Epileptic Seizure Detection Using Brain-Rhythmic Recurrence Biomarkers and ONASNet-Based Transfer Learning

Zhenxi Song, Bin Deng, *Senior Member, IEEE*, Jiang Wang, *Member, IEEE*, Guosheng Yi, and Wei Yue

***Abstract*— *Objective:* The electroencephalogram (EEG) tool has great potential for real-time monitoring of abnormal brain activities, such as preictal and ictal seizures. Developing an EEG-based detection system for patients with epilepsy is vital for clinical management and targeted therapy. *Methods:* This paper proposes a single-channel seizure detection system using brain-rhythmic recurrence biomarkers (BRRM) and an optimized model (ONASNet). BRRM is a direct mapping of the recurrence morphology of brain rhythms in phase space; it reflects the nonlinear dynamics of original EEG signals. The architecture of ONASNet is determined through a modified neural network searching strategy. Then, we exploited transfer learning to apply ONASNet to our EEG data. The combination of BRRM and ONASNet leverages the multiple channels of a neural network to extract features from different brain rhythms simultaneously. *Results:* We evaluated the efficiency of BRRM-ONASNet on the real EEG recordings derived from Bonn University. In the experiments, different transfer-learning models (TLMs) are respectively constructed using ONASNet and seven well-known neural network structures (VGG16/VGG19/ResNet50/InceptionV3/DenseNet121/Xception/NASNet). Moreover, we compared those TLMs by model size, computing complexity, learning capability, and prediction latency. ONASNet outperforms other structures by strong learning capability, high stability, small model size, short latency, and less requirement of computing resources. Comparing BRRM-ONASNet with other existing methods, our work performs better than others with 100% accuracy under the identical dataset and same detection task. *Contributions:* The proposed method in this study, analyzing nonlinear features from phase-space representations using a deep neural network, provides new insights for EEG decoding. The successful application of this method in epileptic-seizure detection contributes to computationally medical assistance for epilepsy.**

***Index Terms*— Seizure detection, EEG biomarker, nonlinear dynamics, neural network, transfer learning.**

Manuscript received November 16, 2021; revised March 6, 2022; accepted April 2, 2022. Date of publication April 5, 2022; date of current version April 19, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62071324 and in part by the Natural Science Foundation of Tianjin under Grant 19JCQNJC01200 and Grant 19JCZDJC36500. *(Corresponding authors: Guosheng Yi; Wei Yue.)*

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

Zhenxi Song, Bin Deng, Jiang Wang, and Guosheng Yi are with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: songzhenxi@tju.edu.cn; dengbin@tju.edu.cn; jiangwang@tju.edu.cn; guoshengyi@tju.edu.cn).

Wei Yue is with the Tianjin Key Laboratory of Cerebral Vascular and Neurodegenerative Diseases, Department of Neurology, Clinical College of Neurology, Neurosurgery, and Neurorehabilitation, Tianjin Medical University, Tianjin Huanhu Hospital, Tianjin 300000, China (e-mail: hhyuewei2008@163.com).

Digital Object Identifier 10.1109/TNSRE.2022.3165060

I. Introduction

EPILEPSY is an acute, recurrent paroxysmal brain dysfunction caused by the over-discharge of brain neurons. During an epileptic seizure, the patient may risk losing consciousness, sustaining an injury, or even death. According to the World Health Organization statistics, more than 60 million people have epilepsy worldwide [1]. As an effective tool for monitoring brain activities at the cerebral cortex level, EEG has the advantages of noninvasiveness, low cost, and high temporal resolution. Due to those advantages, EEG has excellent potential for diagnosing various neurological diseases, such as epilepsy [2], [3]. Manually visual inspection of epileptiform abnormalities in EEG signals is time-consuming and may miss meaningful features behind the observed time series. Developing automatic EEG monitoring systems for epileptic patients is significant for accurate diagnosis of epilepsy, early warning of seizures, and targeted assistance in clinical treatments [4]–[6], such as deploying invasive wearable devices [7].

The EEG-based automatic epileptic-seizure detection strategies have been discussed for over one decade but remain challenging [8]. Automatic seizure detection mainly involves two steps: feature extraction and classification. These features can be obtained by manual quantification or deep learning model coding [8]–[25]. Thus, we boil down those methods into two types: the handcrafted-feature-based traditional machine learning and the deep-learning-based methods.

The handcrafted feature-based machine learning methods require manual feature extraction using the algorithms, such as temporal fluctuation measurements, wavelet transforms, power spectral density analysis, time-frequency distribution analysis, and entropy extractor [8], [17]–[20], [24], [25]. The commonly used classifiers include support vector machines [19]–[21], [24], k-nearest neighbors [17], [21], and the classical artificial neural networks [10], [16], [21]. In addition, some

studies would exploit multiple feature extraction algorithms to ensure the features are sufficient for characterizing EEG. Thus, various feature selection strategies are required for refining the obtained features before feeding them into the classifiers [11], [21].

Deep learning-based methods are designed to automatically learn meaningful features from EEG [26]. Convolutional neural networks (CNNs), as one of the most prominent architectures in deep neural networks, have achieved promising results in seizure detection. Acharya *et al*. [12] implemented a 13-layer deep one-dimensional CNN (1D-CNN) to detect normal, preictal, and seizure patterns. In their study, three hundred single-channel EEG signals were fed into the CNN and finally achieved an overall accuracy of 85.1%. Zhou *et al.* [13] proposed a two-dimensional CNN (2D-CNN) model for the same task on the Freiburg database. By taking the power-frequency graph and the amplitude-time graph as the input, they ended up with an accuracy of 92.3%. Wei *et al*. [14] proposed a three-dimensional CNN (3D-CNN) to identify different stages of seizures. The 3D-CNN analyzes features of multi-channel EEG signals from the dimensions of time, amplitude, and channels simultaneously. They finally reached an accuracy of 90%. In addition, the recurrent neural network (RNN), as a type of model dedicated to processing time series, is also studied and used for detecting epileptiform discharge [23]. A recent review [27] indicates that CNNs are the most widely used method for seizure detection (2D-CNN, 31.48%; 1D-CNN, 22.22%). In this paper, we aim to leverage the capability of CNN to capture invisible features in different EEG patterns.

According to our literature survey, most deep learning-based methods exploit the neural network models to analyze EEG signals in time and frequency domains [27]. Nonlinear analyses are commonly used in handcrafted feature-based machine learning methods. Precisely, the EEG dynamics are quantified as several values to indicate the chaotic and rhythmic patterns concealing in the nonlinear and non-stationary EEG signals [28]. The quantification algorithms include recurrence qualitative analysis (RQA), detrended fluctuation analysis, entropies, fractal dimensions, Hjortj, Hurst exponent, Kolmogorov complexity, largest Lyapunov exponent, and Lempel-Ziv complexity [28]. Of note, entropies have been extensively used in seizure detection [15]; however, just like all the varieties of entropies finally generated the entropy values based on the probability distribution. Certain important nonlinear features would be neglected or lost during the manual extraction of highly refined indicators. To avoid this problem, it motivated us to combine the deep neural network with low-hierarchy nonlinear features to reduce the information loss and characterize the seizure-related nonlinear features more effectively.

The nonlinear features mentioned above are determined by EEG signals' periodicity, chaos, and non-stationary nature. Inspired by the fact that those characteristics are related to the internal structure of time series, we assume that the morphology of EEG trajectory can partly reflect the nonlinear dynamics. For example, RQA [29] was proposed based on a recurrence plot inferred from the EEG trajectory's morphology. However, RQA requires manual quantification to extract features from recurrence plots; the subjective steps in RQA, such as thresholding recurrence plots, would lead to bias and information loss. Meanwhile, the clinical analysis discovered that the rate of interictal spikes would dramatically change at the preictal period of seizure [30], and the brain rhythms drastically altered at the onset of seizures [31]. Therefore, we started this work with the hypothesis that taking the morphology of multi-rhythmic trajectory as the input of CNN models can effectively encode the seizure-free and seizure patterns and distinguish them from healthy patterns.

Moreover, an encoding model with small latency is promising for a real-time application and contributes to the early warning of epileptic seizures. Thus, this paper has investigated various network structures and developed a lightweight structure. The main contributions of this study are summarized as follows: (1) A nonlinear biomarker, brain-rhythmic recurrence map, is proposed to reveal the nonlinear dynamics of EEG signals; (2) An effective model named ONASNet is developed based on the optimized neural network searching (NAS) strategy and the lightweight CNN cells; (3) The effectiveness of biomarkers is validated on various models with different structures, and the performance of their difference is systematically studied in model size, learning capability, and predicted accuracy; (4) The proposed method accurately identifies normal, seizure-free, and seizure patterns with high stability and low prediction latency.

II. Materials

The datasets used in this study are available in the "Klinik fűr Epileptologie, Universität Bonn" repository. The Bonn epileptic EEG dataset contains a total of 500 EEG trials. Each EEG trial is a single-channel EEG signal with a sampling rate of 173.61 Hz and a duration of 23.6 seconds (4097 sample points). These trials have been pre-selected by the data collector by removing the artifacts which result from muscle activity and eye movement. This study defined three categories for further investigation: a normal group, a seizure-free group, and a seizure group. Therefore, the seizure-free and seizure groups are derived from the EEG trails recorded in the epileptogenic areas of the patients with epilepsy. The recordings collected in the interictal period are seizure-free, and those collected during seizures comprise the seizure group. The seizure-free and seizure groups contain 100 EEG trails, respectively. Furthermore, we selected 100 EEG trails from the recordings collected from the healthy subjects in their relaxed states.

According to clinical studies [32], [33], most seizures occur with frequencies ranging from 2Hz to 29Hz. Therefore, the EEG signals were filtered by a 4th-order zero-phase Butterworth bandpass filter with a filtering frequency between 2 and 30 Hz. Moreover, we identified each subject's brain pattern using multiple detection windows to enhance stability. Those detection windows respectively extract a short EEG sequence from each EEG signal without overlap and simultaneously infer the results. Then, the averaged logits of all the detection windows are used to predict the final results. In this study, each detection window extracts 173 sampling points. Hence,

twenty-three detection windows are used to identify a 23.6s EEG signal.

III. Methods

A. The Proposed Brain-Rhythmic Recurrence Map

To exploit the convolutional layers to capture the underlying features in the EEG signals from the perspective of nonlinear dynamics, we first propose a brain-rhythmic recurrence feature map improved from recurrence plots (RPs). Subsequently, we exploit the transfer learning technique to extract features from the brain-rhythmic recurrence maps.

1) The Mathematical Explanation of Recurrence Plots: The trajectory of EEG signals evolves in the high dimensional phase space. The RP method transforms a high-dimensional trajectory into a two-dimensional graph by calculating the recursive features appearing in the phase space to reduce dimensionality. In the following, we provide a brief description of RPs mathematically.

Given a time series $\{x_1, x_2, \ldots, x_n\}$, according to the Takens embedding theorem [34], it can be reconstructed by several state points in the phase space, and each point is defined as the Eq. (1),

$$X_i = \{x_i, x_{i+\tau}, \ldots, x_{i+(m-1)*\tau}\}, \quad i = 1, 2, 3, \ldots, N \tag{1}$$

where $N = n - (m-1)\tau$. The delay time τ and embedding dimension m were estimated using the mutual information method [35] and the false nearest neighbor method [36], [37]. Then, the distance (d_{ij}) between any two elements is defined by the Euclidean norm as Eq. (2). Subsequently, a recursive matrix can be obtained by Eq. (3) and Eq. (4),

$$d_{ij} = \left\| X_i - X_j \right\| \tag{2}$$

$$R_{ij} = H\left(\varepsilon - d_{ij}\right) \tag{3}$$

$$H(x) = \begin{cases} 0 & (x < 0) \\ 1 & (x > 1) \end{cases} \tag{4}$$

where ε is the threshold used for defining useful recurrence features; in general, the value of ε can be defined by the standard deviation of the signals or determined by setting a specific percentage percent of the points as 1. From Eq. (2) to Eq. (4), they delimit a sphere in the phase space by taking one state point X_i as the center and the threshold ε as the radius. Then, the recurrence is captured when other states X_j reach the inside space of the sphere. Fig. 1 illustrates the RPs of the EEG signals, which were randomly sampled from normal, seizure-free, and seizure groups. Benefitting from the high temporal resolution of EEG signals, we could divide each EEG trial into 23 one-second EEG segments. In Fig. 1, each RP corresponds to the exemplified EEG segment depicted in the black-framed sub-panel on the left. The state points on the trajectory in the phase space comprise the axes of each RP matrix. The state points denoted as black indicate that recurrence occurs at those state points corresponding to the horizontal and vertical axes of the coordinate.

In some instances, it is impractical to distinguish different brain patterns based on the RPs directly. Therefore, further quantifications are required to quantify the RP structures, such as RQA [38]. Then, those extracted indicators are used to train the classical machine learning classifiers. However, such a quantification procedure manually discards some natural features that include essential information. With deep learning techniques, CNNs can automatically extract the features from the grid data. Therefore, using a deep neural network model to analyze the RPs promises a greater probability of gaining better performance.

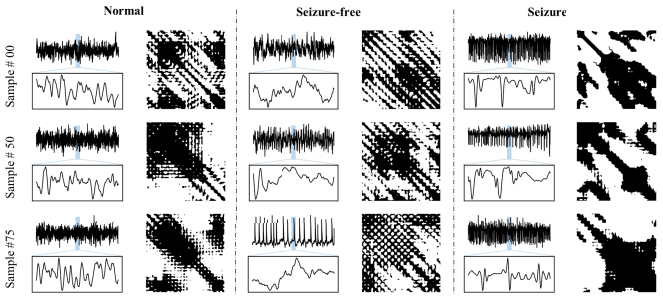

Fig. 1. Illustrations of the recurrence plots of the EEG signals sampled from normal, seizure-free, and seizure groups.

Fig. 2. Examples of the greyscale RPs obtained from different groups.

Moreover, as shown in Fig. 1, the recurrence points (black points) clustered together as several black chunks in seizure patterns. It indicates that the original RP method cannot provide detailed information when dynamic recurrence occurs in the continuous state points. In other words, the original RP method is not appropriate to characterize the recursive structure when the trajectory of EEG folds in the embedding space severely. Hence the corresponding RPs may lose the crucial dynamical information due to binarization.

2) Construction of Brain-Rhythmic Recurrence Feature Map: To address the issue mentioned above, we propose to directly apply Eq. (2) to the state points (Eq. (1)) of each high-dimensional trajectory. Hence a new RP (greyscale RP) with detailed recurrence features can be obtained, as shown in Fig. 2. The illustrations correspond to the exemplified samples' 15th EEG detection window results. The greyscale RP defines the recursive features as a range instead of binary numbers, where the brighter points imply significant recursive features. Thus, the method alleviates the bias and instability induced by the different choices of the parameter ε.

Furthermore, the synchronized electrical pulses of massive neurons would induce rhythmic brain activity, which can be captured at the cerebral cortex level. Hence the recorded EEG signals typically contain five brain rhythms named delta (δ), theta (θ), alpha (α), beta (β), and gamma (γ). Each rhythm corresponding to a specific frequency band frequently occurs under specific brain states (TABLE I). As we stated

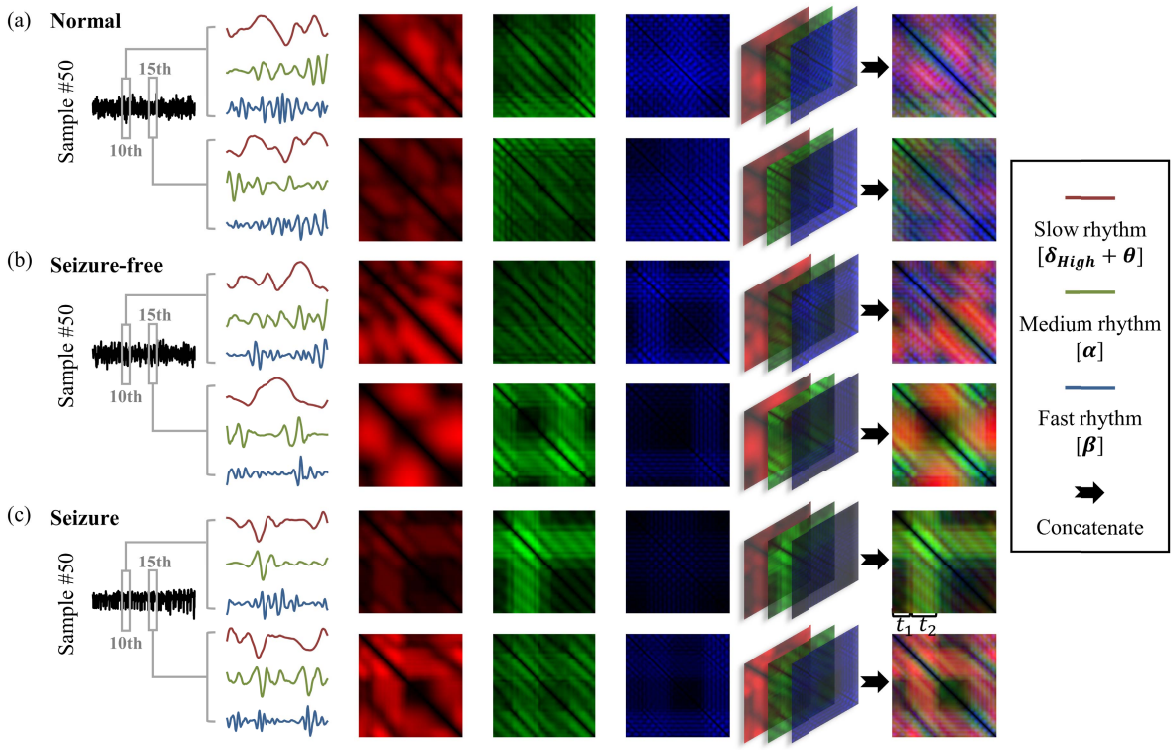

Fig. 3. Construction of the brain-rhythmic recurrence feature maps and the examples derived from different groups.

TABLE I
THE FREQUENCY RANGE OF DIFFERENT BRAIN RHYTHMS

Rhythm	Frequency Band	Occurrence Period
δ	2-4 Hz	Related to the state of deep sleep
θ	4-8 Hz	Related to the twilight zone between waking and sleep
α	8-12 Hz	Related to the state of relaxation (e.g., peacefully imagining)
β	12-30 Hz	Related to the primary state of alertness (e.g., intellectual activity)
γ	30-80 Hz	Related to consciousness and awareness

in Section II, the frequency bands lower than 2Hz and higher than 30Hz have been filtered out during the data preparation to focus on the frequency range most related to epileptic activities. Hence, the filtered signals involve δ, θ, α, β four brain rhythms. We extract the sub-bands from the EEG time series through the multi-level wavelet decomposition and obtain the frequency bands from the reconstructed wavelet coefficients.

Considering that the low brain rhythms (δ and θ) are rare in the awake state of adults [39], we integrate the lower bands with the frequency range from 2Hz to 8Hz as the slow brain rhythm, regard the filtered EEG with the 8-12Hz frequency range as the medium brain rhythm, and set the frequency band higher than 12Hz as the fast brain rhythm. Subsequently, by cooperating with the concept of RGB channels in color images, the brain-rhythmic recurrence map (BRRM) is constructed upon those brain rhythms. We obtained BRRM by concatenating the rhythmic-specific recurrence feature map in depth. It includes three channels corresponding to slow, medium, and fast brain rhythms. In Fig. 3, we illustrate the construction of BRRM and show the BRRMs of two EEG short sequences (the 10^{th} and 15^{th} sequences) split from the exemplified EEG signal (the 50^{th} sample) in normal, seizure-free, and seizure categories. The BRRMs derived from the same EEG signal, but different EEG sequences may have a similar geometric pattern but different color distribution. For example, in the 10^{th} and 15^{th} detection windows of the illustrated seizure sample (Fig. 3(c)), the recursive features are evident at t_1 period but become weak at t_2 period. Of note, the decreased recursive characteristics of the medium brain rhythm (green color) induce this weakened recurrence in the 10^{th} detection. In contrast, such decay in the 15^{th} window is caused by the slow rhythm (red color). However, the geometric patterns of the BRRMs obtained from the different sequences on the same EEG signals also may show differences (e.g., Fig. 3(b)) due to the evolution of the underlying nonlinear dynamics over time. Moreover, compared to the greyscale RPs (Fig. 2), the BRRMs of the normal and seizure-free groups (Fig. 3(a) and 3(b)) in the 15^{th} window are more distinguishable.

Based on the observations above, we propose two hypotheses:

a) the proposed BRRM provides a more precise and comprehensive representation of nonlinear features in EEG signals than other relevant methods, including RQA, original RP, and greyscale RP.

b) the BRRM is a potential biomarker that a deep learning model can process to generate pattern-specific deep representations for automatically identifying epileptic brain states (seizure-free and seizure patterns).

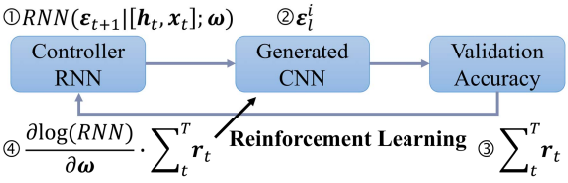

Fig. 4. Neural network searching via RNN and reinforcement learning.

B. Transfer-Learning Method and the Optimized Network

The proposed three-channel BRRM can fully use convolutional layers to process color images. It enables the network to extract and fuse valuable features from different brain rhythms simultaneously. From Fig. 3, we know the diversity of BRRMs between different subjects from the same group. Therefore, we utilized the transfer learning technique and cooperated with the proposed BRRM to extract the pattern-specific deep representations from different samples.

1) The Transfer-Learning Models: Our experiments evaluated eight architectures by taking them as the bone of the transfer-learning model (TLM). Seven of those bottom structures are derived from the acknowledged models, including VGG16 [40], VGG19 [40], ResNet50 [41], InceptionV3 [42], DenseNet121 [43], Xception [44], NASNet [45]. Meanwhile, we introduce another bottom structure named ONASNet, an optimized NASNet produced by a modified neural architecture search strategy and an effective convolutional block with inverted residual and linear bottleneck modules. TLMs are obtained by replacing the respective top layers with a simple structure comprising a global average pooling layer, a dense layer, and a SoftMax activation.

2) The Background Related to the Optimized Network: The neural network search method [46] was proposed to design the architecture of the network by self-learning. NASNet followed this idea to produce the model structure through RNN and reinforcement learning (Fig. 4). Compared with the handcrafted networks (e.g., VGG, ResNet, InceptionV3, Xception), to automatically generate the optimized hyper-parameters of architecture for obtaining the best validation accuracy under the restricted computing cost. As shown in Fig. 4, the controller RNN is a network having recurrent layers, where each layer defines a hyper-parameter for training. Indeed, to reduce the computing cost, we should pre-define the limited candidate CNN operators (such as the convolutional kernels and the filter number) to restrict the search space.

In the function ① of Fig. 4, $\boldsymbol{h}_t$ and $\boldsymbol{x}_t$ represent the state and the input of RNN at time t, respectively. $\boldsymbol{\omega}$ stands for the parameters of recurrent layers and fully convolutional layers. Those parameters determine the probability output of RNN at the next time point, then, the corresponding action of the CNN operator $\boldsymbol{\varepsilon}_{t+1}$ can be sampled from the probability distribution. Each recurrent step (time point) of the controller RNN is used to identify the optimized result of one CNN operator (e.g., kernel size). Detailly, each layer of trained RNN controls the ith operator at the lth layer in CNN (denoted as $\boldsymbol{\varepsilon}_l^i$). The optimization objective is to train the controller RNN for generating an optimized CNN with the highest validation accuracy. However, the objective (validation accuracy) is not a differentiable function of the optimization variable (parameters of RNN $\boldsymbol{\omega}$). Hence, reinforcement learning is required to solve this problem. The symbol ③ stands for the observed validation accuracy from step t to the last step T and is taken as the reward for reinforcement learning. With the state $[\boldsymbol{h}_t, \boldsymbol{x}_t]$ and the action $\boldsymbol{\varepsilon}_{t+1}$, the function ① can be considered as the policy of reinforcement learning. Then, the optimization of the non-differentiable problem can be alternatively solved by policy-gradient algorithms [47], as shown in the equation ④.

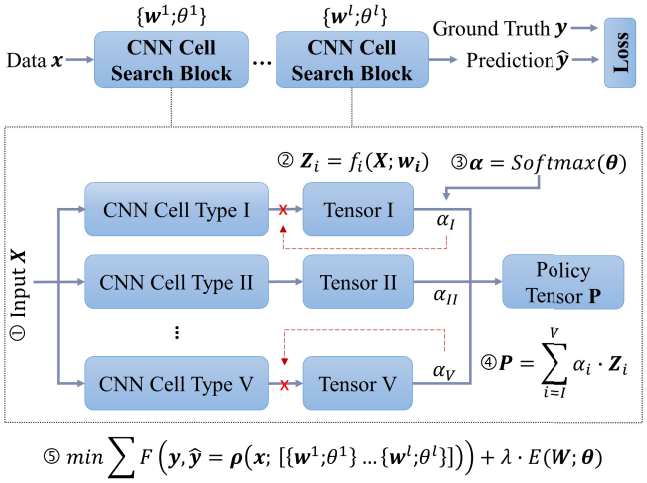

Fig. 5. Neural network searching via gradient descent.

3) The Optimized NASNet (ONASNet): Although the method described in Section III.B.2 could achieve the objective theoretically, training such RNN through reinforcement learning is challenging and computationally costly, induced by the non-differentiable problem. Therefore, making the network search problem computationally differentiable is promising for reducing computing costs. An integrated framework (Fig. 5) with the parallel CNN cells is introduced to obtain the optimal network through the gradient descent method. NASNet implements neural network research in a cascaded way, where the first step is searching for the optimal CNN cells, and the second step is generating the whole network. Fig. 5 integrates two levels in one framework, where a series of blocks construct the optimal network, and each block represents one stage of the network. Specifically, the block contains some parallel pathways, representing one type of CNN cell.

As shown in Fig. 5, for the input $\boldsymbol{X}$, we can attain a tensor $\boldsymbol{Z_i}$ from the ith CNN cells with the parameters $\boldsymbol{w_i}$; then, an output tensor representing the policy of the block is defined by the weighted average of the multiple tensors, as the equation ④ in Fig. 5. The weights $\boldsymbol{\alpha}$ are determined by the parameters $\boldsymbol{\theta}$ and processed by the SoftMax function to normalize those values. Consequently, for the input data $\boldsymbol{x}$ and the ground truth $\boldsymbol{y}$, the optimization objective of training can be defined as the function ⑤ in Fig. 5, where the error between ground truth $\boldsymbol{y}$ and the predicted results $\hat{\boldsymbol{y}}$ is measured through a function F (e.g., cross-entropy), the output $\boldsymbol{\rho}(\cdot)$ is a differential function of the parameters $\boldsymbol{\theta}$ and $\boldsymbol{w}$, the $E(\boldsymbol{W}; \boldsymbol{\theta})$ is the penalty term that is flexible to the requirement

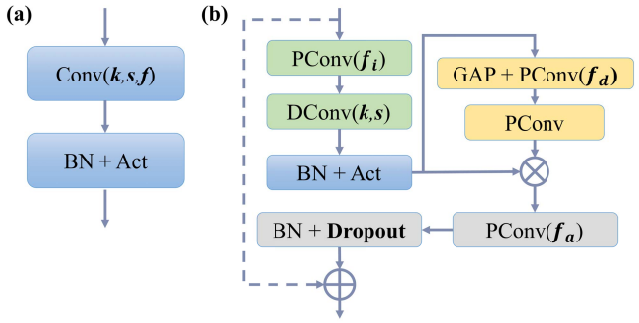

Fig. 6. The candidate CNN cells: (a) the basic type and (b) the effective type.

(e.g., using the elapsed time of each CNN cell for penalty). Finally, the back-propagation of the gradient makes the framework converge to the optimal point. Then, the trained $\boldsymbol{\alpha}$ controls the cutting of the branches in the CNN cell search block to remain only one branch in each block (e.g., the randomly sampling according to the probability distribution of $\boldsymbol{\alpha}$). Consequently, those selected CNN cells construct an optimal network.

For the candidate CNN cells, we introduce two structures (Fig. 6) applied at the first stage/block of the network and the following stages/blocks, respectively. The basic CNN cell is a convolutional layer that follows the batch normalization (BN) layer and a nonlinear activation (Act) function. The green part of the effective CNN cell conducts the convolutions through pointwise convolution (PConv) and depthwise convolution (DConv) to reduce the parameters. The orange part is a weighted activation for highlighting the vital channels, and the grey part controls the output width (filter number) of the CNN cell. The bolded symbols in Fig. 6 indicate the candidate parameters, including the kernel size $\boldsymbol{k}$, the stride of convolution $\boldsymbol{s}$, the number of filters in the initial convolution $\boldsymbol{f}$, the increasing ratio of input width in each cell $\boldsymbol{f_i}$, the decreasing ratio of squeezed width in weighted activation $\boldsymbol{f_d}$, the adjusted number of output width $\boldsymbol{f_a}$, and the dropout rate in the **Dropout** layer. The last candidate parameter is the repeated time $\boldsymbol{p}$ of the same cells. The residual connection added in the effective CNN cell (the dashed line in Fig. 6(b)) is conditionally controlled by an "on-off" switch, which means the connection is added when the stride of the depth-wise convolution equals one and is removed when the stride is 2.

The number of blocks/stages in the network and the candidate CNN cells should be manually defined to delimit the search space. According to the seven models involved in the comparison, we configured the initial convolutional layer with $\boldsymbol{f} = 32$ filters and $\boldsymbol{s} = 2$ stride and a total of eight down-sampling levels. To generate a powerful model and compare it with the other seven models, we conducted neural network searching based on the ImageNet dataset on the server. Moreover, in the following transfer learning experiments, all eight models are initialized using the pre-trained model on ImageNet. Besides the bottom and top layers, the other six CNN blocks are optimized with the candidate parameters ($\boldsymbol{k} \in \{3, 5\}$, $\boldsymbol{s} \in \{1, 2\}$) and the structures with $\boldsymbol{f_i} = 6$, $\boldsymbol{f_d} = 24$, $\boldsymbol{f_a} = 8n$ (n is the integer) and $1 \leq \boldsymbol{p} \leq 4$. The obtained model is termed the optimized NASNet (ONASNet).

TABLE II
A COMPARISON OF THE TLMS IN MODEL SIZE AND COMPUTING COMPLEXITY

TLM Bone	Size (MB)	FLOPs (G)
VGG16	168.54	30.71
VGG19	229.33	39.04
ResNet50	270.55	7.75
InceptionV3	250.63	5.69
DenseNet121	81.88	5.70
Xception	239.11	9.13
NASNet	52.28	1.15
ONASNet	**47.16**	**0.79**

C. The Configuration of the Learning Process

Five-fold cross-validation is exploited to validate the generalizability of different TLMs. Of note, we divided five-folds upon the EEG subjects, not the detection windows, to ensure that there is no overlap between the training and validation set. We optimized the model using the Adam algorithm during training based on the error measured by categorical cross-entropy. The learning rate is configured based on the models to ensure that the model can gradually learn from the data. We adopted two learning rates (1e-3 and 1e-4) to investigate the capability of different TLMs at slow and fast speeds.

Transfer learning progress includes the training phase and fine-tuning phase. Because ONASNet is generated based on the ImageNet dataset, the bottom structures of the other TLMs are initialized using the parameters pre-trained on the ImageNet. The bottom layers are frozen during the training of the top layers. The transfer learning experiments were implemented on a Windows 10 computer with an Intel Core i7-8700K CPU at 3.70 GHz, 32 GB RAM, and an NVIDIA GeForce GTX 1080Ti, based on TensorFlow 2.3.0-GPU and Python 3.7.11. Each model is trained by a batch size of 69 for 25 epochs, yielding 80 steps for each epoch and 2000 steps. The training progress stops early when the categorical accuracy does not improve on the validation data in the five consecutive epochs. Finally, all the parameters in the TLM are unfrozen for further fine-tuning. The fine-tuning is started from the early stop point of the training phase and implemented for ten epochs. The training data is shuffled at the end of each epoch in both phases.

IV. EXPERIMENTAL RESULTS

A. Comparative Study of Different TLMs

We systematically explored the framework performance using different structures as the bone of the transfer-learning model. The comparison includes (1) the size and the complexity of the model (TABLE II), (2) the capability and the

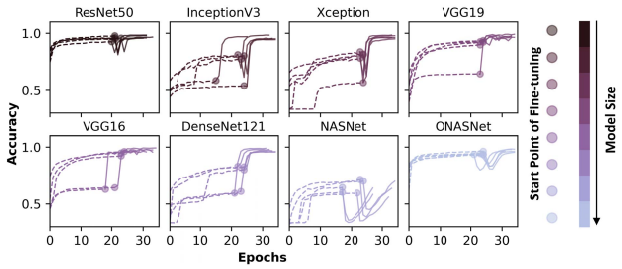

Fig. 7. The comparison of learning performance for different models trained with a slow learning rate (lr = 1e-4).

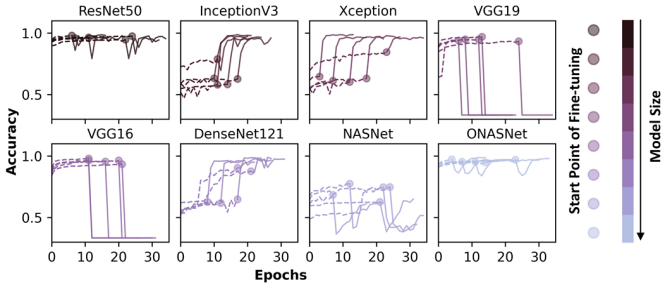

Fig. 8. The comparison of learning performance for different models trained with a fast learning rate (lr = 1e-3).

TABLE III
A COMPARISON OF THE CANDIDATE TLMS IN MODEL LATENCY

TLM Bone	Latency-Mean (ms)	Latency-Std (ms)
VGG16	93.27	3.45
VGG19	108.30	5.82
ResNet50	79.90	2.75
InceptionV3	72.88	3.00
DenseNet121	77.23	2.38
Xception	92.16	3.16
NASNet	75.38	2.49
ONASNet	**66.00**	**2.29**

stability of the model for learning deep representations (Fig. 7 and Fig. 8), and (3) the latency time of model for prediction (TABLE III).

1) Results on Model Size and Complexity: Compared with other models, ONASNet has the smallest model size and model complexity, where the complexity of the model is measured by floating-point operations (FLOPs). The size of FLOPs indicates the computing cost in theory and reflects the requirement of hardware (such as GPU) to run the model. TABLE II demonstrates that the relationship between model size and FLOPs is not linear. For example, the size of ONASNet is only reduced by half compared to DenseNet121, but the value of FLOPs in ONASNet is an order of magnitude lower than DenseNet121. It implies that the dense connections between the feature maps at different levels dramatically increase FLOPs. Therefore, we did not consider this type of structure when designing the candidate structures for the optimized model.

2) Results on Learning Capability and Stability: Each panel of Fig. 7 and Fig. 8 indicates the learning history of one model in five folds. We sorted the panels of figures by the model size and denoted two phases of transfer learning by the dashed line and the solid line, respectively. The circle point means the start point of fine-tuning. ResNet50 and ONASNet show strong learning capability in both phases and achieve convergence faster than other models. Meanwhile, the consistency of five-fold cross-validation is better in ResNet50 and ONASNet. Although the model size of NASNet is comparable to ONASNet, NASNet cannot precisely capture the effective deep representations of our data. It indicates that the generalizability of our CNN cell (Fig. 6) outperforms the *Reduction Cell* proposed in NASNet.

We also investigated the training performance of those models under a slow learning rate (Fig. 7) and a fast learning-rate (Fig. 8). The fast learning rate significantly impacts the stability of learning, and the early stop points of the training are various in different folds. The stability of the VGG structure (VGG16\VGG19) is destroyed when fine-tuning the entire model at a fast rate. It implies that updating the gradient of a large model with a fully cascaded structure is easily influenced by the perturbation. Fortunately, this problem can be relieved by residual and short-cut connections. Under the fast learning rate, ResNet50 and ONASNet still outperform others.

3) Results on Time Consumption: Furthermore, the latency of the model is vital for real-time applications and high-risk alarming. We recorded the latency of different TLMs for predicting one patient in five-fold experiments (TABLE III). With our GPU computing platform, ONASNet outperforms other models by the most negligible average latency (66 ms) and the lowest standard deviation (2.29 ms) in five-fold cross-validation. Other TLMs are also faster than their original version due to the modified lightweight top structure. However, the FLOPs of other models (except NASNet) are at least seven times larger than ONASNet. It means that ONASNet can maintain the performance on a poorly performing computing platform.

TABLE IV
TRADITIONAL CONFUSION MATRIX

	Predicted Positive	Predicted Negative
True Positive	True Positive (TP)	False Negative (FN)
True Negative	False Positive (FP)	True Negative (TN)

TABLE V
THE CONFUSION MATRIX FOR MULTICLASS CLASSIFICATION TASK

	Predicted as 0	Predicted as 1	Predicted as 2
Normal (0)	T_{00}	F_{01}	F_{02}
Seizure-free (1)	F_{10}	T_{11}	F_{12}
Seizure (2)	F_{20}	F_{21}	T_{22}

B. Performance Quantification and Comparison

1) Evaluation Metrics: The typical evaluation metrics include accuracy, precision, recall (sensitivity), and specificity. According to the traditional confusion matrix (TABLE IV), we defined a new confusion matrix (TABLE V) for our

TABLE VI
THE DETECTION PERFORMANCE OF THE PROPOSED METHOD AND NON-RHYTHMIC FEATURE MAPS

Learning Rate	Confusion Matrix							Metrics			
								Accuracy (%)		Precision (%)	
	Labels	BRRM biomarker			Non-rhythmic features (NR)			BRRM	NR	BRRM	NR
1e-4	Normal	**100**	**0**	**0**	100	0	0	**99.67 (±0.67)**	96.33 (±2.21)	**99.68 (±0.63)**	96.81 (±1.85)
	Seizure-free	**1**	**99**	**0**	11	89	0				
	Seizure	**0**	**0**	**100**	0	0	100				
1e-3	Normal	**100**	**0**	**0**	100	0	0	**100 (±0.00)**	92.33 (±4.78)	**100 (±0.00)**	94.15 (±3.34)
	Seizure-free	**0**	**100**	**0**	23	77	0				
	Seizure	**0**	**0**	**100**	0	0	100				

TABLE VII
COMPARISON RESULTS BETWEEN THE PROPOSED METHOD AND OTHER EXISTING METHODS

Studies	Method	Cross-validation	Macro-average Accuracy
Acharya et al. (2018) [12]	Raw data with a 13-layer Deep 1D convolutional neural network (CNN)	10-fold	88.67%
Yavuz et al. (2018) [16]	Cepstral analysis, Mel frequency coefficients feature extraction, artificial neural network (ANN) with generalized regression	5-fold/10-fold	97.80% ± 1.73% (5-fold) 98.07% ± 2.28% (10-fold)
Chandel et al. (2019) [17]	Three-band triadic wavelet decomposition and k-nearest neighbour classifier (KNN)	5-fold	96.00%
San-Segundo et al. (2019) [18]	Fourier transform, wavelet coefficients and empirical mode decomposition (EMD), 2D CNN	5-fold	96.50% ± 0.44%
Gupta et al. (2019) [19]	Fourier–Bessel series expansion (FBSE)-based rhythm separation, weighted multiscale Renyi permutation entropy, least squares-support vector machine (LS-SVM)	10-fold	97.33%
Nabil et al. (2020) [20]	DWT decomposition, approximate entropy, largest Lyapunov exponents and statistical indicators, multiclass SVM	10-fold	96.80%
Tuncer et al. (2020) [21]	An ensemble feature extraction network based on local graph structure and DWT, feature reduction based on ReliefF and neighborhood component analysis, multi-classifiers	5-fold	94.00% (linear discriminant analysis); 96.00% (SVM); 98.67% (ANN); 94.67% (KNN)
Hassan et al. (2020) [22]	Complete ensemble EMD with adaptive noise, normal inverse Gaussian pdf parameters, adaptive boosting	10-fold	98.67%
Khan et al. (2021) [23]	Hilbert vibration decomposition, correlation-based Q-score, long short-term memory (LSTM)-based deep learning	5-fold	96.00% ± 3.84%
Jing et al. (2021) [24]	DWT, energy contribution weighting rules and non-zero processing for signal and feature enhancement, SVM	10-fold	96.59%
Oliva et al. (2021) [25]	105 measurements extracted from the power spectrum, spectrogram, and bi-spectrogram, minimal sequential optimization supported by the polynomial kernel classifier	10-fold	98.00% ± 3.22%
Anuragi et al. (2021) [8]	FBSE-based empirical wavelet transform, 3D phase-space representation, entropy-based features, ensemble learning	10-fold	98.3%
Our work (2021)	Brain-rhythmic recurrence features, ONASNet, transfer-learning	5-fold	99.67% ± 0.67% (lr=1e-4) 100% (lr=1e-3)

multiclass problem. In TABLE V, "T" means true, "F" means False, the first number of the subscript denotes the true label and the second one denotes the predicted label. Three classes are balanced in this study, which indicates that the macro-average accuracy equals the macro-average precision. Therefore, the model's performance is evaluated by the equations (5) and (6), where N is the number of classes.

$$Accuracy = \frac{1}{N}\left(\sum_{i=0}^{N-1} \frac{T_{ii}}{T_{ii} + \sum F_{ij}}\right), \quad j \in \{0, 1, \ldots, N\}, \quad j \neq i \tag{5}$$

$$precision = \frac{1}{N}\left(\sum_{i=0}^{N-1} \frac{T_{ii}}{T_{ii} + \sum F_{ji}}\right), \quad j \in \{0, 1, \ldots, N\}, \quad j \neq i \tag{6}$$

2) Performance Evaluation: Moreover, we evaluated the overall recall and precision metrics for the prediction results obtained based on the BRRM biomarker and ONASNet. Specifically, we compared the proposed biomarker with non-rhythmic feature maps under the same model and training configuration. The results are summarized in TABLE VI, which indicates the contribution of the brain-rhythmic biomarker for classifying seizure-free patterns from normal patterns.

Furthermore, the proposed method finally reaches the best performance of 100% accuracy under the 1e-3 learning rate.

3) Comparison With Other Recent State-of-the-Art Methods: The performance of our proposed detection algorithm is compared with other existing methods published in recent years (TABLE VII). To ensure fairness, we included the studies into comparison under the conditions that (1) experiments were implemented based on the same data (Bonn University epileptic EEG dataset), (2) the results were obtained the same classification task (normal/seizure-free/seizure), (3) the methods were validated through five-fold or ten-fold cross-validation, and (4) the performance was evaluated by the macro-average accuracy.

V. Discussion

The method proposed in this paper is determined based on our pre-experimental research. Firstly, we investigated the effectiveness of the greyscale RPs (Fig. 2) against the original RPs (Fig. 1) and the related quantifications (e.g., RQA). Subsequently, we explored the effects of rhythmic analysis on improving performance. We found that the brain rhythmic recurrence map (Fig. 3) obtained by multi-level wavelet decomposition and recursive phase-trajectory mapping are promising for indicating epileptic patterns. Next, we studied multiple advanced network structures (VGG, ResNet, Inception, DenseNet, Xception, NASNet) to validate the stability of the biomarker on the different models. We also attempted different training configurations to observe their respective learning performance in this step. Besides the effects on learning rate (Fig. 7 and Fig. 8), we also found that the normalization strategy influenced the model accuracy. The cross-rhythm normalization performed better than the inner-rhythm normalization. Finally, we introduced an optimized structure named ONASNet as the bone of the transfer-learning model. We produced the ONASNet structure using the neural network search algorithm (Fig. 5) and the hand-designed CNN cells (Fig. 6). The pre-experiments investigated different CNN blocks and determined two effective and lightweighted CNN cells.

In Section IV.A.2), most TLMs can capture the difference between normal, seizure-free, and seizure patterns through the brain-rhythmic recurrence features and transfer-learning technique, which indicates the effectiveness of the proposed biomarkers and learning configurations. Furthermore, we compared different TLMs in size, complexity, generalizability, stability, and latency, proving the advantages of the proposed optimized network ONASNet. To further discuss the predicted performance of those models, the validation results of five-fold experiments are summarized in the confusion matrices (Fig. 9). Even though the NASNet is the most lightweight model (TABLE II) among the models (a) to (g) in Fig. 9, it is seriously biased due to overfitting. We thought the failure may result from the considerably complex *Inception* connections and *Added* connections applied in the NASNet. In addition, the VGG structure is not appropriate for a faster learning rate. Thus, in our approach, we avoided overcomplicated connections and fully cascading structures like VGG. Referring

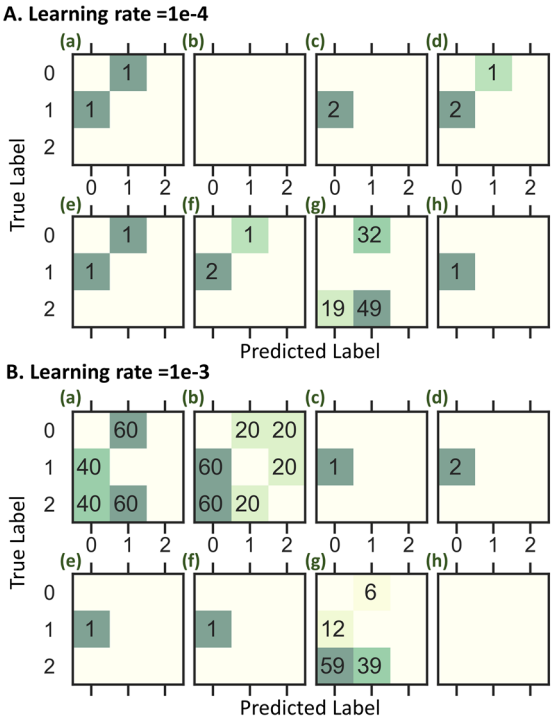

Fig. 9. The misclassified subjects in five-fold cross-validation generated by (a) VGG16, (b) VGG19, (c) ResNet50, (d) Inception V3, (e) DenseNet 121, (f) Xception, (g) NASNet and (e) ONASNet, where normal, seizure-free and seizure patterns are respectively denoted as 0, 1 and 2.

to the usage of separable convolutions in the *Reduction Cell* defined in NASNet, we applied depth-wise convolutions and pointwise convolutions to reduce the parameters in our CNN cell. Referring to the residual learning proposed in the ResNet, we introduced a conditional inverted residual learning with squeeze-and-excitation in our CNN cell (Fig. 6(b)). For the input stage of the model, we applied a simple CNN cell to process the data coarsely (Fig. 6(a)).

Although our method outperforms other methods listed in TABLE VII, we did not involve some seizure detection algorithms developed for different classification tasks or validated through multi-channel EEG datasets. It is valuable to explore those algorithms in the future and extend our method from single-channel EEG signals to multi-channel EEG signals. Moreover, we attained the ONASNet structure through the neural network searching upon a large dataset (ImageNet) and then transferred it from the source domain (image classification) to the target domain (epileptic patterns detection). Therefore, we believe that ONASNet can be further simplified as a smaller model to directly train our small EEG data to reach the same accuracy using less latency.

VI. Conclusion

This paper proposes a new EEG-based seizure detection strategy to decode EEG patterns in the nonlinear dynamic domain through a modified deep learning model. This method (a) decomposes an original EEG signal into brain rhythms, (b) transforms those rhythms as brain-rhythmic recurrence maps (BRRM) inferred from the dynamical

trajectories in phase space, and (c) utilizes an effective neural network (ONASNet) as a feature extractor and classifier to identify normal, seizure-free and seizure patterns. The BRRM-ONASNet framework is partly immune from losing important nonlinear features, which would occur in the complex feature engineering algorithms and traditional machine learning methods.

We divided the brain activities into low, medium, and fast rhythms and then used their corresponding frequency bands (sub-bands) of EEG signals to construct BRRM. Consequently, the constructed BRRM was processed by the multiple channels in the neural network. This method leverages the capability of a multi-channel model in analyzing color images, which avoids using 3D CNN to process sub-bands. The ONASNet is obtained through neural network searching using a complex dataset (ImageNet) and then applied to our small EEG dataset through the transfer-learning technique. To compare with our model, we involved seven acknowledged neural networks (VGG16, VGG19, Reset50, InceptionV3, DenseNet121, Xception, NASNet). Those models were pre-trained on the ImageNet and transferred to the epileptic EEG data with the same learning configuration. The comparative study indicates that BRRM-based biomarkers effectively identify epileptic patterns. The ONASNet-based transfer-learning model outperforms the other seven TLMs in model size, computing complexity, stability, and prediction latency. The BRRM-ONASNet framework achieves an accuracy of 100% in five-fold experiments for predicting three different brain states from Bonn epileptic EEG data. Our performance is better than other reported results in the related studies, but the limitation mentioned in the *Discussion* needs to be further explored and studied in the future.

REFERENCES

[1] J. Saini and M. Dutta, "An extensive review on development of EEG-based computer-aided diagnosis systems for epilepsy detection," *Netw., Comput. Neural Syst.*, vol. 28, no. 1, pp. 1–27, Jan. 2017.

[2] U. R. Acharya, S. V. Sree, G. Swapna, R. J. Martis, and J. S. Suri, "Automated EEG analysis of epilepsy: A review," *Knowl.-Based Syst.*, vol. 45, pp. 147–165, Jun. 2013.

[3] H. Işik and E. Sezer, "Diagnosis of epilepsy from electroencephalography signals using multilayer perceptron and Elman artificial neural networks and wavelet transform," *J. Med. Syst.*, vol. 36, no. 1, pp. 1–13, Feb. 2012.

[4] L. S. Vidyaratne and K. M. Iftekharuddin, "Real-time epileptic seizure detection using EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 2146–2156, Nov. 2017.

[5] A. S. Zandi, M. Javidan, G. A. Dumont, and R. Tafreshi, "Automated real-time epileptic seizure detection in scalp EEG recordings using an algorithm based on wavelet packet transform," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1639–1651, Jul. 2010.

[6] S. Ramgopal *et al.*, "Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy," *Epilepsy Behav.*, vol. 37, pp. 291–307, Aug. 2014.

[7] 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.

[8] A. Anuragi, D. S. Sisodia, and R. B. Pachori, "Epileptic-seizure classification using phase-space representation of FBSE-EWT based EEG sub-band signals and ensemble learners," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103138.

[9] H. Adeli, Z. Zhou, and N. Dadmehr, "Analysis of EEG records in an epileptic patient using wavelet transform," *J. Neurosci. Methods*, vol. 123, no. 1, pp. 69–87, Feb. 2003.

[10] F. Lotte *et al.*, "A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update," *J. Neural Eng.*, vol. 15, no. 3, p. 28, Jun. 2018.

[11] E. B. Assi, D. K. Nguyen, S. Rihana, and M. Sawan, "Towards accurate prediction of epileptic seizures: A review," *Biomed. Signal Process. Control*, vol. 34, pp. 144–157, Apr. 2017.

[12] 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.

[13] M. Zhou *et al.*, "Epileptic seizure detection based on EEG signals and CNN," *Frontiers Neuroinform.*, vol. 12, p. 95, Dec. 2018.

[14] X. Wei, L. Zhou, Z. Chen, L. Zhang, and Y. Zhou, "Automatic seizure detection using three-dimensional CNN based on multi-channel EEG," *BMC Med. Informat. Decis. Making*, vol. 18, no. 5, pp. 71–80, Dec. 2018.

[15] U. R. Acharya, H. Fujita, V. K. Sudarshan, S. Bhat, and J. E. Koh, "Application of entropies for automated diagnosis of epilepsy using EEG signals: A review," *Knowl. Based Syst.*, vol. 88, pp. 85–96, Nov. 2015.

[16] E. Yavuz, M. C. Kasapbaşı, C. Eyüpoğlu, and R. Yazıcı, "An epileptic seizure detection system based on cepstral analysis and generalized regression neural network," *Biocybern. Biomed. Eng.*, vol. 38, no. 2, pp. 201–216, 2018.

[17] G. Chandel, P. Upadhyaya, O. Farooq, and Y. U. Khan, "Detection of seizure event and its onset/offset using orthonormal triadic wavelet based features," *IRBM*, vol. 40, no. 2, pp. 103–112, Mar. 2019.

[18] R. San-Segundo, M. Gil-Martín, L. F. D'Haro-Enríquez, and J. M. Pardo, "Classification of epileptic EEG recordings using signal transforms and convolutional neural networks," *Comput. Biol. Med.*, vol. 109, pp. 148–158, Jun. 2019.

[19] V. Gupta and R. B. Pachori, "Epileptic seizure identification using entropy of FBSE based EEG rhythms," *Biomed. Signal Process. Control*, vol. 53, Aug. 2019, Art. no. 101569.

[20] D. Nabil, R. Benali, and F. Bereksi Reguig, "Epileptic seizure recognition using EEG wavelet decomposition based on nonlinear and statistical features with support vector machine classification," *Biomed. Eng./Biomedizinische Technik*, vol. 65, no. 2, pp. 133–148, Apr. 2020.

[21] T. Tuncer, S. Dogan, F. Ertam, and A. Subasi, "A novel ensemble local graph structure based feature extraction network for EEG signal analysis," *Biomed. Signal Process. Control*, vol. 61, Aug. 2020, Art. no. 102006.

[22] 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, Mar. 2020, Art. no. 105333.

[23] P. Khan, Y. Khan, S. Kumar, M. S. Khan, and A. H. Gandomi, "HVD-LSTM based recognition of epileptic seizures and normal human activity," *Comput. Biol. Med.*, vol. 136, Sep. 2021, Art. no. 104684.

[24] J. Jing, X. Pang, Z. Pan, F. Fan, and Z. Meng, "Classification and identification of epileptic EEG signals based on signal enhancement," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103248.

[25] J. T. Oliva and J. L. G. Rosa, "Binary and multiclass classifiers based on multitaper spectral features for epilepsy detection," *Biomed. Signal Process. Control*, vol. 66, Apr. 2021, Art. no. 102469.

[26] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," *J. Neural Eng.*, vol. 16, no. 3, Jun. 2019, Art. no. 031001.

[27] A. Shoeibi *et al.*, "Epileptic seizures detection using deep learning techniques: A review," *Int. J. Environ. Res. Public Health*, vol. 18, no. 11, p. 5780, May 2021.

[28] U. R. Acharya *et al.*, "Characterization of focal EEG signals: A review," *Future Gener. Comput. Syst.*, vol. 91, pp. 290–299, Feb. 2019.

[29] J. P. Zbilut and C. L. Webber, "Embeddings and delays as derived from quantification of recurrence plots," *Phys. Lett. A*, vol. 171, nos. 3–4, pp. 199–203, Dec. 1992.

[30] P. J. Karoly *et al.*, "Interictal spikes and epileptic seizures: Their relationship and underlying rhythmicity," *Brain*, vol. 139, no. 4, pp. 1066–1078, Apr. 2016.

[31] N. E. Cámpora, C. J. Mininni, S. Kochen, and S. E. Lew, "Seizure localization using pre ictal phase-amplitude coupling in intracranial electroencephalography," *Sci. Rep.*, vol. 9, no. 1, pp. 1–8, Dec. 2019.

[32] M. E. Saab and J. Gotman, "A system to detect the onset of epileptic seizures in scalp EEG," *Clin. Neurophysiol.*, vol. 116, no. 2, pp. 427–442, Feb. 2005.

[33] X. Liu *et al.*, "Clinical implications of scalp ictal EEG pattern in patients with temporal lobe epilepsy," *Clin. Neurophysiol.*, vol. 130, no. 9, pp. 1604–1610, Sep. 2019.

[34] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980* (Lecture Notes in Mathematics), vol. 898, D. Rand and L. S. Young, Eds. Berlin, Germany: Springer, 1981, doi: 10.1007/BFb0091924.

[35] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Phys. Rev. A, Gen. Phys.*, vol. 33, no. 2, pp. 1134–1140, Feb. 1986.

[36] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Phys. Rev. A, Gen. Phys.*, vol. 45, no. 6, pp. 3403–3411, Mar. 1992.

[37] H. D. I. Abarbanel, R. Brown, J. J. Sidorowich, and L. S. Tsimring, "The analysis of observed chaotic data in physical systems," *Rev. Mod. Phys.*, vol. 65, no. 4, pp. 1331–1392, Oct. 1993.

[38] A. K. Chaou, A. Mekhaldi, and M. Teguar, "Recurrence quantification analysis as a novel LC feature extraction technique for the classification of pollution severity on HV insulator model," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 22, no. 6, pp. 3376–3384, Dec. 2015.

[39] P. Bashivan, G. M. Bidelman, and M. Yeasin, "Spectrotemporal dynamics of the EEG during working memory encoding and maintenance predicts individual behavioral capacity," *Eur. J. Neurosci.*, vol. 40, no. 12, pp. 3774–3784, Dec. 2014.

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

[41] 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.

[42] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[43] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.

[44] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1800–1807.

[45] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8697–8710.

[46] B. Zoph and Q. V. Le. (Nov. 1, 2016). *Neural Architecture Search With Reinforcement Learning*. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2016arXiv161101578Z

[47] V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare, and J. Pineau, "An introduction to deep reinforcement learning," *Found. Trends Mach. Learn.*, vol. 11, nos. 3–4, pp. 219–354, Nov. 2018.