A.; Taieb, A.; Wollstein, R. A demographic analysis of cubital tunnel syndrome. *Ann. Plast. Surg.* **2010**, *64*, 177–179. [[CrossRef](#)] [[PubMed](#)]
5. Schnabl, S.M.; Kisslinger, F.; Schramm, A.; Dragu, A.; Kneser, U.; Unglaub, F.; Horch, R.E. Objective outcome of partial medial epicondylectomy in cubital tunnel syndrome. *Arch. Orthop. Trauma Surg.* **2010**, *130*, 1549–1556. [[CrossRef](#)]
6. American Association of Electrodiagnostic Medicine; Campbell, W.W. Guidelines in electrodiagnostic medicine. Practice parameter for electrodiagnostic studies in ulnar neuropathy at the elbow. *Muscle Nerve Suppl.* **1999**, *8*, S171–S205.
7. Vucic, S.; Cordato, D.J.; Yiannikas, C.; Schwartz, R.S.; Shnier, R.C. Utility of magnetic resonance imaging in diagnosing ulnar neuropathy at the elbow. *Clin. Neurophysiol.* **2006**, *117*, 590–595. [[CrossRef](#)]
8. Yoon, J.S.; Walker, F.O.; Cartwright, M.S. Ulnar neuropathy with normal electrodiagnosis and abnormal nerve ultrasound. *Arch. Phys. Med. Rehabil.* **2010**, *91*, 318–320. [[CrossRef](#)]
9. Chen, I.J.; Chang, K.V.; Wu, W.T.; Ozcarar, L. Ultrasound parameters other than the direct measurement of ulnar nerve size for diagnosing cubital tunnel syndrome: A systemic review and meta-analysis. *Arch. Phys. Med. Rehabil.* **2019**, *100*, 1114–1130. [[CrossRef](#)]
10. Chang, K.V.; Wu, W.T.; Han, D.S.; Ozcarar, L. Ulnar nerve cross-sectional area for the diagnosis of cubital tunnel syndrome: A meta-analysis of ultrasonographic measurements. *Arch. Phys. Med. Rehabil.* **2018**, *99*, 743–757. [[CrossRef](#)]
11. Kalia, V.; Jacobson, J.A. Imaging of Peripheral Nerves of the Upper Extremity. *Radiol. Clin. N. Am.* **2019**, *57*, 1063–1071. [[CrossRef](#)] [[PubMed](#)]
12. D'Souza, J.C.; Sultan, L.R.; Hunt, S.J.; Schultz, S.M.; Brice, A.K.; Wood, A.K.W.; Sehgal, C.M. B-mode ultrasound for the assessment of hepatic fibrosis: A quantitative multiparametric analysis for a radiomics approach. *Sci. Rep.* **2019**, *9*, 8708. [[CrossRef](#)] [[PubMed](#)]
13. Weston, A.D.; Korfiatis, P.; Kline, T.L.; Philbrick, K.A.; Kostandy, P.; Sakinis, T.; Sugimoto, M.; Takahashi, N.; Erickson, B.J. Automated abdominal segmentation of CT Scans for body composition analysis using deep learning. *Radiology* **2019**, *290*, 669–679. [[CrossRef](#)] [[PubMed](#)]

14. Shin, H.C.; Roth, H.R.; Gao, M.; Lu, L.; Xu, Z.; Nogues, I.; Yao, J.; Mollura, D.; Summers, R.M. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imaging* **2016**, *35*, 1285–1298. [[CrossRef](#)] [[PubMed](#)]
15. Apostolopoulos, I.D.; Mpesiana, T.A. Covid-19: Automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Phys. Eng. Sci. Med.* **2020**, *43*, 635–640. [[CrossRef](#)]
16. Olczak, J.; Fahlberg, N.; Maki, A.; Razavian, A.S.; Jilert, A.; Stark, A.; Skoldenberg, O.; Gordon, M. Artificial intelligence for analyzing orthopedic trauma radiographs. *Acta Orthop.* **2017**, *88*, 581–586. [[CrossRef](#)]
17. Ma, J.; Wu, F.; Zhu, J.; Xu, D.; Kong, D. A pre-trained convolutional neural network based method for thyroid nodule diagnosis. *Ultrasonics* **2017**, *73*, 221–230. [[CrossRef](#)]
18. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Vegas, NV, USA, 27–30 June 2016; pp. 2921–2929.
19. Lalehzarian, S.P.; Gowd, A.K.; Liu, J.N. Machine learning in orthopaedic surgery. *World J. Orthop.* **2021**, *12*, 685–699. [[CrossRef](#)]
20. Xue, Y.; Zhang, R.; Deng, Y.; Chen, K.; Jiang, T. A preliminary examination of the diagnostic value of deep learning in hip osteoarthritis. *PLoS ONE* **2017**, *12*, e0178992. [[CrossRef](#)]
21. Ashinsky, B.G.; Bouhrara, M.; Coletta, C.E.; Lehallier, B.; Urish, K.L.; Lin, P.C.; Goldberg, I.G.; Spencer, R.G. Predicting early symptomatic osteoarthritis in the human knee using machine learning classification of magnetic resonance images from the osteoarthritis initiative. *J. Orthop. Res.* **2017**, *35*, 2243–2250. [[CrossRef](#)]
22. Shubert, D.J.; Prud'homme, J.; Sraj, S. Nerve conduction studies in surgical cubital tunnel syndrome patients. *Hand* **2021**, *16*, 170–173. [[CrossRef](#)] [[PubMed](#)]
23. Aminu, M.; Ahmad, N.A.; Mohd Noor, M.H. COVID-19 detection via deep neural network and occlusion sensitivity maps. *Alex. Eng. J.* **2021**, *60*, 4829–4855. [[CrossRef](#)]
24. Srinivas, S. A Machine Learning-Based Approach for Predicting Patient Punctuality in Ambulatory Care Centers. *Int. J. Environ. Res. Public Health* **2020**, *17*, 3703. [[CrossRef](#)]
25. Matsuo, H.; Nishio, M.; Kanda, T.; Kojita, Y.; Kono, A.K.; Hori, M.; Teshima, M.; Otsuki, N.; Nibu, K.I.; 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.* **2020**, *10*, 19388. [[CrossRef](#)] [[PubMed](#)]
26. Padua, L.; Coraci, D.; Erra, C.; Pazzaglia, C.; Paolasso, I.; Loreti, C.; Caliandro, P.; Hobson-Webb, L.D. Carpal tunnel syndrome: Clinical features, diagnosis, and management. *Lancet Neurol.* **2016**, *15*, 1273–1284. [[CrossRef](#)]
27. Ng, A.W.H.; Griffith, J.F.; Lee, R.K.L.; Tse, W.L.; Wong, C.W.Y.; Ho, P.C. Ultrasound carpal tunnel syndrome: Additional criteria for diagnosis. *Clin. Radiol.* **2018**, *73*, 214.e11–214.e18. [[CrossRef](#)]
28. Kerasnoudis, A.; Tsvigoulis, G. Nerve Ultrasound in Peripheral Neuropathies: A Review. *J. Neuroimaging* **2015**, *25*, 528–538. [[CrossRef](#)]
29. Tai, T.W.; Wu, C.Y.; Su, F.C.; Chern, T.C.; Jou, I.M. Ultrasonography for diagnosing carpal tunnel syndrome: A meta-analysis of diagnostic test accuracy. *Ultrasound Med. Biol.* **2012**, *38*, 1121–1128. [[CrossRef](#)]
30. Hadjiiski, L.M.; Tourassi, G.D.; Bar, Y.; Diamant, I.; Wolf, L.; Greenspan, H. Deep learning with non-medical training used for chest pathology identification. In Proceedings of the SPIE 9414, Medical Imaging 2015: Computer-Aided Diagnosis, Orlando, FL, USA, 22–25 February 2015; Volume 94140.
31. Jimenez-Sanchez, A.; Kazi, A.; Albarqouni, S.; Kirchhoff, C.; Biberthaler, P.; Navab, N.; Kirchhoff, S.; Mateus, D. Precise proximal femur fracture classification for interactive training and surgical planning. *Int. J. Comput. Assist. Radiol. Surg.* **2020**, *15*, 847–857. [[CrossRef](#)]
32. Kalmet, P.H.S.; Sanduleanu, S.; Primakov, S.; Wu, G.; Jochems, A.; Refaee, T.; Ibrahim, A.; Hulst, L.V.; Lambin, P.; Poeze, M. Deep learning in fracture detection: A narrative review. *Acta Orthop.* **2020**, *91*, 215–220. [[CrossRef](#)]
33. Tomita, N.; Cheung, Y.Y.; Hassanpour, S. Deep neural networks for automatic detection of osteoporotic vertebral fractures on CT scans. *Comput. Biol. Med.* **2018**, *98*, 8–15. [[CrossRef](#)] [[PubMed](#)]
34. Smistad, E.; Johansen, K.F.; Iversen, D.H.; Reinertsen, I. Highlighting nerves and blood vessels for ultrasound-guided axillary nerve block procedures using neural networks. *J. Med. Imaging* **2018**, *5*, 044004. [[CrossRef](#)] [[PubMed](#)]
35. Jabeen, K.; Khan, M.A.; Alhaisoni, M.; Tariq, U.; Zhang, Y.D.; Hamza, A.; Mickus, A.; Damasevicius, R. Breast Cancer Classification from Ultrasound Images Using Probability-Based Optimal Deep Learning Feature Fusion. *Sensors* **2022**, *22*, 807. [[CrossRef](#)] [[PubMed](#)]
36. Tsai, K.J.; Chou, M.C.; Li, H.M.; Liu, S.T.; Hsu, J.H.; Yeh, W.C.; Hung, C.M.; Yeh, C.Y.; Hwang, S.H. A High-Performance Deep Neural Network Model for BI-RADS Classification of Screening Mammography. *Sensors* **2022**, *22*, 1160. [[CrossRef](#)]
37. Meraj, T.; Alosaimi, W.; Alouffi, B.; Rauf, H.T.; Kumar, S.A.; Damasevicius, R.; Alyami, H. A quantization assisted U-Net study with ICA and deep features fusion for breast cancer identification using ultrasonic data. *PeerJ Comput. Sci.* **2021**, *7*, e805. [[CrossRef](#)]
38. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385.
39. Zhang, L.; Li, M.; Zhou, Y.; Lu, G.; Zhou, Q. Deep Learning Approach for Anterior Cruciate Ligament Lesion Detection: Evaluation of Diagnostic Performance Using Arthroscopy as the Reference Standard. *J. Magn. Reson. Imaging* **2020**, *52*, 1745–1752. [[CrossRef](#)]

40. Zhang, H.; Han, L.; Chen, K.; Peng, Y.; Lin, J. Diagnostic Efficiency of the Breast Ultrasound Computer-Aided Prediction Model Based on Convolutional Neural Network in Breast Cancer. *J. Digit. Imaging* **2020**, *33*, 1218–1223. [[CrossRef](#)]
41. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. *arXiv* **2019**, arXiv:1801.04381.
42. 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*, 4304–4315. [[CrossRef](#)]
43. Wang, J.; Liu, Q.; Xie, H.; Yang, Z.; Zhou, H. Boosted EfficientNet: Detection of Lymph Node Metastases in Breast Cancer Using Convolutional Neural Networks. *Cancers* **2021**, *13*, 661. [[CrossRef](#)] [[PubMed](#)]
44. Tan, M.; Le, Q.V. EfficientNet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
45. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 618–626. [[CrossRef](#)]
46. Ahsan, M.M.; 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* **2021**, *9*, 1099. [[CrossRef](#)] [[PubMed](#)]