RESEARCH ARTICLE

Less Parameterization Inception-Based End to End CNN Model for EEG Seizure Detection

KUO-KAI SHYU, (Senior Member, IEEE), SZU-CHI HUANG,
LUNG-HAO LEE, (Member, IEEE), AND PO-LEI LEE, (Member, IEEE)
Department of Electrical Engineering, National Central University, Taoyuan 32001, Taiwan

Corresponding author: Szu-Chi Huang (r310102930719@gmail.com)

ABSTRACT Many deep-learning-based seizure detection algorithms have achieved good classification, which usually outperformed traditional machine-learning-based algorithms. However, the hand-engineered features increase the computational complexity and potentially have an ineffectiveness problem for the category. Therefore, this paper proposes a novel end-to-end deep-learning model comprising an inception module and a residual module to analyze the multi-scales of original EEG signals and realize seizure detection without feature extraction. Experiments were conducted and evaluated on the Bonn dataset and the CHB-MIT dataset. In the subject-dependent experiments, our model achieved an average F1-score of 69.34% on the CHB-MIT dataset. In subject-independent experiments, our method achieved an average accuracy of 99.04% on the Bonn dataset and an average F1-score of 37.31% on the CHB-MIT dataset. A series of analyses confirmed that our proposed model has better classification performance and lower computational complexity than existing end-to-end seizure detection models.

INDEX TERMS Convolutional neural networks, seizure detection, epilepsy, electroencephalography, end-to-end model.

I. INTRODUCTION

Epilepsy is a chronic neurological disease affecting about 50 million people worldwide [1]. Abnormal discharge of brain cells during epileptic seizures causes symptoms such as convulsions, fainting, loss of consciousness, behavior changes, etc. These symptoms may not only cause physical damage to patients but may also cause psychological problems in the long run [2]. Although most patients' epilepsy symptoms can be effectively improved with antiepileptic drugs. However, about 20-40% of patients with severe epilepsy may need surgical treatment, and an accurate diagnosis of epilepsy is a necessary preoperative evaluation [3].

Electroencephalography (EEG) a method for recording brain electrical activity is often used in diagnosing brain diseases [4]. The traditional epilepsy diagnosis method is manual detection by visual observation of EEG signals by neurologists. This method is time-consuming and the accuracy of diagnosis highly depends on the experience and ability

The associate editor coordinating the review of this manuscript and approving it for publication was Larbi Boubchir.

of the diagnostician. Long-term EEG diagnosis is more prone to human negligence resulting in misdiagnosis or delayed diagnosis aggravating the condition [5].

In recent years, many automated seizure detection technologies have been proposed for epilepsy diagnosis assistance systems [6], [7] and seizure alarm systems [8], [9]. The epilepsy diagnostic assistance system can increase the accuracy of epilepsy diagnosis and shorten the diagnosis time. The seizure alarm systems can immediately notify relevant units when a seizure occurs. Patients can quickly obtain treatment and the sudden unexpected death in epilepsy (SUDEP) can be reduced [10].

Seizure detection is regarded as a classification problem. Those related algorithms can be roughly divided into two categories: traditional machine learning and deep learning. The algorithm based on traditional machine learning generally includes pre-processing, feature extraction, feature selection, classifier, and post-processing. For example, Chen et al. [6] used discrete wavelet transformation (DWT) and calculates a variety of different types of entropy (Shannon entropy, sample entropy, fuzzy entropy, etc.) as features. Then, analysis

of variance (ANOVA) is used to sort and select features. Finally, the least squares support vector machine (LS-SVM) was used as a classifier to identify whether the input signal was a seizure. Olokodana et al. [10] used DWT soft threshold technology to filter out signal noise, and then calculate fractal dimension, Hjorth parameter, and singular value decomposition entropy as signal features. Finally, the Kriging model is applied for classification. Tiwari et al. [11] used the difference of Gaussian (DoG) to identify the key positions in the EEG signal, then calculates the local binary pattern (LBP) as a feature. Finally, the support vector machine (SVM) is used as a classifier.

Deep learning techniques have been widely applied in many non-stationary signal processing tasks. For example, Lopac et al. [12] used Cohen's class time-frequency representations and convolutional neural network (CNN) models to detect gravitational waves. Arias-Vergara et al. [13] used a CNN to detect speech deficits and classify phone attributes based on convolutional gated recurrent units (CGRU). Sakib et al. [14] used a CNN model to analyze raw Electrocardiogram (ECG) signals for arrhythmia detection on the embedded system. In recent years, many seizure detection algorithms are also based on deep learning models, such as long short-term memory (LSTM) [15], gated recurrent units (GRU) [16], CNN [17], [18], [19], [20], [21], [22], [23] etc. The CNN-based algorithm is the most commonly used model architecture. The EEG is converted into two-dimensional feature images in the pre-processing and then input into CNN for classification. For instance, Rashed-Al-Mahfuz et al. [17] used short-time Fourier transform (STFT) and continuous wavelet transform (CWT) to convert EEG into a two-dimensional time-frequency map and then used pre-trained VGG16 [24] and ResNet50 [25] for classification. Lai et al. [18] used a bandpass filter to obtain 80-250Hz and 250-500Hz signals in EEG, and then the candidate regions are identified by calculating the short-time energy. Finally, CWT was used to convert the EEG of candidate regions into time-frequency images and input shallow CNN with a depth of 5 layers for classification.

However, hand-engineered feature representations may ignore temporal patterns of the original EEG signals and have ineffectiveness issues for classification [19]. Design an end-to-end model that directly uses the original EEG signal as the model input. For instance, Acharya et al. [20] analyzed raw EEG signals using a 13-layer 1D-CNN. Wang et al. [21] extracted high-level representations from EEG through two convolution blocks with different sizes of the kernels, and output with a fully connected layer. Thuwajit et al. [22] used a multiscale architecture to analyze the different ranges of time domain information of EEG. Duan et al. [23] used the Siamese network architecture to train the embedding module. Through the embedding module, EEG can be converted to the embedding space, and the classification can be performed in the embedding space. The main contributions of this paper are summarized as follows:

- We propose a novel end-to-end deep learning model to detect seizures based on the ResNet and Inception Net [26]. Unlike the previous end-to-end models that use different scale information independently, we use the inception module to integrate multi-scale features effectively.
- Besides using traditional metrics accuracy, sensitivity and specificity for seizure detection evaluation, we use the metric F1-score to measure and compare model performance. Experimental results show the F1-score is a more helpful metric than others, especially on the class imbalanced datasets.
- We conduct experiments and analyses on two datasets to show that our proposed model has promising classification performance and lower computational complexity than existing end-to-end models.

The rest of this paper is organized as follows. Section II describes the details of our proposed method. Section III specifies the experimental setups, including the dataset description, classification tasks, baseline methods, and evaluation metrics. Section IV presents experimental results for model performance evaluation and comparisons. Section V introduces the findings for further analysis and discussion. Conclusions with future directions are finally drawn in Section VI.

II. METHODOLOGY
A. PROPOSED MODEL ARCHITECTURE
This paper uses the inception module, residual module, and global average pooling as feature extractors to generate feature vectors from EEG. Using a two-layer fully connected layer as a classifier to output the EEG is inter-ictal or seizure. The overall model architecture is shown in Fig. 1. The EEG signals will first pass through three inception and residual modules for feature extraction. We then perform the global average pooling to obtain the output using the fully-connected layer. The inception module allows the model to have multi-scale analysis and integration. The residual module effectively maintains the gradient process during the model training phase. Finally, the fully-connected layer classifies the features extracted by the inception and residual modules.

1) INCEPTION MODULE
Many seizure detection studies have achieved better classification performance through multi-scale feature extraction [19], [21], [22]. Multi-scale feature extraction enables the model to analyze EEG signals in different ranges, obtaining comprehensive features with local and global information. We use the inception module to extract multi-scale features through convolutional layers with different convolution kernel sizes. Compared with the previous multi-scale feature extraction, the inception module integrates features in different scales and reduces the number of model parameters [27].
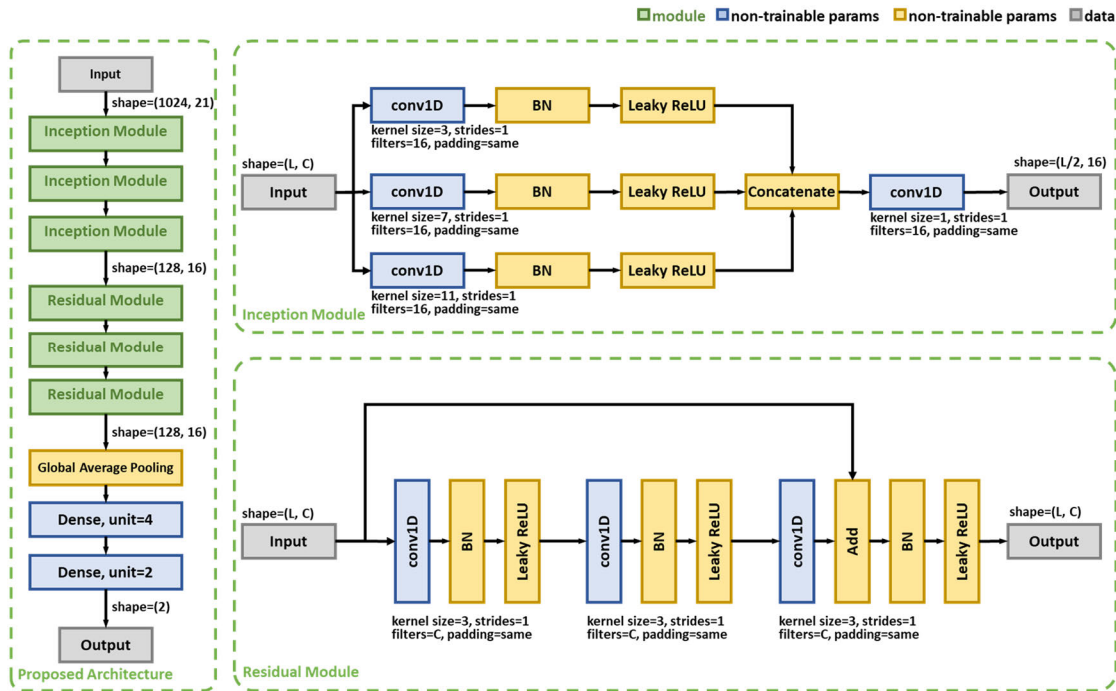
FIGURE 1. Model architecture of our proposed method.

Our used inception module comprises three parallel convolutional blocks and a convolutional layer with a convolution kernel size of one. Each convolutional block is deduced by a convolutional layer (stride is 2, padding mode is the same, and the number of kernels is 16), batch normalization, and leaky ReLu activation function. The convolution kernel sizes of the three blocks are 3, 7, and 11, respectively. The output of the three parallel convolutional layers is concatenated and then passed through the convolutional layer (kernel size is one, and the number of kernels is 16).

2) RESIDUAL MODULE

The shortcut connections architecture of the residual module effectively maintains the gradient process during training to avoid the problem of exploding or vanishing gradients [25]. In addition, the skip connection mitigates the information loss after the signal passes through the convolutional layers so that the extracted features retain the original signal information [28].

Our used residual module is composed of three stacked convolution blocks. We use the output of the third convolution Block in the inception module as the input of our residual module. Each convolution block comprises a convolution layer with a convolution kernel size of three, batch normalization, and leaky ReLu.

3) CLASSIFIER

We use the feature vectors generated by the inception and residual modules to train the classifier comprising a two-layer fully-connected layer for seizure detection. The size of the first layer is four units, and the activation function is ReLu. The size of the second layer is two units, and the activation function is softmax.

During the training phase, if an EEG record is annotated with seizure, the class is assigned as 1, and 0 otherwise. To prevent abnormal stopping caused by unstable loss, we apply the early stopping mechanism when the epoch exceeds the threshold (i.e., I_{min}). The training loss function is categorical cross-entropy. The optimizer is Adam (epsilon=0.1, clipnorm=1), and the learning rate uses cosine decay restart [29]. We summarize the model parameters in Table 1. To detect an instance during the testing phase, we use the class with the more significant probability than the other as model output.

III. EXPERIMENT

A. DATABASE

We use the CHB-MIT and Bonn datasets to verify the model performance. The CHB-MIT dataset contains long-term EEG record files with time characteristics. In addition, its imbalanced class distribution may naturally reflect real-life seizure detection challenges. The Bonn dataset contains the segmented EEG data for fair performance comparisons, which gets rid of the bias caused by different pre-processing. Details are introduced as follows

1) CHB-MIT DATASET

The CHB-MIT dataset, a scalp EEG (sEEG) dataset from Children's Hospital Boston, was sampled using 16-bit

TABLE 1. Training parameters of our proposed method.

	Subject Dependent	Subject-Independent	
	CHB-MIT	CHB-MIT	Bonn
Input Signal Length (samples)	1024	1024	4096
Input Channels	20	20	1
Batch Size	128	128	32
Class weights	Balance	Balance	Balance
CDR[1] initial learning rate	1e-3	1e-5	1e-3
CDR[1] T_0	5	5	5
CDR[1] T_{mult}	2	2	2
Early stopping I_{min}	30	30	30
Early stopping patience	30	30	30

1. CDR: cosine decay restart

resolution and 256Hz [30]. The CHB-MIT dataset is a long-term EEG dataset that is commonly used to evaluate seizure detection algorithms. The entire dataset contained EEG recordings of 24 epileptic patients aged 3-22 years with a total of 198 epileptic events and 644. The edf files were each about 1 hour long.

In this paper, according to the baseline studies [22], the following preprocessing is carried out:

1. Select FP1-F7, F7-T7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, and "T8-P8-0", "P8-O2", "FZ-CZ", "CZ-PZ", "P7-T7", "T7-FT9", "FT9-FT10", and "FT10-T8" 21 common channels.
2. Split the record into 4-second segments with 1-second overlaps between each segment.
3. Use a low-pass filter to remove noise above 64 Hz.

Files chb12_27.edf, chb12_28.edf, and chb12_29.edf were excluded from the experiment because they did not contain any common channels.

2) BONN DATASET

The Bonn dataset, an epilepsy dataset provided by the University of Bonn, was sampled using 12-bit resolution at 173.61Hz [31]. The data set consists of 500 records and each record is a single channel with 4096 sampling points. The duration is about 23.6 seconds and can be divided into five categories: A to E. Class A and Class B signals were collected from sEEG signals of five healthy awake subjects. Class C, Class D, and Class E were collected from intracranial EEG (iEEG) signals of five epileptic patients.

Class A is the signal when the eyes are opened. Class B is the signal when the eyes are closed. Class C and D are the signals when the inter-ictal. Class D measures the epileptogenic zone. Class C measured the hippocampal formation of the opposite hemisphere of the epileptic seizure region. Class E is the signal during an epileptic seizure.

B. CLASSIFICATION TASK

1) EXPERIMENT I–SUBJECT-DEPENDENT

Subject dependent experiment of this paper is validated in the CHB-MIT dataset and leave-one-record-out cross-validation was used. First, one record file is reserved as the testing set and the rest files are used as the training set. Then, 20% of the training set data are randomly sampled as a validation set.

2) EXPERIMENT II–SUBJECT-INDEPENDENT

Subject-independent experiment of this paper is validated in the CHB-MIT dataset and the Bonn dataset. The CHB-MIT dataset uses leave-one-subject-out cross-validation. First, one subject is reserved as the testing set, and the rest subjects are used as the training set. Then 20% of the training set data are randomly sampled as a validation set.

For Bonn Dataset, we use ten-fold cross-validation and eight tasks in Table 2 were performed to evaluate the model. Case1, Case2, and Case5 compare the EEG of healthy subjects with the EEG of seizure. Case3, Case4, and Case6 compared EEG in epileptic patients with inter-ictal and during seizures. Case7 and Case8 contain EEG signals from healthy subjects and people with epilepsy. The epileptic EEG, non-epileptic EEG, and healthy EEG were compared.

TABLE 2. Classification task of the Bonn dataset.

Classification Case	Dataset Combinations	Number of EEG signals
Case 1	E vs A	200
Case 2	E vs B	200
Case 3	E vs C	200
Case 4	E vs D	200
Case 5	E vs AB	300
Case 6	E vs CD	300
Case 7	E vs ABCD	500
Case 8	E vs AB vs CD	500

C. BASELINE METHOD

We compared with one traditional machine learning algorithm and three end-to-end deep learning models to show the classification performance of the proposed method. The following is a brief introduction to each method.

1) FBCSP+SVM

The common spatial pattern (CSP) is a classic feature extraction method that is applied to emotion recognition [32], [33], imaginary movement [34], [35], seizure detection [36], [37], etc. Brain-computer interface (BCI).

The filter bank common spatial pattern (FBCSP) divides the signals into different frequency bands through multiple filter banks, and then multiplies the projection matrix with the

EEG signals by CSP algorithm to maximize the difference between each category of EEG signals. Finally extract the most discriminant part as the feature. The bandpass filter bands of the filter banks used in this paper are respectively 4-8Hz, 8-16Hz, 16-32Hz, and 32-55Hz. The kernel function of SVM is "rbf".

2) EEGNET-8,2

EEGNet [38] is a convolutional neural network with a compact architecture, and its architecture is formed by stacking three convolutional layers. EEGNet makes each layer of the network learn different features by designing different convolution modes and sizes of the convolution kernel. The first layer uses 2D convolution to learn frequency filters, and the second layer uses Depthwise convolution to learn frequency-specific spatial filters. The third layer uses Separable convolution to learn the integration and optimization of time-frequency and spatial features. Finally, it is output as a model through a fully connected layer. This paper uses EEGNet-8,2 which has 8 frequency filters and 2 spatial filters as the baseline model.

3) STACKED 1D-CNN

Stacked 1D-CNN [21] is an end-to-end model, which uses two parallel convolutional blocks to extract features from the EEG and then outputs them through a fully connected layer. Two convolution blocks are formed by repeated stacking of convolutional layers, Batch Normalization, and Max Pooling three times, and the number of convolutional kernels is 32, 64, and 128, respectively. The kernel size of the convolutional layer in convolution block 1 is 3. In convolution block 2, the kernel size of the first two convolutional layers is 5, and the kernel size of the last convolutional layer is 3. Features extracted from two convolution blocks are used as the input of the classifier after a Global Average Pooling. The classifier consists of two fully connected layers, whose unit number is 128 and 2, respectively.

4) EEGWAVENET

EEGWaveNet [22] is an end-to-end model consisting of a multiscale convolution module, a spatial-temporal feature extractor, and a classifier. The multiscale convolution module uses 6-layer depthwise convolution (kernel size=2, strides=2), and the length of the signal will become 1/2 of the original one after each layer. The feature maps generated by depthwise convolution from layer 2 to layer 6 are used as the output of the multi-scale convolution module. The space-time domain feature extractor repeatedly stacks the convolutional layer (kernel size=4, strides=2) and batch normalization three times, and then performs global average pooling. The space-time domain feature extractor outputs a 32-dimensional feature vector at each scale and concatenates the 5-scale feature vectors to 160 - dimensional feature vectors as classifier input. The classifier is a three-layer fully connected layer, and the number of units is 64, 32, and 2, respectively.

D. EVALUATION METRICS

In this paper, Accuracy, Sensitivity, Specificity, and F1-score were used to evaluate the classification performance of the model. Accuracy evaluation model reflects the resolution of the model in the overall data. The calculation method is shown as follows.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + FP + \text{TN} + FN} \qquad (1)$$

TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative.

Sensitivity was calculated through Eq. (2) to indicate the accuracy of a model for a positive sample, while Specificity was calculated through Eq. (3) to indicate the accuracy of a model for a negative sample.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + FN} \qquad (2)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + FP} \qquad (3)$$

F1-Score is the harmonic average of Precision and Sensitivity. For imbalanced data sets, F1-Score can more accurately reflect the classification performance of the model. Its calculation method is shown in Eq. (4).

$$\text{F1Score} = \frac{2TP}{2TP + FP + FN} \qquad (4)$$

Many seizure detection studies used accuracy, sensitivity, and specificity as evaluation metrics for performance comparisons [16], [23], [36], [39]. However, accuracy and specificity may have a bias due to the class imbalance. For example, about 98.4% of the CHB-MIT dataset are negative samples. If a classifier predicts all testing instances as negative class will achieve high accuracy of 98.4% and perfect specificity of 100%. Therefore, the F1-score measure, which is a harmonic mean of recall (i.e., sensitivity) and precision, is usually more helpful than accuracy, especially for data with an uneven class distribution.

IV. RESULT
A. RESULTS OF SUBJECT-DEPENDENT EXPERIMENTS

Table 3 shows the average classification performance of different algorithms in the subject-dependence experiment. Compared with the previous end-to-end model, the proposed method has the best performance in all performance metrics, and the number of model parameters is only more than EEGNet-8,2 [38]. FBCSP+SVM [34] has the highest accuracy and specificity, but its F1-score is the lowest comparing with other methods. The sensitivity achieved by FBCSP+SVM even clearly underperformed (about 10%) other methods, revealing that seizure cases cannot be effectively detected. This confirms that accuracy and specificity are not proper metrics on class imbalance data like the CHB-MIT dataset. The model structure of EEGNet-8,2 is the most concise and has the least number of parameters, the overly simple model structure also causes the classification performance to be slightly lower than other models.

TABLE 3. Comparison of subject-dependent approach performance in the CHB-MIT dataset.

Published Work	Classification Algorithm	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-score (%)	Parameters
Keng et al. (2008) [34]	FBCSP+SVM	**98.83**	60.30	**99.51**	63.57	N/A
Lawhern et al. (2018) [38]	EEGNet-8,2	98.20	69.4	98.63	65.63	**2452**
Wang et al. (2021) [21]	Stacked 1D-CNN	97.66	69.24	98.19	65.98	106306
Thuwajit et al. (2022) [22]	EEGWaveNet	97.99	72.64	98.41	66.63	69502
Our Method	Inception and Residual model	98.34	**73.08**	98.79	**69.34**	28590
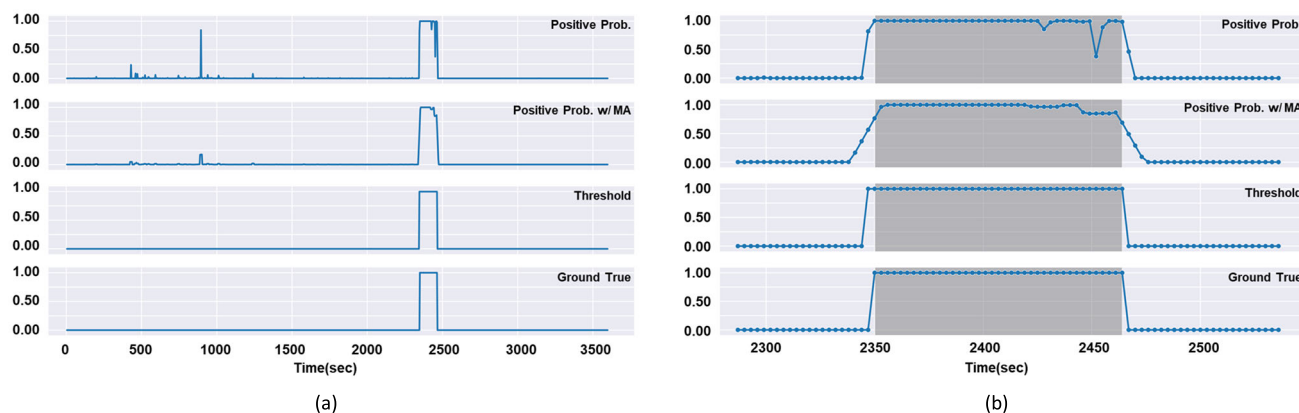
FIGURE 2. The chb05_22.edf file model detection result. (a) Result for all file. (b) Zoom in seizure location.

Fig. 2 shows the model prediction results of the record file chb05_22.edf, which are the seizure probability output by the model, the result after moving average, the result after thresholding, and the real label in sequence. Fig. 2(a) shows the results of the entire file. It can be seen that even if the model has false detections, it can be effectively suppressed by moving average and thresholding. Fig. 2(b) zooms in on the seizure event region, and the gray background is the seizure occurs. The detection errors for this seizure event were only one false positive at the onset of the seizure and one false negative during the seizure. The F1-score of chb04, chb06, chb08, chb13, chb14, chb16, chb23, and chb24 is low than 60%. The characteristics of each patient's data and the reasons for difficult training will be discussed in the following section.

B. RESULTS OF SUBJECT-INDEPENDENT EXPERIMENTS

We also conducted the subject-independent experiment for both CHB-MIT and Bonn datasets. Table 4 shows the average classification performance of different algorithms in subject-independent experiments. Compared with the previous end-to-end model, the method of this paper has the best Accuracy, Specificity, F1-score. Sensitivity is second only to Stacked 1D-CNN [21]. Stacked 1D-CNN has the best Sensitivity, its F1-score is lower than the proposed method, which means that Stacked 1D-CNN has more

false positives. FBCSP+SVM has the same problem as the subject-dependent experiment, and the Sensitivity is 10-20% less than other methods.

Table 5 shows the experimental results of different algorithms on the Bonn dataset. According to Table 5, it can be seen that the proposed method can still have excellent classification performance in datasets with a small amount of data, with an average accuracy of 99.04%. Compared with [21], [22], [38], [40], [41], and [42], the proposed method has more excellent performance. Compared with [17], [43], [44], and [45], the classification accuracy of the proposed method is a little bit worse, but [17], [43], and [45] did not report the accuracy of Case 8, and the difference between Class AB and Class CD cannot be clearly distinguished in its feature representation. Furthermore, those methods use hand-engineered feature representations that have potentially an ineffectiveness problem for classification [19].

The experiments and results on the CHB-MIT dataset experiment are closely similar to the related study conducted by Thuwajit et al. [22]. The lower F1-score was mainly caused by extremely imbalanced class distribution. The F1-scores of the previous studies [10], [19] were higher than 80% on the class-balanced data such as the Bonn dataset. We also analyzed the F1-scores in the experiments of the Bonn dataset. Our proposed method achieved a Marco-averaging

TABLE 4. Comparison of subject-independent approach performance in the CHB-MIT dataset.

Published Work	Classification Algorithm	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-score (%)	Parameters
Keng et al. (2008) [34]	FBCSP+SVM	**98.49**	29.76	**99.73**	32.45	N/A
Lawhern et al. (2018) [38]	EEGNet-8,2	95.07	40.14	96.05	31.35	**2452**
Wang et al. (2021) [21]	Stacked 1D-CNN	95.79	**49.00**	96.55	34.20	106306
Thuwajit et al. (2022) [22]	EEGWaveNet	95.98	43.65	96.87	34.41	69502
Our Method	Inception and Residual model	97.11	45.93	97.95	**37.31**	28590

TABLE 5. Comparison of epilepsy classification performance in the Bonn dataset.

Published Work	Classification Algorithm	Classification Accuracy (%)								
		Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Avg.
Lawhern et al. (2018) [38]	EEGNet-8,2	93.50	90.00	93.00	90.50	96.67	93.33	96.20	94.40	93.45
Gupta et al. (2018) [40]	Discrete Cosine Transform Based Filterbank + Statistical modeling + SVM	94.85	99.00	97.50	94.85	96.35	96.92	97.79	-	96.75
Mamli et al. (2019) [41]	Fourier Synchro-Squeezed Transform + Gray-Level Co-occurrence Matrix + KNN	100	99.38	99.54	96.48	99.73	99.59	97.38	99.26	98.92
Zhao et al. (2020) [42]	1D CNN	99.52	99.11	98.02	97.63	99.38	98.03	98.76	96.97	98.43
Jiang et al. (2020) [44]	Symplectic Geometry Eigenvalues + SVM	100	99.33	99.33	100	100	99.28	99.97	99.80	99.71
Rashed-Al-Mahfuz et al. (2021) [17]	Continuous Wavelet Transform + FT-VGG16	99.38	100	99.69	98.44	100	99.38	100	-	99.56
Yazid et al. (2021) [45]	3 level DWT + Local Binary Pattern Transition Histogram+ LBPMAD + KNN	99.94	99.86	99.88	99.70	99.92	99.72	99.74	-	99.82
Wang et al. (2021) [21]	Stacked 1D-CNN	99.00	99.50	99.00	97.00	99.33	96.67	99.00	96.00	98.19
Thuwajit et al. (2022) [22]	EEGWaveNet	100	99.00	98.00	96.00	99.67	98.67	98.00	97.80	98.39
Yan (2022) et al. [43]	MS-WTC + CNN	100	100	99.80	99.80	-	-	99.90	-	99.90
Our Method	Inception and Residual model	99.55	99.50	99.00	98.75	99.47	98.67	99.20	98.14	99.04

F1 score of 98.85%. In practice, intel-ictal samples and seizure samples are not balanced in real applications, so the results of the CHB-MIT dataset may naturally reflect the real-life seizure detection challenges.

V. DISCUSSION

In this section, we discuss the causes of false positives and missed detections. It can be mainly divided into noise, extreme class imbalance, and heterogeneity.

A. NOISE

We found that the subjects chb04, chb08, chb13, chb23, and chb24 in the CHB-MIT dataset had long-duration, huge amplitude, and multi-band noises. The noise appears in multiple channels at the same time, and it is difficult to filter it out through pre-processing because the noise frequency of each subject is different. Such noise can cause the model to generate a large number of false positives. The results of the short-time Fourier transform (STFT) and model output of the normal record file (chb04_28.edf) shown in Fig. 3 (a), which are the result of STFT, seizure probability output by the model, the result after moving average, and the real label in sequence. The abnormal record file (chb04_05.edf) is shown in Fig. 3 (b). In Fig. 3 (b), high-energy noise appears at about 3.5Hz, 8Hz, 11Hz, 15Hz...etc. A large number of false positives appear in the model output due to high energy noise at the same position.
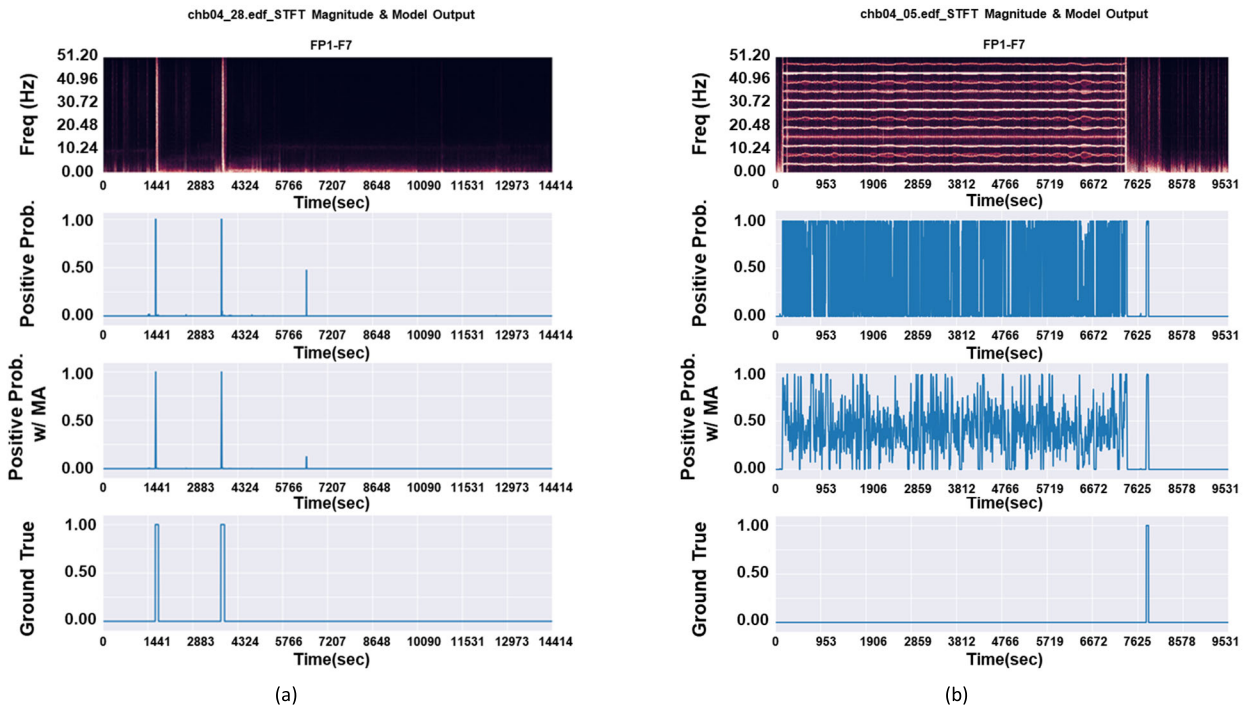
(a)   (b)

FIGURE 3. Short-time Fourier transform result, and the proposed model output. a) result of chb04_28.edf file. b) result of chb04_05.edf file.

B. EXTREME CLASS IMBALANCE

In the experiment of this paper, CHB-MIT is an imbalanced dataset, and the class ratio is about 1:63 (Seizure: Normal). In a few subjects such as chb04, chb06, chb14, etc., the number of normal samples was more than 100 times the number of seizure samples, which caused the model to abandon seizure samples during training. The chb16 only has 24 seizure samples, which made the model unable to learn more general features.

C. HETEROGENEITY

EEG classification of epilepsy has a high degree of heterogeneity within the same category. The EEG of the same subject will vary depending on the subject's current physical state when they are not having a seizure, and depending on the symptoms of epilepsy when the seizure occurs. The same class of EEG from different subjects varied depending on their physiological characteristics. Fig. 4 (a) and (b) show the proposed model dimension reduction distribution of the chb11 training data feature vector in the subject-dependence experiment by the t-SNE algorithm [46]. Fig. 4 (c) and (d) show the proposed model, dimension reduction distribution of the feature vectors chb10, chb18, and chb24 in the training data subject in the subject-independent experiment (chb01 as the test data) using the t-SNE algorithm. Fig. 4 (a) and (c) are labeled according to data categories, and Fig. 4 (b) and (d) are labeled according to data sources. We can find that although the feature vectors extracted by the model can clearly distinguish the states of seizures and non-seizures,
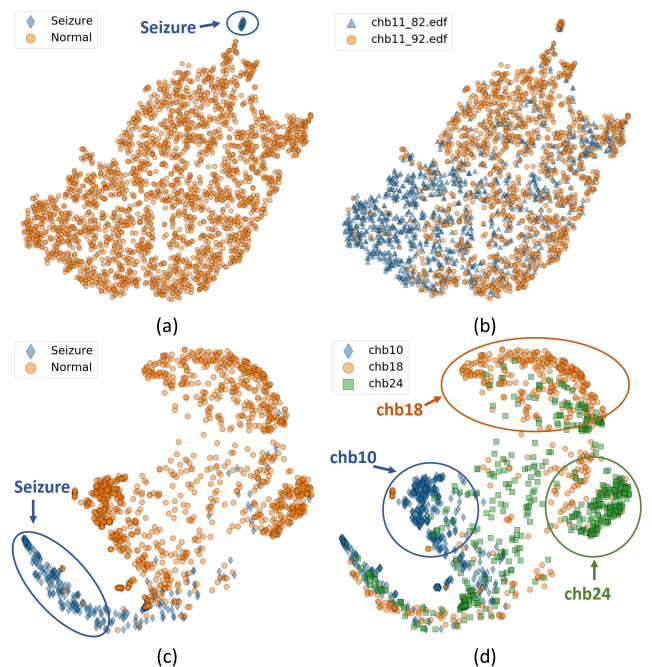

(a)   (b)

(c)   (d)

FIGURE 4. t-SNE result for proposed model feature vector.

they can also easily identify the source of the data. Fig. 4 (b) shows the samples from chb11_82.edf file is mainly distributed on the left side of the figure, and the samples from chb11_92.edf are mainly distributed on the right side of the figure. Fig. 4 (d) shows that the samples from chb18 are

mainly distributed on the upper side of the figure, the samples from chb10 are mainly distributed on the left side of the figure, and the samples from chb24 are mainly distributed on the right side of the figure. According to the results in Fig. 4, the model cannot eliminate the differences caused by different seizure events or different subjects in the process of generating feature vectors, to obtain more normalized EEG features.

VI. CONCLUSION

We propose a novel end-to-end deep learning model based on the inception and residual modules for seizure detection. We do not need domain expert knowledge to design the feature engineering with an end-to-end architecture. Compared to the previous models, our method integrates different scale information through the inception module for feature extraction. In addition, the residual module maintains the gradient process effectively during the model training phase. We conducted subject-dependent and subject-independent experiments on CHB-MIT and Bonn datasets. The previous methods, including FBCSP+SVM, EEGNet-8.2, Stacked 1D-CNN, and EEGWaveNet, were used to compare model performance. The experimental results show that our proposed model has the best classification performance in the CHB-MIT dataset. The macro-averaging f1-scores of the subject-dependent and subject-independent experiments are 69.34% and 37.31%. Our method has a macro-averaging accuracy of 99.04% in the Bonn dataset. Compared with the previous seizure detection methods needing a feature engineering process, our approach uses the original EEG signal as the model input, resulting in fewer parameters and lower computational complexity. The advantages of our proposed method are summarized as follows:

- Use EEG signals directly to avoid the highly complex feature extraction and the ineffectiveness risk of hand-engineered features.
- Our proposed method requires fewer parameters and lower computing resource requirements.
- Integrate different scale information in feature extraction through the inception module to achieve better classification performance.

As error cases we have analyzed and discussed, our model is limited to three aspects: noises, class imbalance, and heterogeneity that cause performance degradation. Therefore, future work is investigated in several directions. First, the denoising autoencoder [47] may help reduce noise. Secondly, for the class imbalance problem, deep metric learning [23] would be an attempt to transform the classification problem into a similarity problem during model training. Finally, domain-adversarial learning [48] that uses domain classifiers against a feature extractor during training may benefit from the heterogeneity limitation.

REFERENCES

[1] *Epilepsy: A Public Health Imperative: Summary*, World Health Organization, Geneva, Switzerland, 2019.

[2] K. Akyol, "Stacking ensemble based deep neural networks modeling for effective epileptic seizure detection," *Expert Syst. Appl.*, vol. 148, Jun. 2020, Art. no. 113239, doi: 10.1016/j.eswa.2020.113239.

[3] J. A. French, "Refractory epilepsy: Clinical overview," *Epilepsia*, vol. 48, no. s1, pp. 3–7, Mar. 2007.

[4] Y. Kumar, M. L. Dewal, and R. S. Anand, "Epileptic seizures detection in EEG using DWT-based ApEn and artificial neural network," *Signal, Image Video Process.*, vol. 8, no. 7, pp. 1323–1334, Oct. 2014.

[5] T. Zhang and W. Chen, "LMD based features for the automatic seizure detection of EEG signals using SVM," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 8, pp. 1100–1108, Aug. 2017, doi: 10.1109/TNSRE.2016.2611601.

[6] S. Chen, Zhang, Xi, L. Chen, and Z. Yang, "Automatic diagnosis of epileptic seizure in electroencephalography signals using nonlinear dynamics features," *IEEE Access*, vol. 7, pp. 61046–61056, 2019, doi: 10.1109/access.2019.2915610.

[7] J. Na, Z. Wang, S. Lv, and Z. Xu, "An extended k nearest neighbors-based classifier for epilepsy diagnosis," *IEEE Access*, vol. 9, pp. 73910–73923, 2021, doi: 10.1109/ACCESS.2021.3081767.

[8] M. R. Khan, W. Saadeh, and M. A. B. Altaf, "A low complexity patient-specific threshold based accelerator for the grand-mal seizure disorder," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Oct. 2017, pp. 1–4.

[9] D. Sopic, A. Aminifar, and D. Atienza, "E-glass: A wearable system for real-time detection of epileptic seizures," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2018, pp. 1–5.

[10] I. L. Olokodana, S. P. Mohanty, E. Kougianos, and R. S. Sherratt, "EZcap: A novel wearable for real-time automated seizure detection from EEG signals," *IEEE Trans. Consum. Electron.*, vol. 67, no. 2, pp. 166–175, May 2021, doi: 10.1109/tce.2021.3079399.

[11] A. K. Tiwari, R. B. Pachori, V. Kanhangad, and B. K. Panigrahi, "Automated diagnosis of epilepsy using key-point-based local binary pattern of EEG signals," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 4, pp. 888–896, Jul. 2017, doi: 10.1109/JBHI.2016.2589971.

[12] N. Lopac, F. Hržic, I. P. Vuksanovic, and J. Lerga, "Detection of non-stationary GW signals in high noise from Cohen's class of time–frequency representations using deep learning," *IEEE Access*, vol. 10, pp. 2408–2428, 2022, doi: 10.1109/ACCESS.2021.3139850.

[13] T. Arias-Vergara, P. Klumpp, J. C. Vasquez-Correa, E. Nöth, J. R. Orozco-Arroyave, and M. Schuster, "Multi-channel spectrograms for speech processing applications using deep learning methods," *Pattern Anal. Appl.*, vol. 24, no. 2, pp. 423–431, May 2021, doi: 10.1007/s10044-020-00921-5.

[14] S. Sakib, M. M. Fouda, Z. Md. Fadlullah, N. Nasser, and W. Alasmary, "A proof-of-concept of ultra-edge smart IoT sensor: A continuous and lightweight arrhythmia monitoring approach," *IEEE Access*, vol. 9, pp. 26093–26106, 2021, doi: 10.1109/ACCESS.2021.3056509.

[15] W. Liang, H. Pei, Q. Cai, and Y. Wang, "Scalp EEG epileptogenic zone recognition and localization based on long-term recurrent convolutional network," *Neurocomputing*, vol. 396, pp. 569–576, Jul. 2020.

[16] Y. Zhang, S. Yao, R. Yang, X. Liu, W. Qiu, L. Han, W. Zhou, and W. Shang, "Epileptic seizure detection based on bidirectional gated recurrent unit network," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 135–145, 2022, doi: 10.1109/TNSRE.2022.3143540.

[17] M. Rashed-Al-Mahfuz, M. A. Moni, S. Uddin, S. A. Alyami, M. A. Summers, and V. Eapen, "A deep convolutional neural network method to detect seizures and characteristic frequencies using epileptic electroencephalogram (EEG) data," *IEEE J. Transl. Eng. Health Med.*, vol. 9, 2021, Art. no. 2000112, doi: 10.1109/JTEHM.2021.3050925.

[18] D. Lai, X. Zhang, K. Ma, Z. Chen, W. Chen, H. Zhang, H. Yuan, and L. Ding, "Automated detection of high frequency oscillations in intracranial EEG using the combination of short-time energy and convolutional neural networks," *IEEE Access*, vol. 7, pp. 82501–82511, 2019, doi: 10.1109/ACCESS.2019.2923281.

[19] Y. Li, Y. Liu, W. Cui, Y. Guo, H. Huang, and Z. 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, doi: 10.1109/TNSRE.2020.2973434.

[20] 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, Sep. 2018, doi: 10.1016/j.compbiomed.2017.09.017.

[21] X. Wang, X. Wang, W. Liu, Z. Chang, T. Kärkkäinen, and F. Cong, "One dimensional convolutional neural networks for seizure onset detection using long-term scalp and intracranial EEG," *Neurocomputing*, vol. 459, pp. 212–222, Oct. 2021.

[22] P. Thuwajit, P. Rangpong, P. Sawangjai, P. Autthasan, R. Chaisaen, N. Banluesombatkul, P. Boonchit, N. Tatsaringkansakul, T. Sudhawiyangkul, and T. Wilaiprasitporn, "EEGWaveNet: Multiscale CNN-based spatiotemporal feature extraction for EEG seizure detection," *IEEE Trans. Ind. Informat.*, vol. 18, no. 8, pp. 5547–5557, Aug. 2022, doi: 10.1109/tii.2021.3133307.

[23] L. Duan, Z. Wang, Y. Qiao, Y. Wang, Z. Huang, and B. Zhang, "An automatic method for epileptic seizure detection based on deep metric learning," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 5, pp. 2147–2157, May 2022, doi: 10.1109/JBHI.2021.3138852.

[24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[26] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[27] H. You, L. Yu, S. Tian, X. Ma, Y. Xing, N. Xin, and W. Cai, "MC-Net: Multiple max-pooling integration module and cross multi-scale deconvolution network," *Knowl.-Based Syst.*, vol. 231, Nov. 2021, Art. no. 107456, doi: 10.1016/j.knosys.2021.107456.

[28] X. Jiang, J. Zhao, D. Bo, A. Panfeng, H. Guo, and Z. Yuan, "MRNet: A multi-scale residual network for EEG-based sleep staging," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2021, pp. 1–8, doi: 10.1109/IJCNN52387.2021.9534133.

[29] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," 2016, *arXiv:1608.03983*.

[30] A. H. Shoeb and J. V. Guttag, "Application of machine learning to epileptic seizure detection," in *Proc. ICML*, 2010, pp. 975–982.

[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, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 64, no. 6, Nov. 2001, Art. no. 061907, doi: 10.1103/PhysRevE.64.061907.

[32] Z. Y. Chin, K. K. Ang, and C. Guan, "Multiclass voluntary facial expression classification based on filter bank common spatial pattern," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2008, pp. 1005–1008, doi: 10.1109/IEMBS.2008.4649325.

[33] A. D. Bigirimana, N. Siddique, and D. Coyle, "Emotion-inducing imagery versus motor imagery for a brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 4, pp. 850–859, Apr. 2020, doi: 10.1109/TNSRE.2020.2978951.

[34] 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.)*, Jun. 2008, pp. 2390–2397, doi: 10.1109/IJCNN.2008.4634130.

[35] S. Park, D. Lee, and S. Lee, "Filter bank regularized common spatial pattern ensemble for small sample motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 498–505, Feb. 2018, doi: 10.1109/TNSRE.2017.2757519.

[36] C. Li, W. Zhou, G. Liu, Y. Zhang, M. Geng, Z. Liu, S. Wang, and W. Shang, "Seizure onset detection using empirical mode decomposition and common spatial pattern," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 458–467, 2021, doi: 10.1109/TNSRE.2021.3055276.

[37] T. N. Alotaiby, F. E. Abd El-Samie, S. A. Alshebeili, K. H. Aljibreen, and E. Alkhanen, "Seizure detection with common spatial pattern and support vector machines," in *Proc. Int. Conf. Inf. Commun. Technol. Res. (ICTRC)*, May 2015, pp. 152–155, doi: 10.1109/ICTRC.2015.7156444.

[38] 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, Oct. 2018, Art. no. 056013.

[39] X. Tian, Z. Deng, W. Ying, K. Choi, D. Wu, B. Qin, J. Wang, H. Shen, and S. Wang, "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, doi: 10.1109/TNSRE.2019.2940485.

[40] A. Gupta, P. Singh, and M. Karlekar, "A novel signal modeling approach for classification of seizure and seizure-free EEG signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 5, pp. 925–935, May 2018, doi: 10.1109/TNSRE.2018.2818123.

[41] S. Mamli and H. Kalbkhani, "Gray-level co-occurrence matrix of Fourier synchro-squeezed transform for epileptic seizure detection," *Biocybern. Biomed. Eng.*, vol. 39, no. 1, pp. 87–99, Jan. 2019, doi: 10.1016/j.bbe.2018.10.006.

[42] W. Zhao, W. Zhao, W. Wang, X. Jiang, X. Zhang, Y. Peng, B. Zhang, and G. Zhang, "A novel deep neural network for robust detection of seizures using EEG signals," *Comput. Math. Methods Med.*, vol. 2020, pp. 1–9, Apr. 2020, doi: 10.1155/2020/9689821.

[43] X. Yan, D. Yang, Z. Lin, and B. Vucetic, "Significant low-dimensional spectral-temporal features for seizure detection," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 668–677, 2022, doi: 10.1109/TNSRE.2022.3156931.

[44] Y. Jiang, W. Chen, and M. Li, "Symplectic geometry decomposition-based features for automatic epileptic seizure detection," *Comput. Biol. Med.*, vol. 116, Jan. 2020, Art. no. 103549, doi: 10.1016/j.compbiomed.2019.103549.

[45] M. Yazid, F. Fahmi, E. Sutanto, W. Shalannanda, R. Shoalihin, G. Horng, and Aripriharta, "Simple detection of epilepsy from EEG signal using local binary pattern transition histogram," *IEEE Access*, vol. 9, pp. 150252–150267, 2021, doi: 10.1109/access.2021.3126065.

[46] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 1–27, 2008.

[47] Y. Qiu, W. Zhou, N. Yu, and P. Du, "Denoising sparse autoencoder-based ictal EEG classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 9, pp. 1717–1726, Sep. 2018, doi: 10.1109/TNSRE.2018.2864306.

[48] Y. Ganin, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2030–2096, 2017.

KUO-KAI SHYU (Senior Member, IEEE) received the B.S. degree in electrical engineering from the Tatung Institute of Technology, Taipei, Taiwan, in 1979, and the M.S. and Ph.D. degrees in electrical engineering from National Chung Kung University, Tainan, in 1984 and 1987, respectively.

In 1988, he joined National Central University, Jhongli, where he is currently a Professor with the Department of Electrical Engineering. He is currently a Professor with the Pervasive Artificial Intelligence Research (PAIR) Laboratories, Taiwan. He was the Chairperson of the Department, from 2004 to 2007. From 1988 to 1999, he was a Visiting Scholar with the Electrical and Computer Engineering Department, Auburn University, Auburn, AL, USA. His teaching and research interests include variable structure control systems and signal processing with applications in motor control, power electronics, and biomedical systems.

SZU-CHI HUANG was born in Taichung, Taiwan, in 1999. He received the B.S. degree in electronic engineering from the National Taipei University of Technology, Taipei, Taiwan, in 2021. He is currently pursuing the M.S. degree in electrical engineering with National Central University, Taoyuan, Taiwan. His research interests include signal processing and machine learning with applications in biomedical systems.

LUNG-HAO LEE (Member, IEEE) received the B.S. degree in statistics from National Taipei University, Taiwan, in 2003, the M.S. degree in information management from Yuan Ze University, Taoyuan, Taiwan, in 2005, and the Ph.D. degree in computer science and information engineering from National Taiwan University, in 2015. From 2015 to 2018, he was a Postdoctoral Fellow with National Taiwan Normal University. He joined as an Assistant Professor with the Department of Electrical Engineering, National Central University, Taiwan, in 2018. He is currently an Associate Professor. His research interests include natural language processing, information retrieval and extraction, biomedical and health informatics, and artificial intelligence technologies.

PO-LEI LEE (Member, IEEE) was born in 1973. He received the B.S. degree in electrical engineering from National Cheng Kung University, Taiwan, in 1995, and the Ph.D. degree from the Institute of Biomedical Engineering, National Yang Ming Chiao Tung University, Taiwan, in 2000.

From 2001 to 2005, he was a Postdoctoral Fellow with the Taipei Veterans General Hospital, researching the signal and image-analysis procedures for electroencephalography and magnetoencephalography signals. He joined the Department of Electrical Engineering, National Central University, Taiwan, in 2005. He is currently a Professor with the Pervasive Artificial Intelligence Research (PAIR) Laboratories, Taiwan. His research interests include the signal and image processing of EEG and MEG signals and designing the EEG-based brain–computer interfaces.

● ● ●