

# Fully-Automated Spike Detection and Dipole Analysis of Epileptic MEG Using Deep Learning

Ryoji Hirano<sup>1</sup>, Takuto Emura, Otoichi Nakata, Toshiharu Nakashima, Miyako Asai<sup>2</sup>, Kuriko Kagitani-Shimono, Haruhiko Kishima<sup>3</sup>, and Masayuki Hirata

**Abstract**—Magnetoencephalography (MEG) is a useful tool for clinically evaluating the localization of interictal spikes. Neurophysiologists visually identify spikes from the MEG waveforms and estimate the equivalent current dipoles (ECD). However, presently, these analyses are manually performed by neurophysiologists and are time-consuming. Another problem is that spike identification from MEG waveforms largely depends on neurophysiologists' skills and experiences. These problems cause poor cost-effectiveness in clinical MEG examination. To overcome these problems, we fully automated spike identification and ECD estimation using a deep learning approach fully automated AI-based MEG interictal epileptiform discharge identification and ECD estimation (FAMED). We applied a semantic segmentation method, which is an image processing technique, to identify the appropriate times between spike onset and peak and to select appropriate sensors for ECD estimation. FAMED was trained and evaluated using clinical MEG data acquired from 375 patients. FAMED training was performed in two stages: in the first stage, a classification network was

learned, and in the second stage, a segmentation network that extended the classification network was learned. The classification network had a mean AUC of 0.9868 (10-fold patient-wise cross-validation); the sensitivity and specificity were 0.7952 and 0.9971, respectively. The median distance between the ECDs estimated by the neurophysiologists and those using FAMED was 0.63 cm. Thus, the performance of FAMED is comparable to that of neurophysiologists, and it can contribute to the efficiency and consistency of MEG ECD analysis.

**Index Terms**—Magnetoencephalography (MEG), epilepsy, spike detection, dipole analysis, deep learning.

## I. INTRODUCTION

MAGNETOENCEPHALOGRAPHY (MEG) is used as a clinical epilepsy examination to evaluate the localization of epileptic activity. MEG is a useful technique for evaluating pre-epileptic surgery because it can evaluate the epileptic focus non-invasively. To determine the epileptic focus, an equivalent current dipole (ECD) estimation method is commonly used. ECD estimates the characteristic signal source by solving the inverse problem [1]. ECD estimation is widely recognized as an effective method to determine the epileptic focus [2]–[4]. The general ECD analysis flow is as follows: first, visual detection of the interictal spikes, then the estimation of the ECD based on the rise of the spike waves, and finally, evaluation of the applicability of epilepsy surgery based on the localization of the ECD cluster formed by performing this analysis on multiple spikes.

However, visual inspection of spikes and ECD analyses are time-consuming [5]. Additionally, visual identification of spikes may vary among neurophysiologists. Automation of these procedures is time-saving and contributes to the consistency of the quality of results [6]. Many automated spike detection algorithms have been proposed for electroencephalography (EEG) [5]. Based on the recent progress in artificial intelligence (AI), AI-based high-performance detection algorithms using deep learning, such as convolutional neural networks (CNNs), have also been proposed [7]–[10]. Conversely, there have been few reports on MEG, including methods other than deep learning [11]–[14]. Moreover, all of these EEG and MEG studies have focused on detecting spikes but never included other algorithms necessary to complete ECD analysis, including not only detection of spikes but also determining the optimal time between spike onset and peak,

Manuscript received 14 February 2022; revised 25 April 2022; accepted 2 May 2022. Date of publication 10 May 2022; date of current version 30 September 2022. This work was supported in part by the Collaborative Grant Between RICOH Corporation and the Department of Neurological Diagnosis and Restoration (NDR), Graduate School of Medicine, Osaka University; in part by the Japan Agency for Medical Research and Development (AMED) under Grant BrainMind Beyond 19dm0307103; and in part by the Grants-in-Aid for Scientific Research from KAKENHI under Grant JP18H04085. (Corresponding authors: Ryoji Hirano; Masayuki Hirata.)

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Review Committee of the Osaka University Hospital under Application No. 18329-3.

Ryoji Hirano is with CoMIT, Department of Neurological Diagnosis and Restoration, Graduate School of Medicine, Osaka University, Suita, Osaka 565-0871, Japan, and also with Ricoh, Osaka 563-8501, Japan (e-mail: hirano@ndr.med.osaka-u.ac.jp).

Takuto Emura and Masayuki Hirata are with CoMIT, Department of Neurological Diagnosis and Restoration, and the Department of Neurosurgery, Graduate School of Medicine, Osaka University, Suita, Osaka 565-0871, Japan (e-mail: t-emura@nsurg.med.osaka-u.ac.jp; mhirata@ndr.med.osaka-u.ac.jp).

Otoichi Nakata, Toshiharu Nakashima, and Miyako Asai are with Ricoh, Kanagawa 243-0460, Japan (e-mail: otoichi.nakata@jp.ricoh.com; toshiharu.nakashima@jp.ricoh.com; miyako.asai@jp.ricoh.com).

Kuriko Kagitani-Shimono is with the United Graduate School of Child Development, Osaka University, Suita 565-0871, Japan (e-mail: kuriko@ped.med.osaka-u.ac.jp).

Haruhiko Kishima is with the Department of Neurosurgery, Graduate School of Medicine, Osaka University, Suita, Osaka 565-0871, Japan (e-mail: hkishima@nsurg.med.osaka-u.ac.jp).

Digital Object Identifier 10.1109/TMI.2022.3173743

selecting the sensors used for ECD analysis, and quantitative evaluation of the cluster size of ECDs. Importantly, an appropriate selection of sensors is recommended before estimating the ECD [15], [16]. To the best of our knowledge, there is only one paper mentioning the feasibility of automated ECD analysis using independent component analysis (ICA) [17]. To date, there are no published papers that used AI for fully automated ECD analysis. However, full automation is clinically indispensable in terms of the efficiency of MEG examination.

Here, we hypothesized that machine learning methods, including deep learning, can learn more spike patterns than other methods for spike detection using ICA and threshold processing [17] and thus can improve performance. Therefore, we developed an AI-based fully automated spike detection and ECD analysis of clinical MEG examination (Fully Automated AI-based MEG Interictal Epileptiform Discharge Identification and Dipole Estimation: FAMED) for epilepsy that is time-saving and may contribute to quality control. In this study, we propose a novel network applying semantic segmentation for automated ECD analysis. Semantic segmentation is an image processing task that classifies each pixel of an image into a class [18], [19], and it is applied to time-series anomaly detection [20]. We trained the network for time series semantic segmentation using as many as 406 clinical MEG datasets for epilepsy with ECDs analyzed by neurophysiologists. As a result, the proposed deep neural network detected spikes with high specificity and accurately identified spike timings. Finally, we discuss the feasibility of the fully automated ECD analysis using the proposed network.

## II. MATERIALS AND METHODS

### A. MEG Data and Pre-Processing

MEG data used in this study were recorded as a clinical examination of epilepsy at Osaka University Hospital from 2010–2019. This study was approved by the Ethics Review Committee of the Osaka University Hospital (approval number: 18329–3). A total of 516 MEG examinations related to epilepsy were recorded during this period. All these data were analyzed using ECD methods by neurophysiologists engaged in clinical cases of epilepsy whose years of experience of epilepsy varied from one year to more than 20 years. In cases where the years of clinical experience of epilepsy were less than three years, their analysis results were reviewed by expert neurophysiologists with more than 10 years of clinical experience. In addition, all of the MEG data was reviewed using clinical findings by an expert neurophysiologist (MH) having 24 years of experience with epilepsy MEG prior to being included in this study. As a result, a total of 406 MEG datasets (age: 0–79 years, median value: 18; 197 females and 209 males) were used in this study. The summary of 406 patients was shown in TABLE I. The type of disease was extracted based on the information provided in the medical records. As MEG measurements were performed mainly to evaluate the location of epileptic activity, focal epilepsy cases were common. In particular, localization-related epilepsies such as temporal lobe and frontal lobe epilepsy were the most

TABLE I  
CHARACTERISTICS OF PATIENTS USED IN TRAINING AND TEST

		Total N=406	Male N=209	Female N=197
Age	0 to <1	6(1.5%)	2(1%)	4(2%)
	1 to <5	40(9.9%)	24(11.6%)	16(8.1%)
	5 to <13	95(23.4%)	48(26.2%)	47(23.9%)
	13 to <18	60(14.8%)	33(19.2%)	27(13.7%)
	18 to <30	86(21.4%)	46(29.5%)	40(20.3%)
	30 to <50	94(23.2%)	48(42.5%)	46(23.4%)
	50 to <65	18(4.2%)	7(10.8%)	11(5.6%)
	65 to <75	6(1.5%)	0(0%)	6(3%)
	≥ 75	1(0.2%)	1(1.1%)	0(0%)
Type of disease	Localization-related epilepsy			
	Temporal lobe epilepsy	87(21.4%)	37(17.7%)	50(25.4%)
	Frontal lobe epilepsy	33(8.1%)	17(8.1%)	16(8.1%)
	Others	10(2.5%)	3(1.4%)	7(3.6%)
	Congenital brain malformation			
	Tuberous sclerosis	24(5.9%)	16(7.7%)	8(4.1%)
	Cortical dysplasia	21(5.2%)	13(6.2%)	8(4.1%)
	Others	9(2.2%)	4(1.9%)	5(2.5%)
	Age-related epilepsy			
	West syndrome	26(6.4%)	10(4.8%)	16(8.1%)
	Lennox-Gastaut syndrome	11(2.7%)	6(2.9%)	5(2.5%)
	Encephalitis	24(5.9%)	16(7.7%)	8(4.1%)
	Brain tumor	18(4.4%)	10(4.8%)	8(4.1%)
	Vascular malformation	7(1.7%)	3(1.4%)	4(2%)
	Epilepsy suspected	79(19.5%)	46(22%)	33(16.8%)
	Non epileptic seizures	14(3.4%)	8(3.8%)	6(3%)
Others with few cases	43(10.6%)	20(9.6%)	23(11.7%)	

common, accounting for 32% of the cases, whereas generalized epilepsies such as West syndrome and Lennox–Gastaut syndrome (LGS) accounted for 9.1%. Thirty-one patients were measured multiple times over several years, so the total number of patients was 375. Eighty-six cases out of these data included no spikes, whereas 320 cases included at least one spike and its corresponding ECD. The age distribution of patients was biased towards the younger age group. To train more diverse waveforms, data selection by epilepsy subtype, age, sleeping conditions, or medication conditions were not performed.

The spontaneous-state MEG data were recorded using a whole-head MEG equipped with 160 axial gradiometers housed in a magnetically shielded room (MEGvision NEO; Yokogawa Electric Corporation, Kanazawa, Japan). The measurement conditions were as follows: sampling frequency of 1000 Hz or 2000 Hz, low pass filter at 100 Hz, 200 Hz or 500 Hz, high pass filter at 0.1 Hz, notch filter at 60 Hz, or nothing. The duration of one recording session was either 240 s or 300 s, and multiple sessions were carried out. During measurement, patients were instructed to keep resting, but child patients and patients with mental retardation were not instructed to do so. Instead, child patients were sedated by intravenous administration of hydroxyzine pamoate and pentazocine in addition to orally administration of triclofos sodium or thiopental sodium.

Each data set was registered to the training data set through preprocessing and window extraction. A bandpass filter with a 3 Hz and 35 Hz cutoff was applied to all data. Down sampling from 2000 Hz to 1000 Hz was applied only for the 2000 Hz data. This preprocessing was performed to

normalize data and remove noise. The cutoff frequency of the bandpass filter was used in routine analysis of Osaka University Hospital and did not immoderately deviate from the guideline [15]. For data marked as spikes, in other words, data with added ECD estimation results, a 2048 ms window (estimated ECD time  $\pm 1024$  ms) was extracted and registered in the spike positive dataset. If ECD analysis was performed on another spike during this approximately 2-second window, that spike was excluded from the training data. This was done to avoid duplication of training data. Simultaneously, a spike mask as the ground truth of the same window size was created to train the segmentation network. The mask was registered 60 ms before and after the ECD estimated time in the sensor selected during estimation as 1 and sets other than that as 0. This number was calculated from the duration of spikes in MEG [21]. Furthermore, we extracted a window of the same length as the spike-positive data approximately every 3 seconds from the data where no spikes were detected. These data were registered as the spike-negative dataset. The MEG system used in the present study was designed to insert 160 against 234 sensor holders. There were different types of MEG systems with respect to the number of sensors and sensor locations. To apply the same model to different types of sensor configurations for general purposes, we expanded the number of sensors of the training data to 234 and zero paddings were applied to the gaps. before window extraction. Finally, the neurophysiologist (MH) visually reviewed all training data and excluded atypical spikes (e.g., slow waves, small spikes, sharp waves, spikes contaminated by artifacts and spikes associated with seizures). This means that only typical spike waves were used as training data.

## B. MEG Spike Detection Algorithm

We applied a semantic segmentation method, which is an image processing technique, to identify the appropriate times between spike onset and peak and to select appropriate sensors for ECD estimation. An overall picture of the learning process is shown in Fig. 1. The weight of the classification network learned beforehand was transferred to the segmentation network for detecting spikes, as shown in Figs. 1(b) and 1(c). We trained the classification network to improve the learning efficiency of the segmentation network. We also expected improved performance using transfer weight. In detail, the classification network is an application of the 26-layer SE-ResNet [22]. We changed the convolution layer used in the ResBlock to a dilated convolution layer [23] for the time dimension, and the SE-module was changed to a scSE-module [24]. We aim to improve the segmentation performance because of the extraction of the proper time and spatial features. The DropConnect layer [25] was added to the ResBlock to improve the generalization performance. The segmentation network used in the present study was designed based on DeepUNet [19], which has an encoder–decoder structure and the same structure for encoders as the classification network. The implementation of each model is publicly available online via CodeOcean.<sup>1</sup>

Two types of augmentation were applied to train the two networks: (1) random cropping of a 1024 ms window from a 2048 ms window of each dataset to avoid time dependency of the spike, and (2) random sensor sorting to avoid sensor dependency of the spike. We applied these two augmentations under constant probabilities (80% and 20%, respectively). We normalized the magnetic flux density by setting the mean to 0 and the standard deviation to 1 for the 1024 ms window.

We applied the segmentation network to the continuous measurement data using a sliding window to create a confidence map of the same size as that of the data. The confidence map represented the probability of spike existence and was created by applying the sigmoid function to the output of the segmentation network. This confidence map was used to identify the appropriate time and sensors for ECD analysis. Appropriate sensor selection for ECD estimation improves the signal-to-noise ratio and helps the ECD to best describe the observed signal [15]. Here, we describe the mathematical basis of ECD estimation using a spike confidence map. We considered the automated analysis of the measurement data of  $N_t \in \mathbb{N}$  (time) points of  $N_s \in \mathbb{N}$  (sensors). We binarized the confidence map  $\mathbf{X} \in \mathbb{R}^{N_s \times N_t}$  by means of thresholding:

$$\mathbf{X}_b = \begin{cases} 1, & x_{ij} > x_0 \\ 0, & x_{ij} \leq x_0 \end{cases}, \quad (1)$$

where  $x_{ij}$  is  $j$ -th confidence value in  $i$ -th sensor and  $x_0$  is the threshold about the confidence value. The binarized map is then summed for each sensor to obtain a 1D confidence map:

$$w_j = \sum_{i=1}^{N_s} (\mathbf{X}_b)_{ij} \quad (2)$$

where  $(\mathbf{X}_b)_{ij}$  is the  $j$ -th value in the  $i$ -th sensor of  $\mathbf{X}_b$ . A Gaussian filter was applied to this 1D map  $w_j$  to obtain the candidate spike time:

$$z_j = f(w_j) \quad (3)$$

where  $f$  is the Gaussian filter with kernel size = 120 ms,  $\sigma = 15$  ms. Spike candidates  $T_c$  were determined by detecting the peaks of  $z_j$ . The spike times were determined by thresholding the 1D map  $w_j$  for each candidate time:

$$T_{\text{spike}} = \{t | w_t > w_0, t \in T_c\} \quad (4)$$

where  $w_0$  is the threshold for the number of sensors for which spikes are observed. At the same time, the sensors were selected from the binary confidence map. For example, the sensor selection for the  $k$ -th time of candidate times  $T_{\text{spike}}$  is shown below:

$$S_k = \{i | (\mathbf{X}_b)_{ik} = 1, i \in \{1, 2, \dots, N_s\}\} \quad (5)$$

Thus, we obtained the time  $T_{\text{spike}}$  and the selected sensor  $S_k$  for ECD estimation by processing the output of the segmentation network as described earlier. Using these post-processing techniques, ECD analysis was performed fully-automatically.

<sup>1</sup>[Online]. Available: <https://codeocean.com/capsule/4434883/tree/v1>



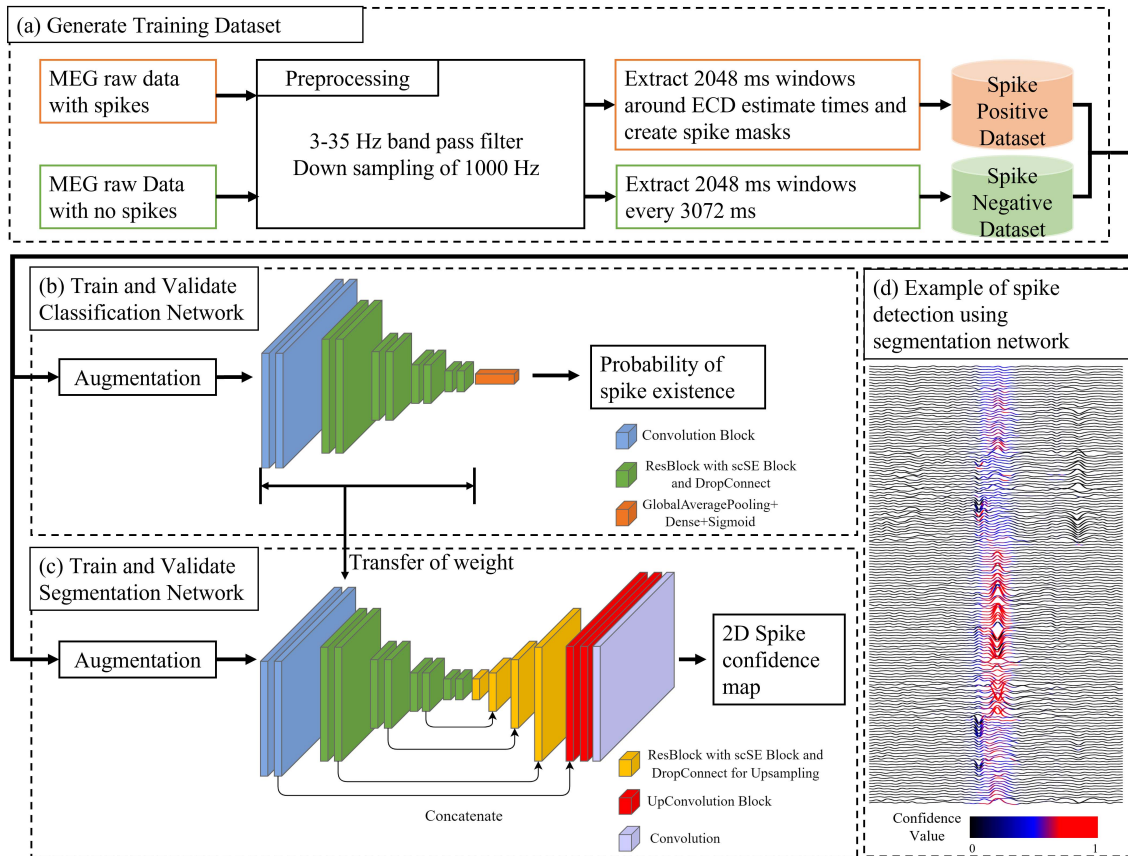


Fig. 1. Overview of training process. First, a training dataset was created (a). In this part, Meg raw data were preprocessed for normalization and removal of noise. Moreover, spike or not-spike labels for classification and spike masks for segmentation were created simultaneously when extracting time windows. Second, the classification network was trained (b). We designed an architecture based on SE-ResNet [22] for the classification network. Finally, the segmentation network was trained by transferring the weight in the classification network (c). We designed an architecture based on DeepUNet [19] regarding the segmentation network. An example of the spike detection result using the segmentation network is shown (d). In this part, 2D spike confidence map, the output of the segmentation network was superimposed on the waveform.

### C. Training and Evaluation Process

We used 5-fold and 10-fold patient-wise cross-validation for training and evaluating both the classification and segmentation networks. We evaluated this method in two steps to reduce the time taken for evaluation. First, we used 5-fold patient-wise cross-validation for preliminary evaluation of the model. Then, we used 10-fold patient-wise cross-validation for the final evaluation of the method. In this cross-validation in a previous study [9], the dataset was divided into groups. Then, the data derived from a unique patient were registered into one of train, validation, or test data, and the machine learning process was run iteratively with different training, validation, and test data. The training dataset was used to optimize the model parameters, the validation dataset for tuning and early stopping of hyperparameters, and the test dataset to evaluate the model. We adopted this validation method to evaluate the generalized performance correctly regarding unseen new patients. There were individual differences in the medical data. If the data from a unique patient were divided into train, validation, and test, the prediction performance would have been overestimated.

We used three metrics for evaluating the classification network: (1) the area under the curve (AUC) of the receiver operating characteristic (ROC) curve, (2) sensitivity, and

(3) specificity for the classification network. We used the intersection over union (IoU) for the segmentation network. The IoU value is the area of overlap between the predicted spike area and the ground truth divided by the area of union between the predicted spike area and the ground truth.

In addition, we used continuous measurement data to evaluate the time points of the detected spikes and the positions of the estimated ECDs in comparison to the neurophysiologist's analysis. To evaluate the detected time points of spikes, we counted auto-detected spike times for 120 ms before and after the spike time detected by the neurophysiologists, based on the method of Ossadtchi *et al.* [17]. To evaluate the positions of the ECDs, we measured the distance between the location of the ECD that was automatically estimated and that of the nearest ECD that was manually estimated by the neurophysiologists. ECD clustering was also applied during the evaluation of the ECD positions, in line with the typical analysis. The clustering method employed was DBSCAN [26], which can determine outliers according to the density of ECDs. We obtained clusters of ECDs estimated by the neurophysiologists and those by FAMED and performed the same evaluation as described above in pairs of clusters where the distance between cluster centers was below the threshold.



TABLE II  
THE EFFECTIVENESS OF EACH METHOD USED IN THIS STUDY

Method	Classification Network Metrics			Segmentation Network Metric
	AUC	Sensitivity	Specificity	IoU
SE-ResNet	0.9786 ± 0.0119	0.6628 ± 0.0633	0.9949 ± 0.0073	0.3670 ± 0.0278
+sSE module	0.9758 ± 0.0092	0.6747 ± 0.0415	<b>0.9960 ± 0.0022</b>	0.3959 ± 0.0656
+dilated convolution	0.9794 ± 0.0089	0.7481 ± 0.0779	0.9905 ± 0.0104	0.4298 ± 0.0387
+DropConnect	<b>0.9804 ± 0.0081</b>	<b>0.7542 ± 0.0592</b>	0.9924 ± 0.0070	0.4352 ± 0.0409
+Transfer Weight	N/A	N/A	N/A	<b>0.4534 ± 0.0118</b>

### III. RESULTS

#### A. Datasets, Training Parameters and Environment

The number of the spike positive data was 5,401 and that of the spike negative data was 17,776. Thus, the total number of training data points was 23,177. The average number of spikes per case was 20 (minimum: 1, maximum: 103). The number of patients from the spike positive dataset and that from the spike negative dataset were 268 and 80, respectively. We trained the classification network and segmentation network using these data. The conditions for training the classification network were as follows: the initialization of the parameters was based on He's algorithm [27], the maximum number of epochs was 50, early stopping was applied when an epoch was over 10, the initial learning rate was 1e-4, the learning rate reduced on plateau strategy was used, and the optimizer was based on AdamW [28] with a batch size of 64 (parameter beta\_1, beta\_2, epsilon, and weight decay were 0.9, 0.999, 1e-8, and 5e-4). The conditions for training the segmentation network were as follows: gradient accumulation was applied every 2 epochs, the batch size was 16, the maximum number of epochs was 60, early stopping was applied when an epoch was over 20, and the other hyperparameters were the same as the classification network. Regarding the loss function of the training procedure, we used the focal loss [29]. We fixed the parameters of the encoder of the segmentation network using the first five epochs. We determined these hyperparameters using the automated hyperparameter tuning tool *optuna* [30]. All implementations were performed in Python using Ubuntu 18.04.3 LTS machine with an Intel Core i9-9900K CPU clocked at 3.60 GHz with 64 GB of RAM with NVIDIA GeForce RTX 2080 Ti GPU.

#### B. Classification and Segmentation Performance

First, the effectiveness of the proposed architecture was evaluated using 5-fold patient-wise cross-validations. We compared the performance of the basic SE-ResNet by adding the sSE module, changing the convolution layer to dilated convolution, and applying DropConnect in order. (Adding the sSE module to SE-ResNet makes it a scSE-module.) For the segmentation network, we also investigated the effectiveness of transfer weight of the trained classification network. The evaluation results are shown in TABLE II. This table shows the statistics (mean and standard deviation) of each metric that was calculated for each fold. Regarding the segmentation metric, we used the median IoU value of spike-positive data in each

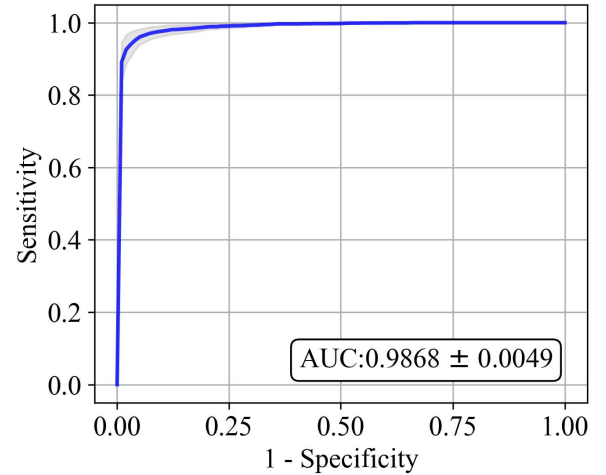


Fig. 2. ROC curve and its AUC. The mean value of the result of the 10-fold patient-wise cross-validation, and its standard deviation are indicated.

fold. The binarizing threshold of the classification network output was 0.7 (if output values were over 0.7, the input contained spikes). And the binarizing threshold of segmentation network output to obtain the IoU value was 0.6. The proposed architecture achieved the highest performance for all metrics except for specificity in the classification metrics. In addition, we showed that the proposed architecture could achieve higher performance by transfer weight.

Then, we evaluated the classification network of the 10-fold patient-wise cross-validation as shown in Fig. 2. The mean value and standard deviation of each fold were  $0.9868 \pm 0.0049$ , and the AUC of whole folds was 0.9888. When we used the same binarizing threshold value 0.7, the sensitivity and specificity were  $0.7952 \pm 0.0910$  and  $0.9971 \pm 0.0022$ , respectively. We compared our method with other recently published methods using deep learning proposed in previous studies: 2D-CNN [8], SpikeNet [9], and EMS-Net [14]. As there are differences in the modalities and the way the networks are applied, the networks were modified to fit to our dataset, but the concepts of those methods were kept intact to the extent possible. The results of the comparison are shown in TABLE III. This table shows the statistics (mean and standard deviation) of metrics that were calculated for each fold and patient. Our method achieved the best performance in AUC and sensitivity and the lowest variability both between folds and patients. Similarly, the segmentation result of the 10-fold patient-wise cross-validation was quantitatively evaluated using IoU values with the different thresholds of the confidence map.

TABLE III  
COMPARISON OF THE PERFORMANCE WITH PREVIOUS STUDIES REGARDING THE CLASSIFICATION OF MEG / EEG SIGNALS

Method	Modality	Params	Folds			Individuals	
			AUC	Sensitivity	Specificity	Sensitivity	Specificity
2D-CNN [8]	EEG	4.4M	0.9816±0.0173	0.5318±0.1634	<b>0.9988±0.0025</b>	0.4986±0.3693	<b>0.9956±0.0316</b>
SpikeNet [9]	EEG	9.1M	0.9862±0.0062	0.7772±0.0872	0.9970±0.0060	0.6921±0.3579	0.9947±0.0264
EMS-Net [14]	MEG	2.7M	0.9835±0.0076	0.6472±0.0583	0.9964±0.0060	0.6155±0.3427	0.9951±0.0222
Ours	MEG	1.2M	<b>0.9868±0.0049</b>	<b>0.7952±0.0910</b>	0.9971±0.0022	<b>0.7066±0.3341</b>	0.9955±0.0081

The IoU value = 1.0 means that the output of the segmentation network is completely matched to the segmentation mask. The median values of IoUs using practical thresholds 0.4, 0.5, and 0.6 were 0.464, 0.446, and 0.310, respectively. Two examples of segmentation results are shown in Fig. 3. The left waves show the original wave with the segmentation mask (green area). The right waves show the corresponding portion with the output of the segmentation network, where green colored sensors indicate those selected for ECD analysis by neurophysiologists. All spikes detected by AI for the test dataset were reviewed by an expert neurophysiologist (MH). This revealed that almost all of the false positive detection consisted of sharp waves, slow waves, or small spike-like waves (Fig. 3(c) and (d)). Moreover, the false positives did not include cardiac artifacts,  $\mu$  waves, humps, and sleep spindles of children. False negative detection included spikes with a rather small amplitude, which otherwise might be judged as small spike-like waves.

### C. FAMED Performance

A total of 268 clinical cases were used to evaluate ECD locations. For each case, we applied FAMED using the model trained with the dataset that did not contain its data. We applied the segmentation network to the time windows of 1024 ms, which were cut out every 256 ms. The outputs of the segmentation network from these overlapped time windows were averaged to create spike confidence maps for the entire measured data. By applying the process using formulas 1–5 to this map, the candidate time points and the sensors at the time points used for ECD analysis were selected. Here, we set the thresholding values in section II-B as  $x_0 = 0.6$ ,  $w_0 = 40$ . The sets of candidate timepoints and sensors obtained were used for subsequent automated ECD estimation. Regarding the automated ECD estimation, we mimicked the workflow of routine manual ECD analysis and applied the following process to exclude artifacts and false positives [31]. First, we performed ECD estimation at 15 timepoints before and after 7 ms of the detected candidate time point and excluded those ECDs with Goodness of Fit (GoF) < 0.95 and an ECD intensity smaller than 50 nAm, or larger than 400 nAm. When the distance between each remaining ECD position and its center of gravity did not exceed 2 cm, the point with the highest intensity among the candidate points was adopted as the point at which the

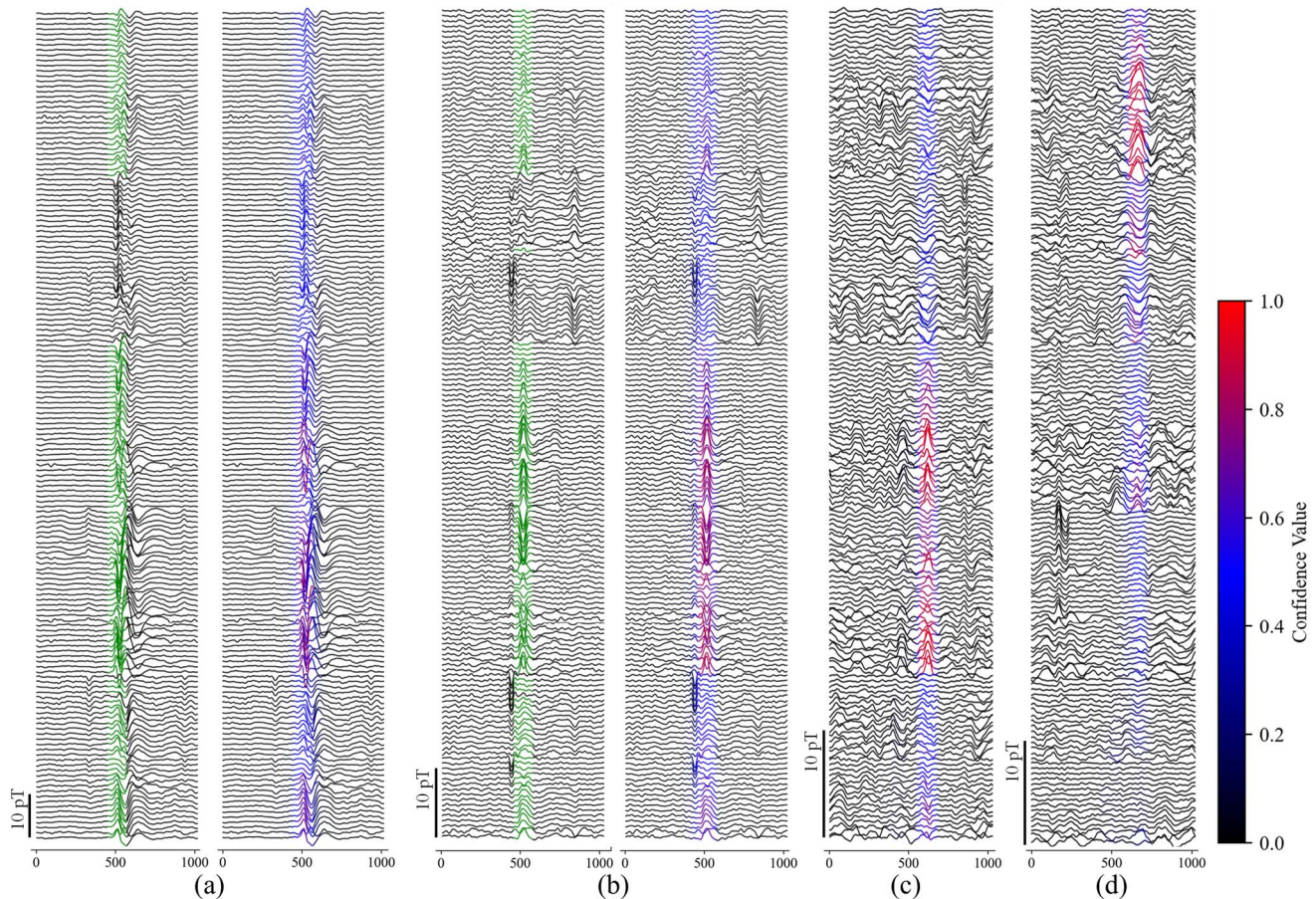
final ECD estimation should be performed. We set each of the parameters described above concerning the results of the segmentation performance and previous studies [31]–[33]. The parameters for applying DBSCAN clustering were set to eps as 0.03, and minPts as 4. The threshold for the inter-cluster distance was set to 1.5 cm. In other words, if the distance between the neurophysiologist-estimated ECD cluster and the FAMED-estimated ECD cluster was less than 1.5 cm, both were evaluated as detecting the same cluster. These clustering parameters were empirically determined.

The evaluation results at the time of the automatically detected ECDs are shown in Fig. 4. This figure shows a histogram of the difference in time between the spikes detected by the AI and the neurophysiologists in the range of the neurophysiologists' detected time  $\pm 120$  ms. In this range, spikes detected by FAMED within  $\pm 50$  ms,  $\pm 30$  ms,  $\pm 15$  ms were 95.5%, 83.0%, and 51.1%, respectively. Out of the 268 spike positive cases, FAMED was able to detect at least one spike in 89.8% of the cases. In addition, we evaluated the time difference between the spike peak and the time detected by FAMED (Fig. 5). Overall, it can be seen that the ECD estimation was performed near the peak of the spike. The percentage of spikes analyzed before their peaks was 65.3%.

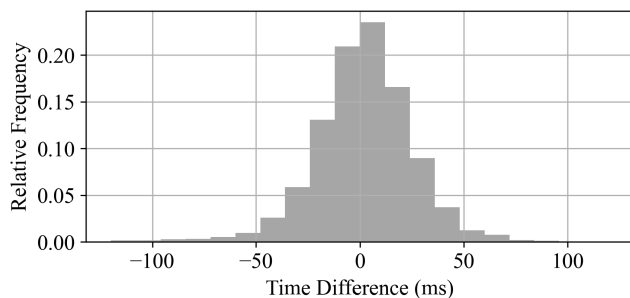
We also evaluated the false positives using the continuous data that the neurophysiologists determined no spikes. A total of 2,137 min of MEG data was evaluated. We calculated the number of false positives per minute per patient as well: 75% of the values were lower than 0.1, and the average was 0.036.

We evaluated the accuracy of ECD localization. The cumulative ratio of the distance between the ECDs estimated by FAMED and the nearest neighborhoods estimated by neurophysiologists are shown in Fig. 6. Out of 268 spike positive cases, there were 111 cases in which the FAMED and neurophysiologists determined the same clusters. There were 27 cases in which the FAMED did not detect any ECDs. A total of 130 cases were excluded by clustering. Here, 13, 31, 34, and 52 cases did not form a cluster based on the neurophysiologist's analysis, FAMED analysis, both analysis, and both analysis without clusters coinciding, respectively. ECD clusters were consistent with the neurophysiologists' analysis in 111 cases. These cases included 37 cases (33.3%) of localization-related epilepsy, 21 cases (18.9%) of age-related epilepsy, 18 cases (16.2%) of congenital brain malformation, 15 cases (13.5%) of epilepsy suspected epilepsy, and 10 cases





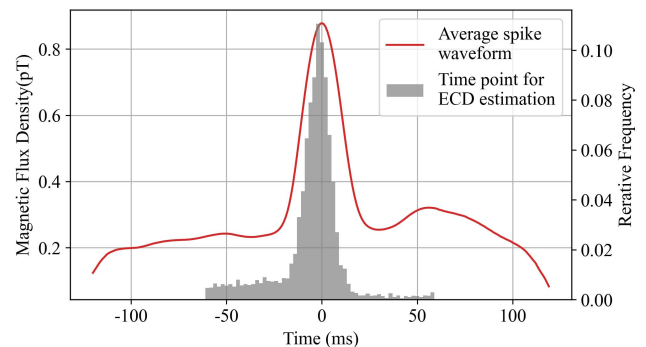
**Fig. 3.** Examples of timepoints and sensors selected by the segmentation network of FAMED. In the waves of (a) and (b), the green waves represent the timepoints and sensors selected by neurophysiologists. The other colored waves represent the output of FAMED segmentation network. The more certain the spike is, the more the color changes from black to blue to red. The IoU values of (a) and (b) were 0.588 and 0.778. Waves of (c) and (d) represent false positives. These data were selected from the spike-negative dataset (see Fig. 1), which was confirmed as no spikes by the expert neurophysiologist (MH). Each false positive detection shows: (c) sharp wave, and (d) slow wave.



**Fig. 4.** Difference in time of spikes between FAMED and neurophysiologists.

(9.0%) of encephalitis. The rest included others with few cases. The median nearest-neighbor distance per case was 0.63 cm for ECDs with clustering and 0.84 cm for ECDs without clustering. The proportion of ECDs analyzed by FAMED that met the nearest neighbor distance of less than 1 cm was 72.9% with clustering and 59.7% without clustering.

We also compared the difference in profiles of ECD clusters by FAMED between focal and non-focal epilepsy using spike positive cases (TABLE IV). For focal epilepsy, we used data from localization-related epilepsy, brain tumors,



**Fig. 5.** Difference in time between the spike peak and the time detected by FAMED. The red line represents the averaged spike waveform.

and cerebrovascular malformations, whereas for non-focal epilepsy, we used data from West syndrome and LGS cases. We excluded from this analysis cases with fewer than four ECDs analyzed by neurophysiologists (43 focal epilepsy cases, 2 non-focal epilepsy cases). In both epilepsy types, there were cases where FAMED could not detect the ECD cluster. The cause of this was FAMED detected only a few ECDs. The cluster size in focal epilepsy is significantly smaller than in



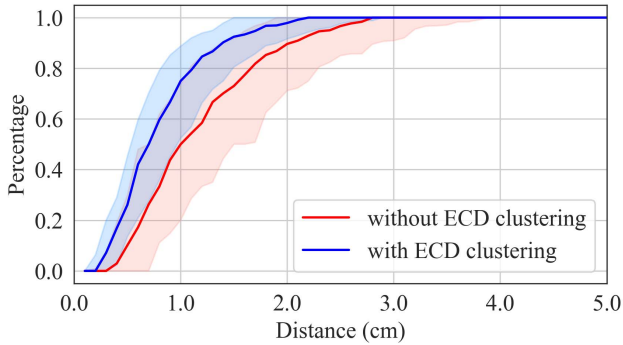


Fig. 6. Cumulative plot of difference in ECD location between FAMED and neurophysiologists. The red line and area show the median and quartile range of all results of FAMED. Those in blue show the results with ECD clustering of FAMED.

TABLE IV

COMPARISON OF AUTOMATED ANALYSIS RESULTS FOR FOCAL AND NON-FOCAL EPILEPSY

	Focal epilepsy N=50	Non focal epilepsy N=30	P Value
# Spike and Dipoles by Neurophysiologist	15 [9, 29]	23 [14, 37]	0.148
# Spikes by FAMED	67 [18, 119]	162 [68, 351]	0.002
# Dipoles by FAMED	12 [5, 25]	17 [8, 33]	0.139
Cluster detection rate (%)	74.00	73.33	NA
Cluster size (cm <sup>3</sup> )	8.40 [4.01, 26.71]	121.60 [82.40, 327.23]	< 0.001

Note: Each value except the cluster detection rate represents the median and the quartiles are shown in parentheses. The p value indicates the result of the test by permutation test.

non-focal epilepsy. We obtained the cluster size by fitting ECDs to a volume of 95% equal probability ellipsoid assuming a multivariate normal distribution. Furthermore, in focal epilepsy, the GoF and Confidence Volume of ECD were compared with the neurophysiologist's results (TABLE V). Both values were higher in automated analysis than in the neurophysiologist's analysis.

The linear regression equation for processing time of the FAMED was  $0.8 \times (\# \text{ detected spikes}) + 32.9 \text{ sec}$  ( $R^2 = 0.986$ ). Since the most time-consuming part of whole FAMED processing is ECD estimation, the processing time depends on the number of detected spikes. If FAMED detected 100 spike candidates, we predicted FAMED processing with 112.9 sec.

#### D. Representative Clinical Cases

Representative clinical results of FAMED are shown in Fig. 7. In case Fig. 7 (a), the patient was a 23-year-old man who had intractable epilepsy and was diagnosed with left temporal lobe epilepsy with hippocampal sclerosis. Note that FAMED analysis showed a clear focal cluster in the left temporal lobe (Fig. 7 (a) upper images), which was consistent with that of a neurophysiologist (Fig. 7 (a) lower images). The patient underwent selective amygdalohippocampectomy. In case Fig. 7 (b), the patient was an 8-year-old girl who had focal onset epilepsy coexisted with autism spectrum disorder

TABLE V  
COMPARISON IN ECD BETWEEN NEUROPHYSIOLOGISTS AND FAMED REGARDING FOCAL EPILEPSY

	Neurophysiologists	FAMED	P Value
# Senros	106.0 [72.0, 160.0]	60.0 [50.0, 70.0]	< 0.001
GoF	0.88 [0.79, 0.93]	0.96 [0.95, 0.97]	< 0.001
Confidence volume (cm <sup>3</sup> )	2.11 [0.81, 6.00]	1.60 [0.68, 3.86]	< 0.001

Note: Each value represents the median and the quartiles are shown in parentheses. The p value indicates the result of the test by permutation test.

and suspected multiple sclerosis. The neurophysiologist's analysis showed only a right frontal cluster consisted of high amplitude spikes (Fig. 7 (b) lower images), but FAMED detected not only a right frontal cluster but also a left frontal cluster consisted of spikes smaller than those in the right side (Fig. 7 (b) upper images). The numbers of spikes detected by spike detection of FAMED were 398 on the right side and 132 on the left. Out of these spikes, 45 spikes on the right side and 52 on the left were used for ECD clustering under the condition of GOF > 0.95. Comparison of detected spikes between neurophysiologists and FAMED are shown in Fig. 8 and Fig. 9.

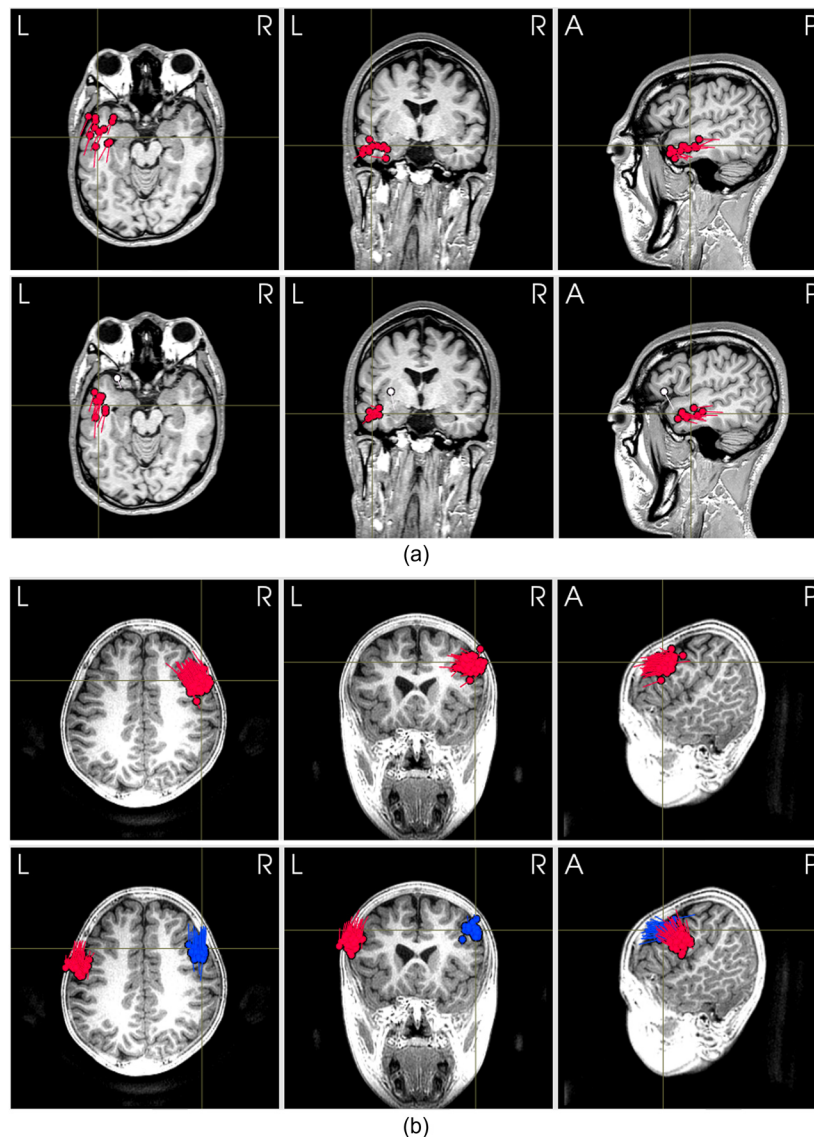
We developed graphical user interfaces (GUIs) for FAMED to display the results concisely but necessarily and sufficiently to help neurophysiologists quickly and accurately diagnose them. The examples of FAMED GUIs showing the results of cases Fig. 7 (a) and Fig. 7 (b) are shown in Fig. 10. FAMED detects spikes, estimates ECDs, and evaluates ECD clusters, which neurophysiologists usually analyze in the diagnosis of epilepsy.

## IV. DISCUSSION

We aimed to develop a fully automated spike detection and ECD analysis using deep learning to minimize the effort and time consumption required for epileptic MEG analyses, and to evaluate its feasibility using a large clinical dataset obtained from a single institution. As a result, we developed a fully automated spike detection and ECD clustering analysis using deep learning (FAMED). Furthermore, we demonstrated that the performance of FAMED is comparable with that of neurophysiologists by training FAMED with as many as 348 patients' data. This is the first reported study to demonstrate fully automated spike detection and ECD clustering analysis of epileptic MEG. Here, we discuss the performance and clinical significance of FAMED, in particular, its potential to automate the skilled and time-consuming work of neurophysiologists.

#### A. Performance of Spike Detection

Most previous studies on AI-based automated analysis of MEG or EEG spikes focused on classifying whether the data was epileptic or not, or whether the data included spikes or not, but did not focus on spike time determination and sensor selection, which is required for ECD analyses. This is insufficient to completely replace time-consuming and laborious works required for the present human-based manual MEG



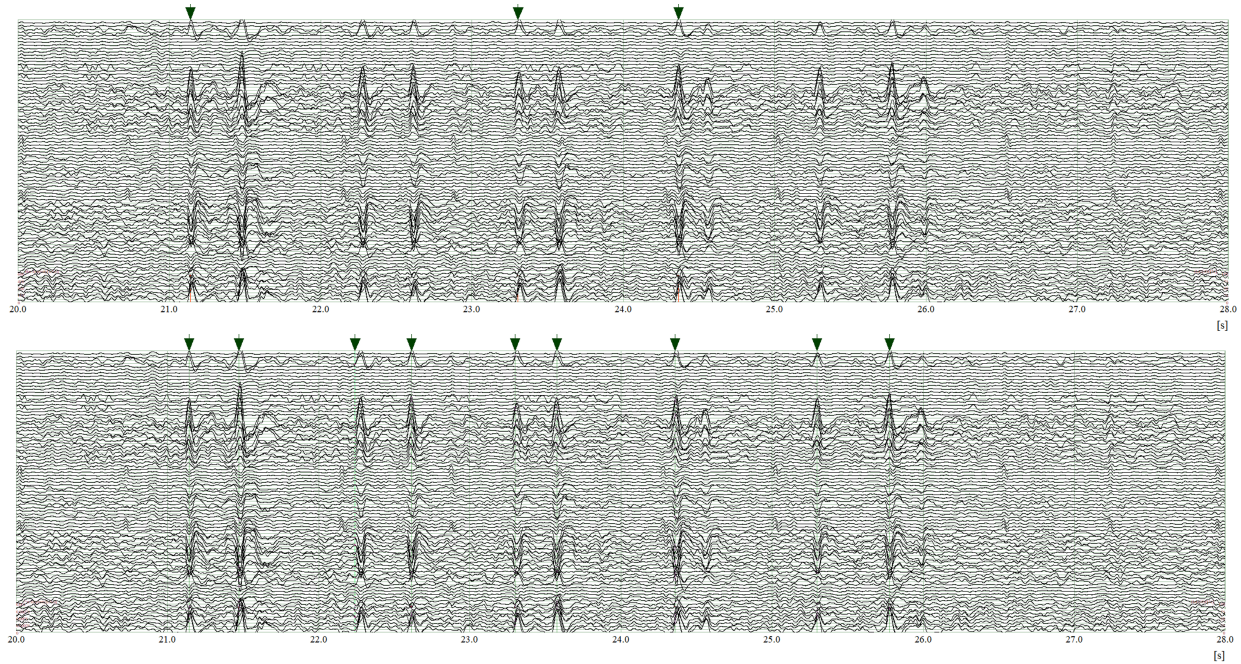
**Fig. 7.** Comparison of ECD localization between the neurophysiologist's analysis (upper images) and the automated analysis using FAMED (lower images). The ECDs within the same cluster are shown by the same color. The cluster with the most numerous ECDs is shown in red. White-colored ECDs represent noise ECDs judged by the DBSCAN algorithm. In case (a), FAMED analyses showed a clear ECD cluster consistent with that by a neurophysiologist. In case (b), neurophysiologist's analysis showed only a right frontal cluster but FAMED pointed out not only a right frontal cluster but also a left frontal cluster.

analysis. Both accurate spike time detection to identify the time from the onset to the peak of a spike waveform and appropriate sensor selection are indispensable to accurately calculate the equivalent current ECD. In this study, we successfully developed a method to identify the accurate spike time and select the appropriate sensors by introducing a network for semantic segmentation. Semantic segmentation has been reported to be used to automatically extract road areas from visual information of an on-vehicle camera or to extract brain tumor areas from brain MR images [18], [19]. We designed the network to successfully extract the appropriate sensors that include a spike waveform using the time-series signals recorded from MEG sensors as input data to the network.

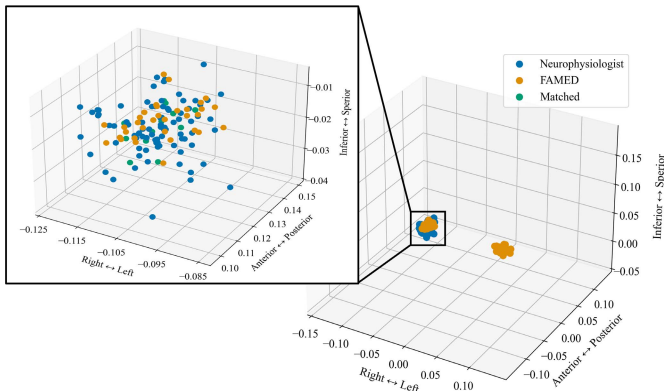
In this study, we applied scSE-module, dilated convolution and DropConnect for spike detection. These methods were

originally proposed mainly for image analysis. However, they were effective in improving the spike detection performance as well as the task for images, as shown in TABLE II. The scSE-module was considered to contribute to the acquisition of spatio-temporal features of spikes. Dilated Convolution is a method to learn a wider range of features. Learning not only the spikes but also the temporally longer features before and after the spikes improved accuracy in spike classification and detection. Regarding DropConnect, we reduced the sensitivity of the classification metrics; however, it was effective for segmentation. Moreover, we showed that the use of the transfer weight further improved the performance of the segmentation network. As a result, these modifications improved the performance in IoU, which is a measure of segmentation performance. This indicates that the proposed





**Fig. 8.** Comparison of spike detection between neurophysiologist and FAMED in case (b) of Fig 7. Arrows above each signals show the detections of spikes. Only right-side sensors are shown. The upper image shows spike detection by the neurophysiologist. The lower image shows the spike detection by FAMED for the same spikes in the upper image.



**Fig. 9.** Comparison of ECD locations between neurophysiologist and FAMED in case (b) of Fig 7. Blue points show the ECDs detected by the neurophysiologist, while orange points show those detected by FAMED. The green points show ECDs that both neurophysiologist and FAMED detected.

architecture can improve the reproducibility of the neurophysiologists' analysis. The performance of our method was equal to or better than that of previous studies on spike classifier learning (Fig. 2, TABLE III). Furthermore, the spike time discrimination performance of FAMED was comparable to that of neurophysiologists (Fig. 4).

We aimed to develop FAMED with high specificity and sufficient sensitivity. In terms of diagnostic significance in the clinical practice of epilepsy, specificity is prior to sensitivity. Of course, false positive detection of spike-like waveforms, such as cardiac artifacts,  $\mu$  waves, humps, and sleep spindles of children, should be minimized. False positive diagnosis of epilepsy should be avoided because the diagnosis leads to a disadvantage to patients by unnecessary long-term

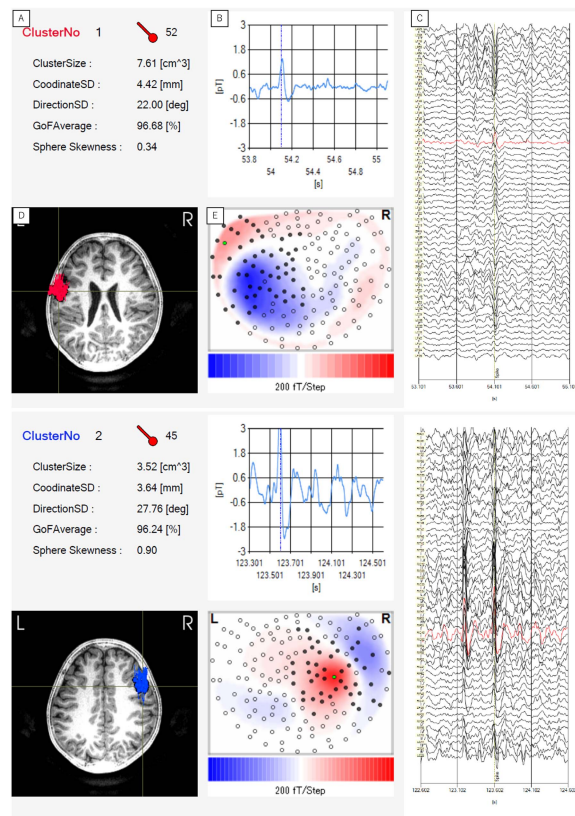
administration of anti-epileptic drugs. In addition, we must not detect all spikes in the data, but only part of the spikes sufficient to quantitatively evaluate spike location and clustering. We suggest that learning using the results of neurophysiologists' analyses resulted in high specificity (0.9971) and sufficient sensitivity (0.7952) required for ECD analyses. This also most likely contributed to minimizing the false positive detection of cardiac artifacts,  $\mu$  waves, humps, and sleep spindles of children as epileptic spikes. In the evaluation, FAMED was able to detect one or more spikes in 89.8% of the 268 spike positive cases. In addition, the average number of false positives per case in unit time (minutes) was 0.036.

### B. Performance of ECD Localization and Clustering

The times of the spikes detected by FAMED were consistent with those by the neurophysiologists. Spikes detected by FAMED were slightly before the spike peak time. This may reflect the fact that neurophysiologists usually perform ECD analysis at rising slope to spike peak. In fact, there are many cases where the location of the ECD is almost the same at the rise and the peak. FAMED reliably learned and followed the clinical heuristics of the neurophysiologists. Regarding ECD localization, FAMED showed comparable performance to neurophysiologists. The evaluation of ECD location showed a performance of 0.63 cm. This was the assessment of the distance between ECDs detected by neurophysiologists and those by FAMED under the condition of applying ECD clustering.

We aimed to develop a method that was applicable for both focal and non-focal epilepsies and that differentiates them. ECD clusters were detected in over 70% cases of focal and non-focal epilepsies each. However, the size of ECD clusters of





**Fig. 10.** Graphical user interface (GUI) of FAMED showing necessary and sufficient results of an ECD analysis. **A.** ECD cluster information. (Cluster size: Volume of 95% equal probability ellipsoid assuming a multivariate normal distribution from the ECD coordinates. CoordinateSD: standard deviation of each ECD location. Direction SD: standard deviation of each ECD direction. GoFAverage: average value of GoFs. Sphere skewness: Deformation of the probability ellipsoid compared to a sphere). **B.** Representative sensor time course. **C.** Sensor time course used in ECD estimation. **D.** MRI image with the ECDs. **E.** Topographic map at spike time of **B** and **C**. This figure corresponds to the case in Fig. 7(b).

focal epilepsy was sufficiently small (TABLE IV). Therefore, FAMED was able to clearly separate cases for focal epilepsy and non-focal epilepsy. In focal epilepsy, both the GoF and the Confidence Volume were better than in the neurophysiologists' analysis. FAMED used about 60 sensors to calculate ECD, fewer than in the neurophysiologists' analysis (TABLE V). This indicates that FAMED detects the ECD pattern more accurately and thus, obtained more stable results by using a more appropriate sensor selection method for ECD analysis than neurophysiologists.

As shown in the case of temporal lobe epilepsy with hippocampal sclerosis (Fig. 7(a)), ECD localization by FAMED was consistent with that of the neurophysiologist. In Fig. 7(a), a cluster of multiple posteriorly directed ECDs was localized in the temporal lobe. This is a typical finding observed in temporal lobe epilepsy with hippocampal sclerosis. In addition, as shown in a case of focal onset epilepsy coexisted with autism spectrum disorder and suspected multiple sclerosis (Fig. 7(b)), FAMED detected two ECD clusters in the right and left frontal lobe. FAMED could analyze spikes derived from the ECD cluster in the right frontal lobe with the quality equal to or greater than that of the neurophysiologist (Fig. 8, Fig. 9). As shown in Fig. 10, smaller but distinct spikes

derived from the ECD cluster in the left frontal lobe were missed by a neurophysiologist. These spikes were confirmed as true positives by two neurophysiologists (MH and KS). In this patient, the MRI showed a white matter lesion in the left frontal lobe, and the small spikes were consistent with this anatomical abnormality. Regarding the numbers of right and left spikes, those of spike detection of FAMED were consistent with the retrospective visual inspection by the neurophysiologist rather than those of ECD clustering of FAMED. This is most probably because the number of right spikes used for ECD clustering was considerably decreased by high GOF ( $> 0.95$ ) to accurately evaluate ECD location. This implies it might be better to use the spike count of spike detection of FAMED to evaluate spike frequency and to use the spike location of ECD clustering of FAMED to evaluate spike location.

### C. Clinical Significance of Automated Analysis

First, AI-based automated analyses free from human-based manual analyses contribute to saving neurophysiologists' time-consuming and laborious efforts, thereby improving the cost-effectiveness of MEG examination. Visual inspection of spikes and manual ECD analyses by neurophysiologists are time-consuming and laborious [5]. It takes several hours to complete these procedures. This time-consuming manual procedure is one of the biggest problems regarding the cost-effectiveness of clinical MEG examination. FAMED fully automatically completes all these procedures in approximately two min in the case of MEG data with 100 spikes. This may help institutions increase the amount of MEG examination and help neurophysiologists minimize the time spent on MEG analyses. The cost-effectiveness of MEG examination is expected to dramatically improve.

Second, AI-based automated analyses free from human-based manual analyses contribute to both consistency and independency, which are important requirements for clinical examinations. AI-based analyses are consistent independent of institutions or neurophysiologists, although human-based manual analyses depend on the experience and ability of neurophysiologists. In addition, neurophysiologists use not only MEG data but also frequently refer to various other information such as clinical symptoms, neuroimaging (MRI, PET, etc.), and EEGs when they inspect MEG waveforms to detect epileptic spikes. However, reference to other clinical information deteriorates independence as a clinical examination. In this respect, FAMED uses only MEG data independent of other clinical information. Taking consistency and independency into consideration, AI-based automated analyses are superior to human-based manual analyses if the performance is equal to that of standard neurophysiologists.

### D. Limitations and Perspectives

We demonstrated that automated ECD analysis by FAMED is feasible as an alternative to manual ECD analysis by neurophysiologists. However, the present study has some limitations. The MEG data were obtained from a single institution. It is not clear whether the performance of FAMED is maintained

across other institutions. As recommended in [34], a multi-institutional study with larger data sets under different clinical conditions is needed. This may contribute to establishing the universality of the performance of FAMED, independent of differences in institutions.

## V. CONCLUSION

In this study, we developed a fully automated spike detection and ECD cluster analysis using deep learning with a semantic segmentation network, FAMED. FAMED was shown to be a potential alternative to neurophysiologists' manual and time-consuming analysis in the clinical situation, contributing to saving neurophysiologists' efforts and improving the cost-effectiveness and quality consistency of MEG examination.

## ACKNOWLEDGMENT

Regarding conflicts of interest related to this research, NDR obtained the grant from 2019 to 2021. The authors would like to thank Editage (www.editage.com) for English language editing.

## REFERENCES

- [1] M. Scherg, "Fundamentals of dipole source potential analysis," in *Auditory Evoked Magnetic Fields and Potentials*, vol. 6, M. Hoke, F. Grandori, and G. L. Romani, Eds. Basel, Switzerland: Karger, 1989.
- [2] M. Lau, D. Yam, and J. G. Burneo, "A systematic review on MEG and its use in the presurgical evaluation of localization-related epilepsy," *Epilepsy Res.*, vol. 79, nos. 2–3, pp. 97–104, May 2008.
- [3] A. C. Papanicolaou *et al.*, "Toward the substitution of invasive electroencephalography in epilepsy surgery," *J. Clin. Neurophysiol.*, vol. 22, no. 4, pp. 231–237, Aug. 2005.
- [4] R. C. Knowlton *et al.*, "Effect of epilepsy magnetic source imaging on intracranial electrode placement," *Ann. Neurol.*, vol. 65, no. 6, pp. 716–723, Jun. 2009.
- [5] S. B. Wilson and R. Emerson, "Spike detection: A review and comparison of algorithms," *Clin. Neurophysiol.*, vol. 113, no. 12, pp. 1873–1881, 2002.
- [6] M. L. Scheuer, A. Bagic, and S. B. Wilson, "Spike detection: Inter-reader agreement and a statistical Turing test on a large data set," *Clin. Neurophysiol.*, vol. 128, no. 1, pp. 243–250, 2017.
- [7] A. Antoniadis *et al.*, "Detection of interictal discharges with convolutional neural networks using discrete ordered multichannel intracranial EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 12, pp. 2285–2294, Dec. 2017.
- [8] M. C. Tjepkema-Cloostermans, R. C. V. de Carvalho, and M. J. A. M. van Putten, "Deep learning for detection of focal epileptiform discharges from scalp EEG recordings," *Clin. Neurophysiol.*, vol. 129, no. 10, pp. 2191–2196, Oct. 2018.
- [9] J. Jing *et al.*, "Development of expert-level automated detection of epileptiform discharges during electroencephalogram interpretation," *JAMA Neurol.*, vol. 77, no. 1, p. 103, Jan. 2020.
- [10] M. A. Jaoude *et al.*, "Detection of mesial temporal lobe epileptiform discharges on intracranial electrodes using deep learning," *Clin. Neurophysiol.*, vol. 131, no. 1, pp. 133–141, Jan. 2020.
- [11] M. I. Khalid, T. N. Alotaiby, S. A. Aldosari, S. A. Alshebeili, M. H. Alhameed, and V. Poghosyan, "Epileptic MEG spikes detection using amplitude thresholding and dynamic time warping," *IEEE Access*, vol. 5, pp. 11658–11667, 2017.
- [12] T. N. Alotaiby, S. R. Alrshoud, S. A. Alshebeili, M. H. Alhameed, and W. M. Alshabhan, "Epileptic MEG spike detection using statistical features and genetic programming with KNN," *J. Healthcare Eng.*, vol. 2017, pp. 1–7, Oct. 2017.
- [13] A. Chahid, F. Albalawi, T. N. Alotaiby, M. H. Al-Hameed, S. Alshebeili, and T.-M. Laleg-Kirati, "QuPWM: Feature extraction method for epileptic spike classification," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 10, pp. 2814–2824, Oct. 2020.
- [14] L. Zheng *et al.*, "EMS-net: A deep learning method for autodetecting epileptic magnetoencephalography spikes," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1833–1844, Jun. 2020.
- [15] A. I. Bagić, R. C. Knowlton, D. F. Rose, and J. S. Ebersole, "American clinical magnetoencephalography society clinical practice guideline 1," *J. Clin. Neurophysiol.*, vol. 28, no. 4, pp. 348–354, Aug. 2011.
- [16] H. Shirozu *et al.*, "Analysis of ictal magnetoencephalography using gradient magnetic-field topography (GMFT) in patients with neocortical epilepsy," *Clin. Neurophysiol.*, vol. 128, no. 8, pp. 1504–1512, Aug. 2017.
- [17] A. Ossadtchi, S. Baillet, J. C. Mosher, D. Thyerlei, W. Sutherling, and R. M. Leahy, "Automated interictal spike detection and source localization in magnetoencephalography using independent components analysis and spatio-temporal clustering," *Clin. Neurophysiol.*, vol. 115, no. 3, pp. 508–522, 2004.
- [18] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Net-Works for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*, (Lecture Notes in Computer Science), vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241.
- [19] R. Li *et al.*, "DeepUNet: A deep fully convolutional network for pixel-level sea-land segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 3954–3962, Nov. 2018.
- [20] T. Wen and R. Keyes, "Time series anomaly detection using convolutional neural networks and transfer learning," in *Proc. IJCAI*, 2019, pp. 1–8.
- [21] R. Nowak, M. Santiuste, and A. Russi, "Toward a definition of MEG spike: Parametric description of spikes recorded simultaneously by MEG and depth electrodes," *Seizure*, vol. 18, no. 9, pp. 652–655, 2009.
- [22] J. Hu, L. Shen, and G. Sun, "Squeeze- and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
- [23] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. Int. Conf. Learn. Represent.*, 2016, pp. 1–13.
- [24] A. G. Roy, N. Navab, and C. Wachinger, "Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks," in *Medical Image Computing and Computer-Assisted Intervention* (Lecture Notes in Computer Science), vol. 11070. Cham, Switzerland: Springer, 2018, pp. 421–429.
- [25] L. Wan, M. Zeiler, S. Zhang, Y. L. Cun, and R. Fergus, "Regularization of neural networks using dropconnect," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 1–9.
- [26] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. Knowl. Discovery Data Mining*, 1996, pp. 1–6.
- [27] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.
- [28] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2017, *arXiv:1711.05101*.
- [29] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.
- [30] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 2623–2631.
- [31] A. Ochi, C. Y. Go, and H. Otsubo, "Clinical MEG analyses for children with intractable epilepsy," in *Magnetoencephalography*, W. P. Elizabeth, Ed. Rijeka, Croatia: InTech, 2011, pp. 177–192.
- [32] R. Alkawadri, R. C. Burgess, Y. Kakisaka, J. C. Mosher, and A. V. Alexopoulos, "Assessment of the utility of ictal magnetoencephalography in the localization of the epileptic seizure onset zone," *JAMA Neurol.*, vol. 75, no. 10, p. 1264, Oct. 2018.
- [33] I. Westmijse, P. Ossenblok, B. Gunning, and G. van Luijckelaar, "Onset and propagation of spike and slow wave discharges in human absence epilepsy: A MEG study," *Epilepsia*, vol. 50, no. 12, pp. 2538–2548, Dec. 2009.
- [34] M. B. Westover, J. J. Halford, and M. T. Bianchi, "What it should mean for an algorithm to pass a statistical Turing test for detection of epileptiform discharges," *Clin. Neurophysiol.*, vol. 128, no. 7, pp. 1406–1407, Jul. 2017.