








EEGWaveNet: Multiscale CNN-Based Spatiotemporal Feature Extraction for EEG Seizure Detection

Punnawish Thuwajit, Phurin Rangpong, Phattarapong Sawangjai, Phairot Autthasan ,
Rattanaphon Chaisaen , Nannapas Banluesombatkul , Puttaranun Boonchit ,
Nattasate Tatsaringkansakul , Thapanun Sudhawiyangkul ,
and Theerawit Wilaiprasitporn , Member, IEEE

Abstract—The detection of seizures in epileptic patients via Electroencephalography (EEG) is an essential key to medical treatment. With the advances in deep learning, many approaches are proposed to tackle this problem. However, concerns such as performance, speed, and subject-independency should still be considered for practical application. Thus, we propose EEGWaveNet, a novel end-to-end multiscale convolutional neural network designed to address epileptic seizure detection. Our network utilizes trainable depth-wise convolutions as discriminative filters to simultaneously gather features from each EEG channel and separate the signal into multiscale resolution. Then, the spatial-temporal features are extracted from each scale for further classification. To demonstrate the effectiveness of EEGWaveNet, we evaluate the model in three datasets: CHB-MIT, TUSZ, and BONN. From the results, EEGWaveNet's performance is comparable to other baseline methods in the subject-dependent approach and outperforms the others in subject-independent approaches. EEGWaveNet also has time complexity comparable to the compact EEGNet-8.2. Moreover, we transfer the model

trained from the subject-independent approach and fine-tune it with a 1-h recording, significantly improving sensitivity and F1-score (Binary) compared to without fine-tuning. This article indicates the possibility of further developing this model and the fine-tuning methodology toward healthcare 5.0, where the AI aid clinicians in a manner of man-machine collaboration.

Index Terms—Convolutional neural network (CNN), deep learning, seizure Electroencephalography (EEG), spatiotemporal neural network, transfer learning (TL).

I. INTRODUCTION

EPILEPSY is a prevalent neurological disorder that has been the leading cause of neurological impairment, with 65 million people affected [1]. It can be identified by the recurrence of at least two unprovoked seizures [2]. Electroencephalography (EEG) has been known as the most common method for assessing patients with epilepsy [3]. Conventionally, clinicians manually diagnose EEG seizures by using the patient history of epilepsy in combination with observing abnormal EEG characteristics such as epileptiform spikes, which is an exhaustive process [4]. Therefore, several studies have proposed computational intelligence to automatically identify seizures from various EEG databases [3].

Deep convolutional neural network (DCNN) is a variant of artificial neural networks that can extract the represented features of input data without relying on high complexity feature engineering. Previous findings have shown the implementation of DCNN in an EEG seizure prediction model and achieved more than 90% accuracy, outperforming traditional machine learning (ML) methods in seizure classification [5]. However, there are few flaws that prevent these methods from being practical in clinical use.

In most of the proposed seizure classification models, the models are trained and evaluated in a subject-dependent setting. The subject-dependent approach can provide more accurate classification than a subject-independent approach due to the high intersubject variance of the EEG. However, it requires a large portion of an individual's EEG recording to train the model before reaching its optimal performance. Labeling the

Manuscript received August 27, 2021; revised October 7, 2021, November 4, 2021, and November 17, 2021; accepted December 1, 2021. Date of publication December 10, 2021; date of current version May 6, 2022. This work was supported in part by TSRI under Grant SRI62W1501, in part by MHESI under Grant RGNS63-252, and in part by NRCT under Grant N41A640131. Paper no. TII-21-3718. (Corresponding authors: Thapanun Sudhawiyangkul; Theerawit Wilaiprasitporn.)

Punnawish Thuwajit is with the Enrolling High-School, Suankularb Wittayalai School, Bangkok 10200, Thailand (e-mail: konkuad2@gmail.com).

Phurin Rangpong, Phattarapong Sawangjai, Phairot Autthasan, Rattanaphon Chaisaen, Nannapas Banluesombatkul, Thapanun Sudhawiyangkul, and Theerawit Wilaiprasitporn are with the School of Information Science and Technology (IST), Vidyasirimedhi Institute of Science and Technology (VISTEC), Rayong 21210, Thailand (e-mail: poorin31632@gmail.com; phattarapong.saw@gmail.com; phairot.a_s17@vistec.ac.th; rattanaphon.c_s18@vistec.ac.th; nannapas.b_s18@vistec.ac.th; thapanun.s@vistec.ac.th; theerawit.w@vistec.ac.th).

Puttaranun Boonchit and Nattasate Tatsaringkansakul are with the School of Information Science and Technology (IST), Vidyasirimedhi Institute of Science and Technology (VISTEC), Rayong 21210, Thailand, and also with the Tokyo Institute of Technology, Meguro-ku 152-8550, Japan (e-mail: puttaranun_jinny@hotmail.com; stampck2@hotmail.com).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TII.2021.3133307>.

Digital Object Identifier 10.1109/TII.2021.3133307

whole EEG recording only to train the model for a specific subject can be a laborious and time-consuming task, which is inappropriate in a real-world application. To tackle this problem, some researchers have proposed a fully subject-independent model that generalized across subjects [6]–[8] by utilizing a pretrained model from multiple subjects on an unseen subject. Although these research works show promising classification results in subject-independent classification, these methods are not in an end-to-end fashion. They require extensive data preprocessing and feature engineering, which is still time-consuming, impractical in real-time usages, and needs expert knowledge.

Another approach to tackle the aforementioned problems is to employ a transfer learning (TL) technique that transfers the model pretrained from multiple subjects and fine-tune it with a small part of recording from an individual subject. Yizhang *et al.* [9] have integrated transductive TL in an existing Takagi–Sugeno–Kang (TSK) model and conducted an experiment comparing different ML methods with TL against non-TL. Their result shows that TL-based models significantly outperformed non-TL-based across different groups of patient in a seizure classification task.

In this article, we propose an automatic seizure detection model, EEGWaveNet, a novel end-to-end multiscale DCNN architecture. EEGWaveNet exhibits high subject-independent performance that can be further improved by implementing TL with the fine-tuning method.

The contributions of this article are as follows.

- 1) We propose a novel end-to-end architecture that can capture and decompose EEG features from multiscale resolution. The proposed architecture is end-to-end, fast, and compact.
- 2) We investigate our model on three datasets: CHB-MIT, TUSZ, and BONN. For low variance CHB-MIT and high variance TUSZ datasets, the proposed model outperforms previous state-of-the-art models in a subject-independent approach. Furthermore, despite the limited number of data in the BONN dataset, the proposed model still performs comparably to other state-of-the-art models in a subject-dependent approach.
- 3) We introduce a TL approach to enhance the subject-independent model. By fine-tuning the model with data from less than one recording, we improve the model's performance while maintaining the practicability of the subject-independent approach.

The rest of this article is organized as follows. Section II explains some backgrounds and related work. Section III describes the datasets, EEG preprocessing, our proposed method's architecture, and the benchmark studies. Experimental results of the proposed method are presented in Section IV and discussed in Section V. Finally, Section VI concludes this article. We also provide an appendix of the study on the Bonn dataset. The code for our model is publicly available on GitHub.

II. RELATED WORK

There are mainly two types of epileptic classification; prediction and detection [10], [11]. For the prediction, it is to classify between the phase of interictal, preictal, and ictal, which is

crucial for taking precautions and interventions before seizure onset [12]. Meanwhile, seizure detection is to classify between seizure and nonseizure states (interictal and preictal are the nonseizure states [11]). The benefit of seizure detection is to reduce the workload of the clinicians in the diagnosis step. In this work, we will focus mainly on seizure detection.

This section briefly describes the studies related to epileptic seizure detection, which consisted of ML approaches, DL approaches, and TL.

A. ML in Epileptic Seizure Detection

ML is a traditional method for seizure detection and classification. To effectively utilize EEG signals in ML, assistance from preprocessing techniques and feature extraction methods are needed. Wavelet packet decomposition (WPD) and filter bank common spatial pattern (FBCSP) are widely used feature extraction techniques for the EEG signals. Tang *et al.* [13] utilized WPD in the very first step to decompose raw EEG signal according to the frequency band of the EEG (Delta to Gamma band). After more feature extraction processes, the obtained features then be classified by support vector machine (SVM). Kai *et al.* [14] proposed FBCSP to autonomously distinguish the key spatial-temporal EEG features for classification with ML algorithm. FBCSP is an advanced CSP algorithm that autonomously selects a subject-specify frequency range of EEG signals, maximizing the variances between two classes of EEG signals. Although ML approaches demonstrate reasonable classification performance, these methods need a certain degree of prior knowledge to hand-design the features extraction methods.

B. Deep Learning in Epileptic Seizure Detection

Recently, DL has been increasingly adopted in EEG seizure detection due to its ability to learn distinguishable features from raw EEG signals without relying on high complexity feature engineering. There are two approaches for epileptic seizure detection: subject-dependent and subject-independent.

The subject-dependent approach studies a model for a particular subject, which is tuned for each specific subject but not generalized. Tian *et al.* [15] generated the spatial-temporal feature domain by WPD and Fast Fourier transform (FFT), followed by the epileptic seizure detection using multiview Takagi–Sugeno–Kang fuzzy system (MV-TSK-FS). Their model provides high accuracy with the tradeoff of being a large and non-end-to-end model. Later, Li *et al.* [16] embedded a squeeze-and-excitation network with a deep spectral-temporal features extraction network, which performed efficiently on the multiple datasets. Although the abovementioned subject-dependent seizure models can provide excellent accuracy, it requires a long calibration session when using with a new patient.

The subject-independent approach designs a model to analyze seizures across multiple subjects, which is practical for clinical usages. However, only a few researchers studied this approach, and the performance is not as high as the subject-dependent approach due to the high intersubject variability and limited data availability. Yuan *et al.* [6] proposed a multiview DL framework for seizure detection based on CNN with autoencoders. They utilized multichannel EEG spectrograms and considered both inter

and intra correlation of the EEG channels. Transforming time-domain EEG signals to time–frequency domain spectrograms provides the model with more detailed contextual information. However, it also increases the input dimension resulting in a large model that requires high computation resources.

C. Transfer Learning

TL technique is generally used in DL to enhance the model's performance and shorten the calibration session. It has been emphasized on EEG for various healthcare-related applications such as the sleep stage classification [17], brain–computer interface [18], and the epileptic seizure classification [8], [9], [19].

Agrawal *et al.* [19] pretrained the image processing model on the ImageNet database. Then, they transferred the model to work on RGB images that were converted from the size-transformed matrices of the EEG dataset by a function in MATLAB. With TL, the model required significantly shorter training time. Theekshana *et al.* [8] transferred the subject-independent model to the subject-dependent seizure classification with the fine-tuned method. First, they removed one patient out and trained the model with the rest of the data. Then, they fine-tuned the trained model with the samples of the removed patient. Additionally, they varied the number of fine-tuning samples while using the same validation dataset to evaluate the performance of TL. With the improved performance of TL-based models in seizure detection [8], [9], [19], it provides a promising tool to reduce the clinician's workload. Instead of training the model in subject-dependent manner, they only need to select a few samples from a patient to fine-tune the pretrained model.

III. METHOD

This section describes the dataset used in our article. Then, the design of our proposed method, along with the baseline methods, is described. Last, we present the evaluation methods to demonstrate the effectiveness of EEGWaveNet.

A. Data Acquisition

1) **CHB-MIT Dataset:** We evaluated our proposed method and baseline methods on the public CHB-MIT dataset [20]. The CHB-MIT dataset was collected from the Children Hospital Boston. It consists of 256 Hz EEG recordings of approximately 1 h each from 23 pediatric patients with intractable seizures (5 males and 17 females), which are grouped into 24 cases. We fail to read some channel data for patient number 12, therefore, this patient's EEG signals are not included. Male patients are in the age group ranging from 3 to 22 years, while female patients are in the age group ranging from 1.5 to 19 years. There is an average of 6.62 ± 4.21 recordings with seizure per patient, while the ratio of seizure to normal signal is approximately 1:60. In this article, the preprocessing includes the utilization of a lowpass filter to filter out noises unrelated to seizure events with frequency higher than 64 Hz [16]. Only the recordings with seizure events were considered. In our work, 21 common channels with seizure are used consisting of $FP_1 - F_7, F_7 - T_7, P_7 - O_1, FP_1 - F_3, F_3 - C_3, C_3 - P_3, P_3 - O_1, FP_2 - F_4, F_4 - C_4, C_4 - P_4, P_4 - O_2, FP_2 - F_8, F_8 - T_8,$

TABLE I
EEGWAVENET ARCHITECTURE

Module	Layer	Kernel	Output	Activation	Parameters
a	Input		(C, I, T)		
	Conv2D	$(1, 2)$ stride 2 group 21	$(C, 1, T/2)$	Linear	42
	Conv2D	$(1, 2)$ stride 2 group 21	$(C, 1, T/4)$	Linear	42
	Conv2D	$(1, 2)$ stride 2 group 21	$(C, 1, T/8)$	Linear	42
	Conv2D	$(1, 2)$ stride 2 group 21	$(C, 1, T/16)$	Linear	42
	Conv2D	$(1, 2)$ stride 2 group 21	$(C, 1, T/32)$	Linear	42
b	Conv2D	$(1, 2)$ stride 2 group 21	$(C, 1, T/64)$	Linear	42
	Input		$(C, I, T/2^k)$		
	Conv2D	$32 \times (1, 4)$	$(32, 1, T/2^{k-3})$		2688
	BatchNorm2D				
	Activation			Leaky ReLU	
	Conv2D	$32 \times (1, 4)$	$(32, 1, T/2^{k-3})$		4096
c	BatchNorm2D				
	Activation			Leaky ReLU	
	Conv2D	$32 \times (1, 4)$	$(32, 1, T/2^{k-9})$		4096
	BatchNorm2D				
	Activation			Leaky ReLU	
	Global average pooling		32		
c	Input		160		
	FC		64	Leaky ReLU	
	FC		32	Sigmoid	
	FC		2	Linear	
	Classifier		1	Log Softmax	

Where C is the Number of Channels, T is the Number of Timepoints and k Denotes the Order of Layer in Module B.

$T_8 - P_8, P_8 - O_2, F_Z - C_Z, C_Z - P_Z, P_7 - T_7, T_7 - FT_9, FT_9 - FT_{10},$ and $FT_{10} - T_8$. We preorganized the dataset by 4 s segmentation, as presented in our baseline studies [16], along with 1 s overlapping to increase the number of samples.

2) **TUSZ Dataset:** In order to observe the generalization of EEGWaveNet, we evaluated our method on the TUSZ dataset [21], which is one of the world's largest annotated datasets for EEG seizure classification. This dataset was a portion of the Temple University Hospital EEG Corpus and contains a large variety of seizure morphology. TUSZ dataset accumulated over the years, including subjects of all ages, and recorded with different equipment generations. Thus, the variance is much more present than CHB-MIT. Furthermore, the recording length of each patient is inconsistent and mostly not sufficient for subject-dependent training. In our article, the preprocessing is similar to the CHB-MIT dataset in which we used identical channels and filter out noises with higher frequency than 64 Hz. Note that only 28 patients with total recordings of length more than 1 h are chosen. Due to the diversity of sampling frequency in the TUSZ dataset, the EEG signals are down-sampled equally to 256 Hz to match the CHB-MIT setup and reduce computational cost. Then, the same set of channels are chosen, and the same segmentation setups are applied.

B. Proposed Model Architecture: EEGWaveNet

The EEGWaveNet consists of three modules: Multiscale convolution, spatial-temporal feature extraction, and classifier as illustrated in Fig. 1. Further details can be found in Table I.

1) **Multiscale Convolution:** Examining EEG in multiple scales has been proven effective by various studies [13], [16], [22]. Inspired by the feature pyramid technique that is widely used in image processing [23], the Multiscale convolution module divides the input signal into different scales of resolution in a channel-wise manner by using depthwise convolution. This module contains six consecutive layers, where each is designed to capture the features in half the scale of the previous layer's resolution. The module can automatically learn the weight to

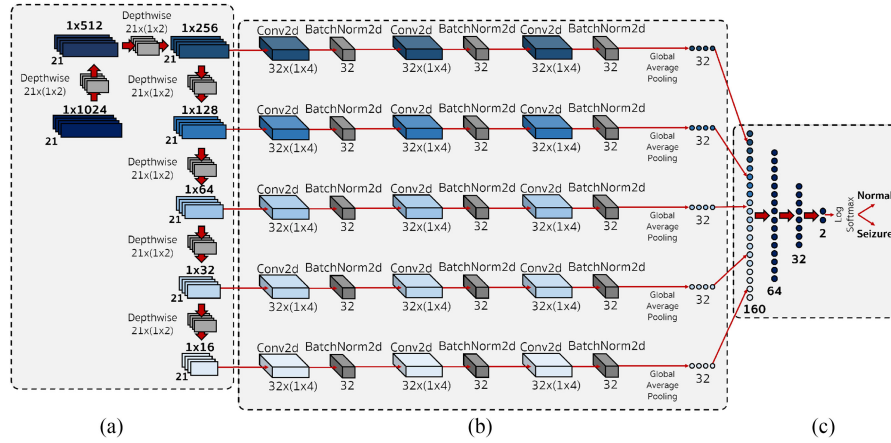


Fig. 1. Overall visualization of EEGWaveNet architecture. (a) Multiscale convolution module. (b) Spatial-temporal feature extraction module. (c) Classifier module consists of FC layers and activation for two-class classification.

effectively distinguish the valuable intrachannels features while down-sampling the signal to half.

In this module, the input data are shaped as $(C, 1, T)$, where T and C denote the number of timepoints and number of channels. C is set to 21 and T is set to 1024, corresponding to Fig. 1 throughout all experiments in this article. Each layer performs a depthwise convolution to each channel with the same kernel size of $(1, 2)$, a stride of 2, and no padding. By this configuration, the number of timepoints from the k layer becomes half of the $k - 1$ layer giving the output size from the k layer to be $(C, 1, \frac{T}{2^k})$. In other words, the resolution of the k layer's signal is reduced to half of the $k - 1$ layer's signal. Depthwise convolution independently extracts features with no crossing of information between channels. Outputs from the second to the sixth layers will advance to the next module for further capturing spatial-temporal features from different field of views.

2) Spatial-Temporal Feature Extraction: In this module, we extract spatial-temporal features from each output's scale of the multiscale convolution module. This module utilizes three regular convolution layers with the same kernel size of $(1, 4)$, a stride of 2, and no padding. The kernel can temporally convolve four timepoints, which result in temporal features from these timepoints. As each layer's input has a different resolution, each module's layer will have a different field of view. A deeper layer will provide a wider field of view for temporal feature extraction. With the kernel size of $(1, 4)$, the field of views of the first to fifth layers are 0.0625, 0.125, 0.25, 0.5, and 1 s, respectively. In contrast to depthwise convolution used in the multiscale convolution module, regular convolution runs through all of the channels extracting spatial-temporal features of the input. Each input feature from multiscale convolution is connected to an identical structure composed of layers of regular convolution and batch normalization in order to normalize the transformed wave of each scale, each followed by a leaky ReLU activation function. Finally, the output is pooled by global average pooling into 32 features of each scale. Therefore, a total of 160 features (5 scales, 32 features each) are extracted from this module.

3) Classifier: This module contains four fully connected (FC) layers with the size of 160, 64, and 32, respectively. The last layer use Log softmax as the binary classifier. Log softmax

calculates the logarithm of the prediction probability of each class, as illustrated in (1). Thus, log softmax heavily penalize the highly incorrect class while stabilizing gradient descent calculation, further optimizing the training time.

$$\hat{y}(x_i) = \log \left(\frac{\exp(x_i)}{\sum_j \exp(x_j)} \right). \quad (1)$$

Finally, This model uses cross entropy to compute classification loss. This loss function, as illustrated in (2), computes the similarity between two sets of prediction probability values.

$$\mathcal{L}_{CE}(y, \hat{y}) = - \sum_{k=1}^{|class|} y_k \log \hat{y}_k. \quad (2)$$

C. Network Training

The proposed method was implemented using the PyTorch v1.7.1 framework. The training process was implemented using NVIDIA RTX2080ti with 11 GB memory. In each iteration of training, the loss function was optimized by utilizing Adam optimizer with a learning rate of 10^{-3} . We set a batch size of 512 samples for all experiments. Moreover, to mitigate the effects of class imbalance, we applied class weights in computing training and validation loss. Finally, the number of training iterations relied on the early stopping strategy so that the training process was stopped in case there was no reduction of the validation loss for ten consecutive epochs. The weights from the epoch with the minimal validation loss are chosen for evaluation. The early stopping implementation is according to PyTorch API [24].

D. Baseline Method

We compare our method with other five state-of-the-art methods to demonstrate the effectiveness of our EEGWaveNet.

1) FBCSP+SVM: FBCSP is created base on the concept of CSP algorithm to autonomously select the key spatial-temporal discriminative EEG features, which were extracted from multiple bands. In this work, FBCSP is applied to decompose EEG signals into eight bands with bandwidth of 4 Hz, beginning from 4 to 64 Hz. Then, SVM cooperating with grid search

algorithm is implemented to classify the obtained features. The hyperparameters for SVM classifier are consisted of kernel (radial bias function (RBF), sigmoid, linear), C (0.001, 0.01, 0.1, 1, 10, 100, 1000), specifically for RBF, γ (0.01, 0.001), and class _ weight (balanced). The grid search algorithm is used to evaluate the prediction on the validation set of the classification in order to obtain the optimal set of hyperparameters. Finally, the SVM classifier with the optimal parameters is used for testing.

2) **EWT+RF**: Empirical wavelet transform (EWT) is used in a nonstationary signal analysis, which can create an adaptive wavelet-based filter specific to the input signal. Abhijit *et al.* [25] proposed the multivariate approach on EWT for EEG seizure detection. Different channels of EEG signal were decomposed into different numbers of sub-bands and frequency ranges. They utilized a channel selection algorithm to select five channels for decomposing into multiple frequency sub-bands by EWT. The instantaneous amplitude and frequency information of each sub-bands belong to a similar oscillatory level were then combined to create joint instantaneous amplitude and frequency functions for further feature extraction and classification. Our article uses their RF classifier's result for comparison, which gives the best performance among other ML classifiers.

3) **MV-TSK-FS**: Proposed by Tian *et al.* [15], MV-TSK-FS is a multiview classifier based on the classical TSK Fuzzy system [9] that can interpret each input features' importance according to the weight of each view. In this framework, raw EEG signal, frequency domain signal from FFT, and time–frequency domain signal from WPD were utilized as initial features. Then, CNN was applied further to extract deep features from each view's initial features. Last, the extracted deep features from all views were weighted and classified by MV-TSK-FS.

4) **STFT-mConvA**: STFT-mConvA is an autoencoder-based model that can learn multiview features from each channel's spectrogram. In yuan *et al.* [6], they utilized the model autoencoders for self-supervised learning of input's latent representations from both inter and intra channels of EEG. Then, the latent representations from inter-and intra channels were combined and used to classify the seizure state.

5) **EEGNet-8,2**: EEGNet-8,2 was proposed as a compact CNN model designed to capture discriminative EEG features [26]. Despite its compactness, EEGNet-8,2 demonstrated high robustness by achieving outstanding performance in various EEG-based brain–computer interface paradigms. It was widely accepted as the state-of-the-art method in various EEG classification tasks. However, EEGNet-8,2's performance on seizure classification was not explored as of now. In this article, EEGNet-8,2 is reproduced for the seizure detection task, keeping the optimal set of hyperparameters as recommended in the original publication [26].

6) **CE-stSENet**: CE-stSENet is an end-to-end CNN-based model that can simultaneously extract multilevel spectral and multiscale temporal features of EEG signals [16]. In Hu *et al.* [27], they used this model which utilized squeeze and excitation (SE) block to selectively magnified the discriminative features while suppressing redundant features. Moreover, maximum mean discrepancy-based information maximizing loss was used along with cross-entropy loss to mitigate the overfitting

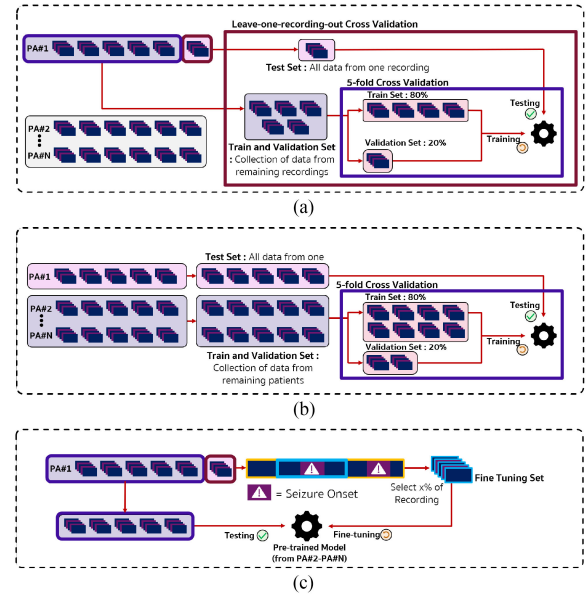


Fig. 2. Overall training framework of our training procedure. (a) Subject-dependent approach. (b) Subject-independent approach (similarly for TUSZ dataset). (c) Procedure of fine-tuning with $x\%$ fine-tune percentage.

problem resulting from insufficiency of seizure events in EEG recording.

7) **MinMaxHist+XGB**: MinMaxHist was developed as a feature extraction algorithm for measuring the topological patterns of EEG signals [28]. It can quantitatively interpret the waveform characteristics by determining the scale of changes in the EEG signals. After MinMaxHist was applied with other feature extraction algorithms, the extracted features then underwent a comprehensive series of feature screening to optimize the classification performance. Finally, extreme gradient boosting (XGB) algorithm was utilized for classification, which provides optimal performance compared to the other subject-independent ML approach.

E. Experimental Evaluation

In our article, we conducted three main experiments: subject-dependent, subject-independent, and TL as shown in Fig. 2. We conducted experiment on three different dataset: CHB-MIT, TUSZ, and BONN. In the first experiment, we employed our model on the CHB-MIT dataset. We employed two datasets on the second experiment: CHB-MIT and TUSZ. As for the BONN dataset, we discuss the experiment and its results in the Appendix to avoid the out-of-scope study because this dataset is too small.

We evaluated the performance of each model with accuracy, sensitivity, specificity, binary F1-scores, and weighted F1-scores. Accuracy computes the percentage of EEG segments that are computed correctly as in

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

where TP , TN , FP , and FN denote true positives, true negatives, false positives, and false negatives.

Sensitivity provides the true positive rate of the relevance class (seizure) as demonstrate in (4), while specificity provides the true negative rate of the nonrelevance class (nonseizure) as shown in (5).

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{Specificity} = \frac{TN}{TN + FP}. \quad (5)$$

Binary F1-scores takes account of the precision (positive predictive value) and sensitivity of the relevance class (seizure), which can be expressed in TP , TN , FP , and FN as in (6). In clinical circumstances, seizure events are much rarer than nonseizure (approximately 1:60 in this dataset), making the binary F1-score a viable measure to compare each model's viability

$$\text{F1-score (Binary)} = \frac{2TP}{2TP + FP + FN}. \quad (6)$$

Finally, we compared the weighted F1-score, computed by weighting the average of the F1-score relative to the two classes as in (7). It provides an overall performance evaluation even when dealing with class imbalance

$$\text{F1-score (Weighted)} = \frac{\sum_{\text{class}} n_{\text{class}} F1_{\text{class}}}{\sum_{\text{class}} n_{\text{class}}} \quad (7)$$

where n_{class} denotes the number of samples in each class, and $F1_{\text{class}}$ is the binary F1-score of each class.

1) Experiment I–Subject-Dependent Approach: In a subject-dependent approach, we trained, validated, and tested the model on the CHB-MIT dataset by using the data from an individual subject. We used leave-one-recording-out cross validation technique as the evaluation method. After leaving one recording as a test set, we trained and validated the model by using five-fold cross validation on the remaining data. The performance of one patient is the average across $N \times 5$ trials, where N is the number of recordings for the patient. The averaged performance across all patients is then reported.

In terms of training speed, each patient contains a varying amount of files with has led to an inconsistent speed of training per epoch. Therefore, the average time per patient is reported, which also takes in account of the average time of the model converging to its minimal loss.

2) Experiment II–Subject-Independent Approach: In a subject-independent approach, we trained and validated the model using the data from multiple subjects and tested the model using an individual subject's data. We used leave-one-subject-out cross validation technique as the evaluation method. After leaving one subject as a test set, we trained and validated the model by using five-fold cross validation. The performance of one patient is the average across five trials. The averaged performance across all patients is then reported.

Aside from the CHB-MIT dataset, we also employed the subject-independent experiment on the TUSZ dataset in order to assess the robustness and generalization of EEGWaveNet. The large variety in time of collection, machine setups, and human variations prove the TUSZ dataset an appropriate dataset in this

objective. Note that in the experiment with TUSZ dataset, we compare EEGWaveNet with only EEGNet-8,2 since there is rarely a study that evaluates seizure detection tasks in TUSZ dataset where most of the studies are seizure types classifying tasks.

3) Experiment III–Transfer Learning: To further improve the performance from the subject-independent approach, we employed TL on the subject-independent model. In this experiment, we randomly selected a recording (approximately one hour) from an individual subject. Then, we selected a seizure event in the chosen recording and included normal EEG signals to cover $X\%$ of the recording. This ensures that the selected data for fine-tuning contain the seizure class. The value X will be defined as “fine-tune percentage,” where we varied this value from 10% to 100%. Finally, we fine-tuned the model for 100 epochs using the selected data. Moreover, we froze the weights of the multiscale convolution and spatial-temporal feature extraction modules while fine-tuning only the classifier. In this experiment, we tested the fine-tuning procedure on the CHB-MIT dataset but not on the TUSZ dataset due to the scarcity and insufficient recordings length of each patient.

IV. RESULTS

This section reports the results and statistical analysis of our experiments. The performance of each experiment was reported in accuracy, sensitivity, specificity, weighted F1-score, and binary F1-score with standard deviation [Accuracy (%) \pm SD and F1-score(%) \pm SD].

A. Experiment I–Subject-Dependent Approach

To evaluate EEGWaveNet in a subject-dependent approach, we compared our experimental results with the recent works that exploit ML and DL on the CHB-MIT dataset [15], [16], [25], [26]. **Table II** reports the performance and computation resource used by EEGWaveNet and other baseline methods in a subject-dependent manner. It is observed that EEGWaveNet achieved comparable performance to other state-of-the-art baseline methods in all measures except sensitivity. The binary F1-score was statistically proven via independent t -test to be significantly higher than all compared methods ($p < 0.05$).

In terms of computation resources, EEGNet-8,2 has the lowest number of parameters and fastest inference time, while EEGWaveNet is the second best and comparable to EEGNet-8,2. Considering the training strategy with early stopping, EEGWaveNet outperforms the other baseline models, including EEGNet-8,2, in averaged training time per patient.

B. Experiment II–Subject-Independent Approach

Table III exhibits the performance of EEGWaveNet and other baseline methods [6], [26], [28] in subject-independent approach. Being chosen as the best baseline ML method in the subject-independent approach on the CHB-MIT dataset, MinMaxHist + SVM provides lower accuracy than most of the DL methods. Despite MinMaxHist + SVM provide high sensitivity, it is not appropriate to directly compare the sensitivity to those of other models since the class ratio used in this model

TABLE II

CLASSIFICATION PERFORMANCE (MEAN \pm SD), COMPUTATION RESOURCES, AND VALIDATION METHOD FOR SUBJECT-DEPENDENT APPROACH ON CHB-MIT DATASET OF FIVE DIFFERENT METHODS

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-score (Binary) (%)	F1-score (Weighted) (%)	Training Time per Patient (s)	Parameters	Validation Method
FBCSP + SVM [14]	96.21 \pm 3.76	54.36 \pm 30.63	97.09 \pm 3.48	39.37 \pm 27.30	96.63 \pm 3.17	364.20 \pm 285.26	NA	5-fold CV + LOOCV
EWf+RF [25]	99.41	97.91	99.57	NA	NA	NA	NA	10-fold CV
MV-TSK-FS [15]	98.33 \pm 0.18	96.66 \pm 0.14	99.14 \pm 0.14	NA	NA	NA	2,403,159	5-fold CV
CE-stSENNet [16]	96.15 \pm 5.07	92.41	96.05	53.44 \pm 23.52	97.03 \pm 2.97	NA	290,000	LOOCV
EEGNet-8,2 [26]	96.12 \pm 6.09	81.48 \pm 19.14	96.33 \pm 6.15	60.88 \pm 24.74	96.96 \pm 4.92	45.72 \pm 30.44	2,546	5-fold CV + LOOCV
EEGWaveNet	98.39 \pm 2.39	68.94 \pm 21.12	99.25 \pm 0.85	65.54 \pm 22.23	98.01 \pm 3.12	3.46 \pm 0.20	46,776	5-fold CV + LOOCV

Bold represent the best numerical values.

TABLE III

CLASSIFICATION PERFORMANCE (MEAN \pm SD), COMPUTATION RESOURCES, CLASS RATIO, AND VALIDATION METHOD FOR SUBJECT-INDEPENDENT APPROACH ON CHB-MIT, AND TUSZ DATASET OF FOUR DIFFERENT METHODS

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-score (Binary) (%)	F1-score (Weighted) (%)	Training Time (s/epoch)	Class Ratio (Seizure : Normal)	Validation Method
A. CHB-MIT dataset								
MinMaxHist + XGB [28]	91.80	89.67	93.93	NA	NA	NA	1 : 1.8	5-fold CV
STFT-mConvA [6]	94.37 \pm 1.19	NA	NA	85.34 \pm 3.35	NA	NA	6 : 1	5-fold CV
EEGNet-8,2 [26]	88.41 \pm 15.23	65.12 \pm 31.00	88.73 \pm 15.78	31.12 \pm 20.70	91.49 \pm 11.51	24.03	1 : 60	5-fold CV
EEGWaveNet	96.17 \pm 2.95	56.83 \pm 24.44	96.97 \pm 3.13	38.26 \pm 21.10	96.94 \pm 2.03	22.90	1 : 60	5-fold CV
B. TUSZ dataset								
EEGNet-8,2 [26]	66.07 \pm 12.11	29.91 \pm 32.55	90.22 \pm 18.76	47.45 \pm 24.88	68.60 \pm 11.56	4.54	1 : 1.8	5-fold CV
EEGWaveNet	67.68 \pm 13.79	59.21 \pm 22.23	75.30 \pm 13.97	47.55 \pm 23.04	69.07 \pm 13.71	4.83	1 : 1.8	5-fold CV

Bold represent the best numerical values.

TABLE IV

COMPARISON OF CLASSIFICATION PERFORMANCE (MEAN \pm SD) AND FINE-TUNING TIME FOR TL ON CHB-MIT DATASET OF EEGNET-8,2 AND EEGWAVENET

Model	Fine-tune Percentage	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-score (Binary) (%)	F1-score (Weighted) (%)	Fine-tuning Time (s)
EEGNet-8,2 [26]	No fine-tune	88.41 \pm 15.23	65.12 \pm 31.00	88.73 \pm 15.78	31.12 \pm 20.70	91.49 \pm 11.51	0.00
	10	88.32 \pm 7.42	82.63 \pm 18.09	88.34 \pm 7.54	25.52 \pm 18.18	92.31 \pm 4.81	2.62
	30	92.28 \pm 4.52	79.35 \pm 19.25	92.48 \pm 4.55	31.76 \pm 18.76	94.75 \pm 2.83	6.95
	50	93.47 \pm 4.15	78.54 \pm 18.97	93.72 \pm 4.18	35.58 \pm 19.85	95.46 \pm 2.59	11.06
	70	94.06 \pm 3.88	76.73 \pm 19.82	94.37 \pm 3.87	37.13 \pm 20.53	95.81 \pm 2.48	14.83
	90	94.33 \pm 3.89	77.35 \pm 18.20	94.65 \pm 3.90	39.14 \pm 21.25	95.99 \pm 2.42	17.97
	100	96.69 \pm 3.59	77.75 \pm 18.57	95.02 \pm 3.57	40.29 \pm 21.59	96.21 \pm 2.28	19.36
EEGWaveNet	No fine-tune	96.17 \pm 2.95	56.83 \pm 24.44	96.97 \pm 3.13	38.26 \pm 21.10	96.94 \pm 2.03	0.00
	10	86.78 \pm 14.22	86.49 \pm 15.54	81.48 \pm 13.43	29.11 \pm 21.29	90.78 \pm 11.70	1.20
	30	94.42 \pm 6.06	80.30 \pm 20.35	91.45 \pm 7.58	43.20 \pm 23.10	95.96 \pm 3.98	2.82
	50	97.05 \pm 2.68	78.72 \pm 19.06	93.09 \pm 7.88	49.36 \pm 26.04	97.27 \pm 2.70	4.35
	70	96.41 \pm 4.03	76.29 \pm 20.79	94.15 \pm 7.04	51.24 \pm 27.23	97.45 \pm 2.85	5.78
	90	96.75 \pm 4.06	75.52 \pm 19.89	95.40 \pm 5.25	53.04 \pm 26.39	97.50 \pm 2.72	6.95
	100	96.74 \pm 4.04	75.32 \pm 20.46	95.96 \pm 4.38	53.87 \pm 27.11	97.50 \pm 2.72	7.41

Bold represents the best numerical values.

is different. Although the STFT-mConvA was reported with an exceptional F1-score, the class ratio in model training is biased toward the seizure stage. Due to the lack of open source code, we could not reproduce the two aforementioned models to compare with EEGWaveNet directly. EEGWaveNet outperformed other baseline methods on accuracy, specificity, binary F1-score, and weight F1-score. However, the sensitivity of EEGWaveNet is lower than that of EEGNet-8,2. Via independent *t*-test, the accuracy, specificity, binary F1-score of EEGWaveNet is shown to be significantly higher than others baseline methods ($p < 0.05$). Moreover, the training time per epoch of EEGWaveNet is competitive to that of EEGNet-8,2.

The performance of EEGNet-8,2 and EEGWavenet subject-independently on the TUSZ dataset is shown in Table III. From the results, EEGWaveNet significantly outperforms EEGNet-8,2 in all measures including the sensitivity. The achieved result demonstrate a decent starting performance on this previously unexplored task on the challenging dataset. Another interesting aspect is that using EEGWaveNet with the TUSZ dataset yields comparable sensitivity compares to those from the CHB-MIT dataset, while the sensitivity from EEGNet-8,2 decreases by

more than half. These results proved the robustness of the EEGWaveNet.

C. Experiment III–Transfer Learning

Table IV shows both models yields decreased accuracy, specificity, and binary F1-score until reaching 30% fine-tuning. We suspected that an insufficient and unbalanced training sample might cause overfitting to the model. Starting from 30%, all of the metrics mentioned above from both models are improved. In contrast to the other metrics, sensitivity decrease when the fine-tune percentage increase. This is because we center our fine-tuning data selection on the seizure event, as mentioned in Section III-E. At a lower fine-tuning percentage, the majority of selected data are seizure class (positive class), thus enhancing the model's sensitivity. At 100% (one recording), EEGWaveNet reached the highest performance and outperformed EEGNet-8,2 in all metrics except sensitivity. Via paired *t*-test, sensitivity and binary F1-score is shown to be significantly higher ($p < 0.05$) than without fine-tuning. Aside from the classifying performance, we also compare the fine-tuning time of the EEGWaveNet and EEGNet-8,2.

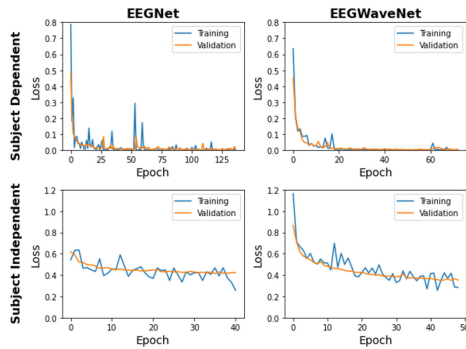


Fig. 3. Training and validation loss of the implemented EEGNet-8,2 and proposed model EEGWaveNet from CHB-MIT dataset (upper) and TUSZ dataset (lower). The plots show loss from the fold with lowest validation loss before early stopping.

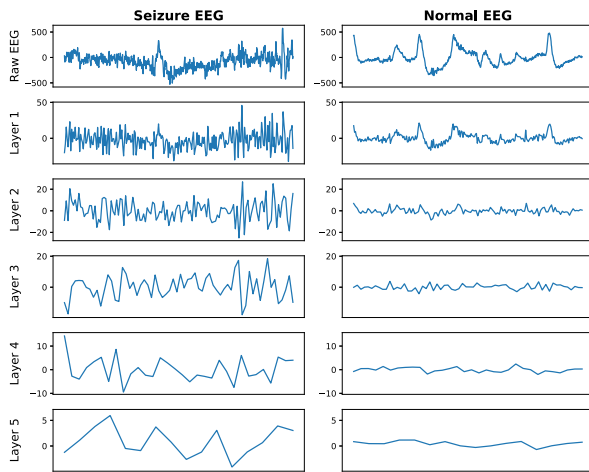


Fig. 4. Comparison of raw EEG and outputs from each layer of multiscale convolution module. Both are from patient #07 of CHB-MIT dataset, and from 1 fold of subject dependent task.

V. DISCUSSION

A. Effectiveness of Multiscale Convolution Module

To illustrate the efficacy of the multiscale convolution module, we arbitrarily picked a seizure and a nonseizure signal sample (4 s duration) of the same recording from the CHB-MIT dataset and passed them through a fully trained multiscale convolution module. Fig. 4 shows a comparison of a seizure and a nonseizure signal with a similar amplitude signature. The convolution produces more response to a seizure sample than a nonseizure one which can be observed through the amplitude of the output signal, especially in the deeper layer. This suggests that the multiscale convolution module not only reduce the resolution of the input signal, but also provide more discriminative features for seizure detection.

B. TL Performance

As demonstrated in Table IV, fine-tuning EEGWaveNet with more than 50% of a 1-h recording improves all metrics except specificity. Interestingly, the change in accuracy, specificity, and weight F1-score are insignificant, but the difference can be clearly seen in the binary F1-score and sensitivity. Since

binary F1-score and sensitivity are viable measures in clinical circumstances, a significant improvement in these metrics are precious for clinical usage. The improvement in a similar manner can be observed on EEGNet-8,2 as well. This suggests that the proposed fine-tuning methodology can be generally applied, which might be a key to improve the feasibility of AI toward healthcare 5.0.

Another essential point from the result is that fine-tuning with 10% of a 1-h recording significantly improves the model's sensitivity. However, this comes with the tradeoff of decreasing in all other metrics. Depending on the application, greatly enhance the sensitivity despite the tradeoff can be useful.

C. Model Compactness

EEGWaveNet can be considered lightweight comparable to EEGNet-8,2. Despite having more parameters, the training time in the subject-dependent approach is comparable to EEGNet-8,2. In Fig. 3, we trained the models until early stop occurred (validation loss reach minimum value and stop reducing for 10 epochs). Fig. 3 (upper) demonstrates that the EEGWaveNet can achieve optimal performance earlier than EEGNet-8,2 in the subject-dependent approach under a certain hyperparameter configuration. As shown in Table III, EEGNet-8,2's and EEGWaveNet's training time per epoch are almost equal; thus, training with more epoch results in a longer training time. The same goes for the TL experiment, where EEGWaveNet provides more than two times faster fine-tuning time than EEGNet-8,2. These results support that EEGWaveNet exhibits faster learning capability compared to EEGNet-8,2.

D. Advantage of EEGWaveNet

We believe that our model has potential for practical usage. Our model's advantages over the other state-of-the-art models are as follows; first, most state-of-the-art models are trained and evaluated only in a subject-dependent manner. On the other hand, our model is proved to perform comparable to others in the subject-dependent approach and outperform the others in the subject-independent approach. Second, our model is an end-to-end network that eliminates most preprocessing of the EEG recording. Most subject-independent models need extensive data preprocessing and feature engineering, which is time-consuming and difficult to generalize. In contrast, our model demonstrates an ability to generalize, proven by the results from three different datasets.

However, the main limitation of EEGWaveNet is achieving a worse performance but with higher complexity in the case of the small dataset. Check the extensive study of this issue in the Appendix.

E. Practical Feasibility in Healthcare 5.0

The proposed fine-tuning allows the model's performance to be significantly improved by training with only a small amount of data from an individual subject. Despite being slower than a fully subject-independent approach, gaining more reliable results is a reasonable tradeoff. The fine-tuning process only took a few seconds in our setup, which is in a reasonable margin.

In a practical scenario using the TL approach, when the clinician needs to diagnose a new patient, the clinician only has to manually label one seizure onset and nonseizure periods, which could be done easily in a matter of seconds. Then, fine-tuning the pre-trained model using the labeled recording takes a few seconds, and the model is ready. This process can reduce the clinician's workload compared to the subject-dependent approach, where the clinician needs to label all of the recordings before training just for an individual patient. In addition, the model pretrained in a subject-independent manner can be retrained when there are more available data to improve the performance further. This demonstrates a good step toward healthcare 5.0, where the AI leverage the workload of clinician rendering the workflow to become faster.

For future works, there is a possibility to improve subject-independent approach and TL performance by increasing the number of training data using data augmentation or including the data from subjects across different datasets. Another interesting option is to further generalize EEGWaveNet in other tasks such as classifying patients with Alzheimer's disease [29] and motor imagery works [30].

VI. CONCLUSION

This article proposed EEGWaveNet, a novel end-to-end multiscale resolution DCNN architecture for automatic seizure detection. EEGWaveNet consist of three modules; multiscale convolution, spatial-temporal feature extraction, and classifier. First, multiscale convolution separated the input signal into multiple resolutions. Then, the spatial-temporal feature extraction was applied to extract spatial-temporal features from the previous extracted output. Difference input resolution provides a different field of view for temporal feature extraction. Last, FC layers were implemented as a binary classifier to detect the seizure. Our proposed model accomplishes a result comparable to other state-of-the-art models in subject-dependent and outperforms other state-of-the-art models in subject-independent approaches. EEGWaveNet also exhibits a decent performance in the high variance dataset, TUSZ, which proves its ability to generalize. With TL from a subject-independent approach, the performance of EEGWaveNet can be improved while sustaining the practicability for practical usage. This proves the application of EEGWaveNet in the possibility for the future, where the AI collaborates with clinicians in clinical diagnosis.

APPENDIX

A. Experiment With BONN Dataset

Aside from CHB-MIT and TUSZ dataset, we also evaluate EEGWaveNet with BONN dataset [31]. However, we decide to discuss the BONN dataset results in this section because the BONN dataset is too small and unable to be used in the subject-independent approach, which would be an out-of-scope study for this article.

1) **BONN Dataset:** This dataset is publicly available EEG signals produced by the university of BONN [31]. The dataset consists of 500 EEG signal segments separated into five groups;

TABLE V
CLASSIFICATION PERFORMANCE (MEAN \pm SD), EVALUATION METHOD, END-TO-END AND FEATURE EXTRACTION ON BONN DATASET OF TWO DIFFERENT METHODS

Performance metric	Case	TQWT [32]	CEEMDAN Based Models [33]–[35]	EEGWaveNet
Accuracy (%)	I (E vs CD vs AB)	98.40	97.6, 86.37, 97.6	95.27 \pm 0.74
	II (E vs ABCD)	99.60	99.2, 98.87, 99.2	98.19 \pm 0.55
	III (E vs A)	100	100, 100, 100	99.89 \pm 0.15
	IV (E vs D vs A)	98.67	99, NA, 98.67	95.74 \pm 0.61
	V (E vs D)	100	97, 97.15, 100	97.29 \pm 0.88
	VI (E vs C)	100	100, 100, 99	98.26 \pm 0.32
Sensitivity (%)	I (E vs CD vs AB)	98.33	98.11, NA, 97.67	95.47 \pm 0.73
	II (E vs ABCD)	99.49	99.17, NA, 98.75	97.38 \pm 0.97
	III (E vs A)	100	100, NA, 100	99.80 \pm 0.31
	IV (E vs D vs A)	96.67	99.2, NA, 98.67	95.75 \pm 0.58
	V (E vs D)	100	99.4, NA, 100	97.11 \pm 1.92
	VI (E vs C)	100	100, NA, 100	98.07 \pm 0.48
Specificity (%)	I (E vs CD vs AB)	98.60	96.93, NA, 97.35	97.45 \pm 0.40
	II (E vs ABCD)	100	99.04, NA, 98.75	98.40 \pm 0.69
	III (E vs A)	100	100, NA, 100	99.97 \pm 0.08
	IV (E vs D vs A)	98	99.5, NA, 98.72	97.84 \pm 0.31
	V (E vs D)	100	98.25, NA, 100	97.11 \pm 1.50
	VI (E vs C)	100	100, NA, 98	98.45 \pm 0.60
Evaluation method (Training : Testing)		50:50	50:50	50:50
End-to-end		No	No	Yes
Feature extraction		Hand-design	Hand-design	Automatic

A, B, C, D, and E. Group A and B represent the surface EEG recordings of five healthy wakeful volunteers with eyes open (A) and close states (B). Group C, D, and E are obtained from five patients with epilepsy. Group C contains the EEG records obtained from the hippocampal formation of the opposite hemisphere of the brain, while group D contains the records in the epileptogenic zone. In this dataset, only group E contains epochs with seizure activities. Each group contains 100 segmented single-channel EEG signals with the sampling rate is 173.61 Hz and the duration of 23.6 s.

2) **Experimental Evaluation:** In this article, we classified into six cases with different groups of study regarding the clinical relevance. Case I and IV investigate three classes of EEG signals: ictal, inter-ictal, and normal, while Case V and VI determine only two classes: ictal and interictal. Besides, Case II and III consider the seizure and nonseizure EEG segments, in which the nonseizure signal is derived from interictal and normal signals in Case II, and only normal signal in Case III. We randomly selected half of the dataset as a training set and the rest as a testing set. After experimenting with 20 runs, we reported the average performance of the model. This evaluation method is similar to two state-of-the-art methods [32], [33].

3) **Experimental Result and Discussion:** Table V shows the performance of EEGWaveNet, and other state-of-the-art models [32]–[35] on the BONN dataset. By comparing the accuracy, sensitivity, and specificity, it is observed that EEGWaveNet's performance is just slightly lower than the others. This is because EEGWaveNet is the DL-based model, which requires a large dataset for excellent performance, while TQWT and the CEEMDAN-based models are the ML-based models, which can perform more efficiently on the small dataset. Despite this weakness, the performance of EEGWaveNet is comparable to the others, proving its high efficiency and ability to generalize.

Aside from the performance, the model needs to be practical for real-world application. The ML-based models such

as the TQWT model and the CEEMDAN-based models need a hand-design feature extraction method to detect the seizure efficiently. This could be difficult to use when changing the dataset since some variables need to be empirically set, such as sub-band selection. In contrast, our proposed end-to-end model has automatic feature extraction, EEGWaveNet is more advantageous than other state-of-the-art models in real-world practical usage.

REFERENCES

- [1] E. L. Johnson, "Seizures and epilepsy," *Med. Clin.*, vol. 103, no. 2, pp. 309–324, 2019.
- [2] R. S. Fisher *et al.*, "ILAE official report: A practical clinical definition of epilepsy," *Epilepsia*, vol. 55, no. 4, pp. 475–482, 2014.
- [3] L. Kuhlmann, K. Lehnertz, M. P. Richardson, B. Schelter, and H. P. Zaveri, "Seizure prediction-ready for a new era," *Nature Rev. Neurol.*, vol. 14, no. 10, pp. 618–630, 2018.
- [4] K. Tse, S. Puttachary, E. Beamer, G. J. Sills, and T. Thippeswamy, "Advantages of repeated low dose against single high dose of kainate in C57BL/6J mouse model of status epilepticus: Behavioral and electroencephalographic studies," *PLoS One*, vol. 9, no. 5, 2014, Art. no. e96622.
- [5] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Comput. Biol. Med.*, vol. 100, pp. 270–278, 2018.
- [6] Y. Yuan, G. Xun, K. Jia, and A. Zhang, "A multi-view deep learning framework for EEG seizure detection," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 1, pp. 83–94, Jan. 2019.
- [7] H. Khan, L. Marcuse, M. Fields, K. Swann, and B. Yener, "Focal onset seizure prediction using convolutional networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 2109–2118, Sep. 2018.
- [8] T. Dissanayake, T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Deep learning for patient-independent epileptic seizure prediction using scalp EEG signals," *IEEE Sensors J.*, vol. 21, no. 7, pp. 9377–9388, Apr. 2021.
- [9] Y. Jiang *et al.*, "Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 12, pp. 2270–2284, Dec. 2017.
- [10] Y. Zhang, Y. Guo, P. Yang, W. Chen, and B. Lo, "Epilepsy seizure prediction on EEG using common spatial pattern and convolutional neural network," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 465–474, Feb. 2020.
- [11] D. Zeng, K. Huang, C. Xu, H. Shen, and Z. Chen, "Hierarchy graph convolution network and tree classification for epileptic detection on electroencephalography signals," *IEEE Trans. Cogn. Develop. Syst.*, vol. 13, no. 4, pp. 955–968, Dec. 2021, doi: [10.1109/TCDS.2020.3012278](https://doi.org/10.1109/TCDS.2020.3012278).
- [12] H. Daoud and M. Bayoumi, "Efficient epileptic seizure prediction based on deep learning," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 5, pp. 804–813, Oct. 2019.
- [13] L. Tang, M. Zhao, and X. Wu, "Accurate classification of epilepsy seizure types using wavelet packet decomposition and local detrended fluctuation analysis," *Electron. Lett.*, vol. 56, no. 17, pp. 861–863, 2020.
- [14] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intell.)*, 2008, pp. 2390–2397.
- [15] Z. Tian *et al.*, "Deep multi-view feature learning for EEG-based epileptic seizure detection," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 10, pp. 1962–1972, Oct. 2019.
- [16] Y. Li, Y. Liu, W. G. Cui, Y. Z. Guo, H. Huang, and Z. Y. Hu, "Epileptic seizure detection in EEG signals using a unified temporal-spectral squeeze-and-excitation network," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 4, pp. 782–794, Apr. 2020.
- [17] N. Banluesombatkul *et al.*, "Metasleeplearner: A pilot study on fast adaptation of bio-signals-based sleep stage classifier to new individual subject using meta-learning," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 6, pp. 1949–1963, Jun. 2021.
- [18] A. Dittaporn, N. Banluesombatkul, S. Kettrat, E. Chuangsuwanich, and T. Wilaiprasitporn, "Universal joint feature extraction for P300 EEG classification using multi-task autoencoder," *IEEE Access*, vol. 7, pp. 68415–68428, 2019.
- [19] A. Agrawal, G. C. Jana, and P. Gupta, "A deep transfer learning approach for seizure detection using RGB features of epileptic electroencephalogram signals," in *Proc. IEEE Int. Conf. Cloud Comput. Technol. Sci.*, 2019, pp. 367–373.
- [20] A. H. Shueb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, 2009.
- [21] V. Shah *et al.*, "The temple university hospital seizure detection corpus," *Front. Neuroinform.*, vol. 12, 2018, Art. no. 83.
- [22] W. Ko, E. Jeon, S. Jeong, and H.-I. Suk, "Multi-scale neural network for EEG representation learning in BCI," *IEEE Comput. Intell. Mag.*, vol. 16, no. 2, pp. 31–45, May 2021.
- [23] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2117–2125.
- [24] Pytorch, "EarlyStopping," Accessed: Nov. 17, 2021. [Online]. Available: https://pytorch.org/ignite/generated/ignite.handlers.early_stopping.EarlyStopping.html
- [25] A. Bhattacharyya and R. B. Pachori, "A multivariate approach for patient-specific EEG seizure detection using empirical wavelet transform," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2003–2015, Sep. 2017.
- [26] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "Eegnet: A compact convolutional neural network for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, 2018, Art. no. 056013.
- [27] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.
- [28] S. Yang *et al.*, "Selection of features for patient-independent detection of seizure events using scalp EEG signals," *Comput. Biol. Med.*, vol. 119, 2020, Art. no. 103671.
- [29] N. Mammone *et al.*, "Brain network analysis of compressive sensed high-density EEG signals in AD and MCI subjects," *IEEE Trans. Ind. Informat.*, vol. 15, no. 1, pp. 527–536, Jan. 2019.
- [30] M. H. Bhatti *et al.*, "Soft computing-based EEG classification by optimal feature selection and neural networks," *IEEE Trans. Ind. Informat.*, vol. 15, no. 10, pp. 5747–5754, Oct. 2019.
- [31] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Phys. Rev. E*, vol. 64, no. 6, 2001, Art. no. 061907.
- [32] R. H. Ahnaf, S. Siuly, and Z. Yanchun, "Epileptic seizure detection in EEG signals using tunable-q factor wavelet transform and bootstrap aggregating," *Comput. Methods Prog. Biomed.*, vol. 137, pp. 247–259, 2016.
- [33] R. H. Ahnaf and S. Abdulhamit, "Automatic identification of epileptic seizures from EEG signals using linear programming boosting," *Comput. Methods Prog. Biomed.*, vol. 136, pp. 65–77, 2016.
- [34] A. R. Hassan and M. A. Haque, "Epilepsy and seizure detection using statistical features in the complete ensemble empirical mode decomposition domain," in *Proc. IEEE Region Ten Conf.*, 2015, pp. 1–6.
- [35] A. R. Hassan, A. Subasi, and Y. Zhang, "Epilepsy seizure detection using complete ensemble empirical mode decomposition with adaptive noise," *Knowl.-Based Syst.*, vol. 191, 2020, Art. no. 105333.



Punyawish Thuwajit is currently a High School Student with Suankularb Wittayalai School, Bangkok, Thailand, Governmental Scholar of the Development and Promotion of Science and Technology Talents (DPST) scholarship, and an intern with the Information Science and Technology, Vidyasirimedhi Institute of Science and Technology, Rayong, Thailand, where he works in the field of biomedical signal processing with artificial intelligence, aimed toward artificial intelligence technologies in healthcare. He is also a Freelance Artificial Intelligence Developer, working mainly in integrating natural language processing techniques with topic landscape and weak signal detection.



Phurin Rangpong was born in Chonburi, Thailand, in 1997. He received the B.E. degree in robotics and automation engineering from the King Mongkut's University of Technology Thonburi, Bangkok, Thailand, in 2019, and the master's degree in engineering sciences and design from the Transdisciplinary Sciences and Engineering Department, School of Environment and Society, Tokyo Institute of Technology, Tokyo, Japan, in 2021.

From 2020 to 2021, he worked as a Research Assistant with Bio-Inspired Robotics and Neural Engineering Laboratory, Vidyasirimedhi Institute of Science and Technology, Rayong, Thailand. His current research interests include human–computer interaction, brain–computer interface, virtual reality technology, biosignal processing, bio-inspired robotics, and mobile robotics.



Phattarapong Sawangjai received the B.E. degree in biomedical engineering from Mahidol University, Salaya, Thailand, in 2018. He is currently working toward the Ph.D. degree in information science and technology with the Bio-Inspired Robotics and Neural Engineering (BRAIN) Laboratory, School of Information Science and Technology, Vidyasirimedhi Institute of Science and Technology (VISTEC), Rayong, Thailand.

His current research interests include biosignal processing, deep learning for time-series data applications, and brain–computer interface (BCI).



Phairot Autthasan received the B.S. degree in pure chemistry from Chemistry Department, King Mongkut's University of Technology Thonburi, Bangkok, Thailand, in 2017. He is currently working toward the Ph.D. degree in information science and technology with the Bio-Inspired Robotics and Neural Engineering (BRAIN) Laboratory, School of Information Science and Technology, Vidyasirimedhi Institute of Science and Technology (VISTEC), Rayong, Thailand.

His current research interests include machine learning for bio-signals, brain–computer interface (BCI), neural signal processing, and human–machine interaction.



Rattanaphon Chaisaen received the B.Sc. degree in computer science from Khon Kaen University, Khon Kaen, Thailand, in 2018. He is currently working toward the Ph.D. degree in information science and technology with the Bio-Inspired Robotics and Neural Engineering (BRAIN) Laboratory, School of Information Science and Technology, Vidyasirimedhi Institute of Science and Technology (VISTEC), Rayong, Thailand.

His current research interests include machine learning approaches applied to biosignals, brain–computer interfaces, and assistive technology.



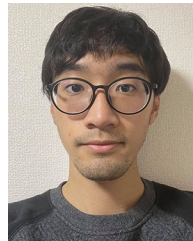
Nannapas Banluesombatkul is currently working toward the Ph.D. degree in information science and technology with the Bio-Inspired Robotics and Neural Engineering (BRAIN) Laboratory, School of Information Science and Technology, Vidyasirimedhi Institute of Science and Technology (VISTEC), Rayong, Thailand.

Her research interests include sleep stages-related topics and sleep diagnosis tools using deep learning and machine learning approaches.



Puttaranun Boonchit was born in Udon Thani, Thailand, in 2000. She is currently working toward the degree majoring in transdisciplinary science and engineering with the School of Environment and Society, Tokyo Institute of Technology, Tokyo, Japan.

With the aim of a sustainable society, she is studying various engineering fields, across multiple disciplinary boundaries. She is going to attend the laboratory about radio wave propagation and the antennas for broadband wireless and mobile communication systems, in 2022. Her research interests include applied data science and machine learning approaches.

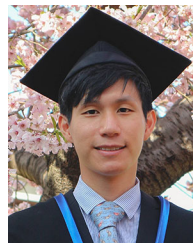


Nattasate Tatsaringkansakul was born in Bangkok, Thailand, in 2000. He is currently working toward the undergraduate degree with the Department of Transdisciplinary of Science and Engineering, Tokyo Institute of Technology, Tokyo, Japan.

The focus of the undergraduate degree is understanding concepts across multiple scientific fields. In 2022, he is going to attend the laboratory about innovation by communication design. His research interests include machine learning

and technology alongside with healthcare.

His research interests include machine learning and technology alongside with healthcare.



Thapanun Sudhawiyaikul received the B.Eng. degree in mechatronic engineering and the M.Eng. degree in control engineering from the King Mongkut's Institute of Technology Ladkrabang (KMITL), Bangkok, Thailand, in 2013 and 2015, respectively, and the D.Eng. degree in mechanical engineering from the Tokyo Institute of Technology, Tokyo, Japan, in 2020.

He is currently a Postdoctoral Researcher with the Vidyasirimedhi Institute of Science and Technology, Rayong, Thailand. His current research interests include AI in healthcare, brain–computer interface, machine monitoring, and MEMS.



Theerawit Wilaiprasitporn (Member, IEEE) received the Ph.D. degree in engineering from the Graduate School of Information Science and Engineering, Tokyo Institute of Technology, Tokyo, Japan, in 2017.

He is currently an Assistant Professor with the School of Information Science and Technology, Vidyasirimedhi Institute of Science and Technology, Rayong, Thailand, where he is also a co-PI of the Bio-Inspired Robotics and Neural Engineering Laboratory. His current research

interests include biomedical signal processing, brain–computer interfaces, and AI in healthcare.

Prof. Wilaiprasitporn serves as a Guest Editor for the IEEE SENSORS JOURNAL and the Track Chair of IEEE SENSORS 2021–2022.