Article

Effective Detection of Epileptic Seizures through EEG Signals Using Deep Learning Approaches

Sakorn Mekruksavanich [1] and Anuchit Jitpattanakul [2,3,*]

1 Department of Computer Engineering, School of Information and Communication Technology, University of Phayao, Phayao 56000, Thailand; sakorn.me@up.ac.th
2 Department of Mathematics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand
3 Intelligent and Nonlinear Dynamic Innovations Research Center, Science and Technology Research Institute, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand
* Correspondence: anuchit.j@sci.kmutnb.ac.th

Abstract: Epileptic seizures are a prevalent neurological condition that impacts a considerable portion of the global population. Timely and precise identification can result in as many as 70% of individuals achieving freedom from seizures. To achieve this, there is a pressing need for smart, automated systems to assist medical professionals in identifying neurological disorders correctly. Previous efforts have utilized raw electroencephalography (EEG) data and machine learning techniques to classify behaviors in patients with epilepsy. However, these studies required expertise in clinical domains like radiology and clinical procedures for feature extraction. Traditional machine learning for classification relied on manual feature engineering, limiting performance. Deep learning excels at automated feature learning directly from raw data sans human effort. For example, deep neural networks now show promise in analyzing raw EEG data to detect seizures, eliminating intensive clinical or engineering needs. Though still emerging, initial studies demonstrate practical applications across medical domains. In this work, we introduce a novel deep residual model called ResNet-BiGRU-ECA, analyzing brain activity through EEG data to accurately identify epileptic seizures. To evaluate our proposed deep learning model's efficacy, we used a publicly available benchmark dataset on epilepsy. The results of our experiments demonstrated that our suggested model surpassed both the basic model and cutting-edge deep learning models, achieving an outstanding accuracy rate of 0.998 and the top F1-score of 0.998.

Keywords: electroencephalography; epileptic seizure detection; deep learning; deep residual network; efficient channel attention

1. Introduction

Epilepsy is an enduring neurological disorder that primarily impacts the central nervous system, with its primary expression manifesting in the brain. It is characterized by the occurrence of sudden and repetitive seizures [1]. This ailment is recognized as the third most prevalent neurological condition, ranking closely behind stroke and Alzheimer's disease [2]. An epileptic seizure is a temporary event characterized by unusual and exaggerated activity of brain neurons, leading to the presence of signs or symptoms [3]. Seizures can be identified by observing and analyzing various physiological indicators, such as brain and muscle activities, heart rate, oxygen levels, synthetic speech, or visual patterns. Techniques like electroencephalography (EEG), electrocardiogram (ECG), electromyography (EMG), movement tracking, or video capturing on a person's head and body are used for this purpose [4].

To detect anomalous behaviors and enable early detection of epileptic seizures before they escalate into more severe states, human activity recognition (HAR) can be effectively

applied [5]. Recent scholarly investigations have explored a range of techniques for identifying anomalous behaviors [6], including wearable technologies, sensor-based approaches, and ambient instrument methodologies. The activated notification system ensures the verification of action identification. However, the accuracy of detecting these behaviors depends on the thorough analysis and precise acquisition of feature patterns [7].

EEG signals are preferred for their cost-effectiveness, portability, and ability to exhibit distinct frequency-dependent patterns [8]. The EEG is a measurement technique that captures the brain's bioelectric activities by recording voltage fluctuations resulting from the ionic flow of neurons [9]. To identify epileptic seizures accurately, capturing signals over an extended duration is necessary, which can introduce complexity due to multiple channels used for storage. Studies have also raised concerns over wearable devices' energy use and data storage limits posing challenges for creating seizure forecasting tools [10]. These portable gadgets tend to swiftly drain batteries and lack capacity to save the huge data flows needed to reliably predict seizures over extended periods. Overcoming these power and memory roadblocks to enable precise ambulatory monitoring and algorithms will require further innovations going forward. Nonetheless, EEG signals can be influenced by disturbances arising from different origins, including the main power supply, movement of electrodes, and muscular vibrations [11]. The presence of noisy EEG signals poses a significant challenge for healthcare professionals in diagnosing epileptic seizures effectively. To address these difficulties, extensive research is currently being conducted to detect and forecast epileptic seizures through the utilization of EEG methods, alongside tools like magnetic resonance imaging (MRI) and artificial intelligence (AI) methodologies [12]. The realm of diagnosing epileptic seizures has been incorporating traditional machine learning (ML) and deep learning (DL) techniques within the framework of AI methods [13].

Numerous ML algorithms have been established for epileptic seizure identification, incorporating statistical, temporal, spectral, time–frequency domain, and nonlinear features [14]. Traditional ML approaches involve a trial-and-error method for selecting features and classifiers [15]. A thorough grasp of signal processing and data mining techniques is essential for constructing accurate models. A recent study [16] utilized three ML models—support vector machines, linear discriminant analysis, and multilayer perceptrons—to differentiate resting state EEG data between healthy subjects and those with psychogenic non-epileptic seizures. Specifically, these algorithms aimed to uncover connections between measures of functional brain connectivity and the eventual categorical diagnosis. By modeling these complex relationships, the systems classified individual data points as belonging to either the healthy control group or the psychogenic seizure group. Initial findings demonstrated promise in using patterns of functional connectivity derived from EEGs to accurately predict which subjects were suffering from non-epileptic events via completely automated ML. However, further validation is still needed, particularly around generalizability to diverse patient subgroups. While these models perform well with small data sets, the field has also implemented sophisticated DL techniques for epileptic seizure identification [17]. DL models, unlike traditional ML methods, require a substantial amount of data during training due to their complex feature mapping spaces. This phenomenon leads to overfitting challenges when confronted with insufficient data.

The primary objective of this research is to utilize DL networks to detect epilepsy by automatically processing EEG signals and recognizing patterns. The aim is to identify the spatial distribution and temporal characteristics of spikes and seizures. Convolutional neural networks (CNN), long short-term memory (LSTM), and gated recurrent unit (GRU) are among the methods employed in this research. We introduce a new deep residual model, ResNet-BiGRU-ECA, to accurately identify epileptic seizures by analyzing EEG data.

To evaluate the performance of our model, we utilized a publicly available epilepsy dataset. This benchmark compilation contained EEG readings segmented into five distinct health categories—one representing active epileptic seizures and the remaining four encompassing normal, non-seizure brain activity. Leveraging this diverse test set enabled robust assessment of the model's ability to accurately differentiate between the pathological

seizure state and healthy function. Through this analysis, we aim to demonstrate several key contributions of the current research, summarized as follows:

- This research presents a framework for epileptic seizure detection (ESD) using EEG data to assess the performance of various DL structures within this particular domain.
- The suggested method introduces a deep residual model that incorporates residual blocks, an efficient channel attention (ECA) module, and a bidirectional GRU (Bi-GRU). This model adeptly captures extended data sequences, extracts spatio-temporal features, and carries out EEG classification.

The paper's organization is as follows: Section 2 offers a summary of pertinent literature and prior studies. Section 3 outlines the proposed DL-based framework for epileptic seizure recognition. The results are showcased and examined in Section 4, and Section 5 wraps up and deliberates on potential avenues for future research.

2. Related Work

In this section, we review several studies that have used DL to classify epileptic and non-epileptic activities. One study introduced a DL model called the pyramidal one-dimensional CNN. This CNN model uses improved parameters that are less amenable to training than those used in conventional CNNs. The model achieved a remarkable accuracy of 99.1% when used to distinguish between different behaviors, including typical behaviors such as eyes closed, eyes open, and pre-ictal, as well as unusual behaviors such as inter-ictal and ictal [18].

Conventional DL models struggle to process lengthy, variable input data like text, sensor readings over time, or video [19]. Yet these sequential data types represent abundant real-world information. Recurrent neural networks (RNNs) have emerged to enable DL on such ordered series with temporal dynamics. Through internal state units retaining context, RNNs can analyze signals that change over timescales like EEG—gaining widespread use in physiology. Standard deep networks fail on such variable, redundant sequential inputs. But by propagating context, RNNs overcome challenges of long inputs with fluctuations and redundancy across temporal or spatial dimensions [7]. This unique feedback architecture provides short-term memory lacking in feedforward networks to better extract patterns from sequential data like measurements over time.

The primary limitation of a basic RNN lies in its inability to retain information effectively over short periods of time. This is mainly because RNNs struggle to propagate information from earlier time steps to later ones, especially when dealing with long sequential data [20]. Another challenge faced by RNNs is the vanishing gradient problem [21], which occurs when the gradients diminish significantly during backpropagation. To overcome the short-term memory issue, researchers devised a solution in the form of LSTM neural networks [22]. LSTM networks address the problem by allowing the model to selectively store and access relevant information, making them more adept at handling long sequences and retaining essential details.

In their study, Golmohammadi et al. [23] evaluated two LSTM designs: one with three layers and the other with four layers, both combined with the Softmax classifier. The researchers reported that their findings were considered appropriate. In another study [24], a three-layer LSTM architecture was used for feature extraction and classification. The final fully connected layer commonly employed the sigmoid activation function for classification purposes. Furthermore, in a study conducted by another group [25], two structures, LSTM and GRU, were investigated. The structure of the LSTM/GRU model included a reshaped layer, succeeded by four layers of LSTM/GRU with an activation function, and, ultimately, a fully connected layer featuring a sigmoid activation function. In summary, these studies explored various LSTM designs and activation functions to improve feature extraction and classification accuracy.

In the realm of sophisticated DL models, it has been observed that employing more intricate and deeper architectures leads to enhanced accuracy compared to the feature learning approach discussed earlier. These prototypes employ CNNs to identify features

autonomously [26]. In particular, the CNN feature extractor is often denoted as the backbone when it comes to object recognition, setting it apart from the complete model architecture. In this investigation, we adopt a CNN-based feature extractor as the foundational framework. DL techniques, such as CNNs and RNNs, have demonstrated their ability to obtain state-of-the-art performance by autonomously learning the underlying characteristics from raw sensor data. The concept of deepening neural networks has evolved with the emergence of hybrid networks. These hybrid models combine diverse architectural designs, leading to improved feature representation and enhancing both computing and network achievements. Moreover, this development opens up opportunities for DL-based techniques in portable electronics.

3. Methodology

In this section, we delineate the structure applied in our study to assess the importance of EEG signals concerning epileptic seizures. The suggested ESD framework encompasses four separate phases: (i) collecting data, (ii) preparing data, (iii) processing data, and (iv) performing classification. To visually represent the approach used to detect seizures, refer to Figure 1.
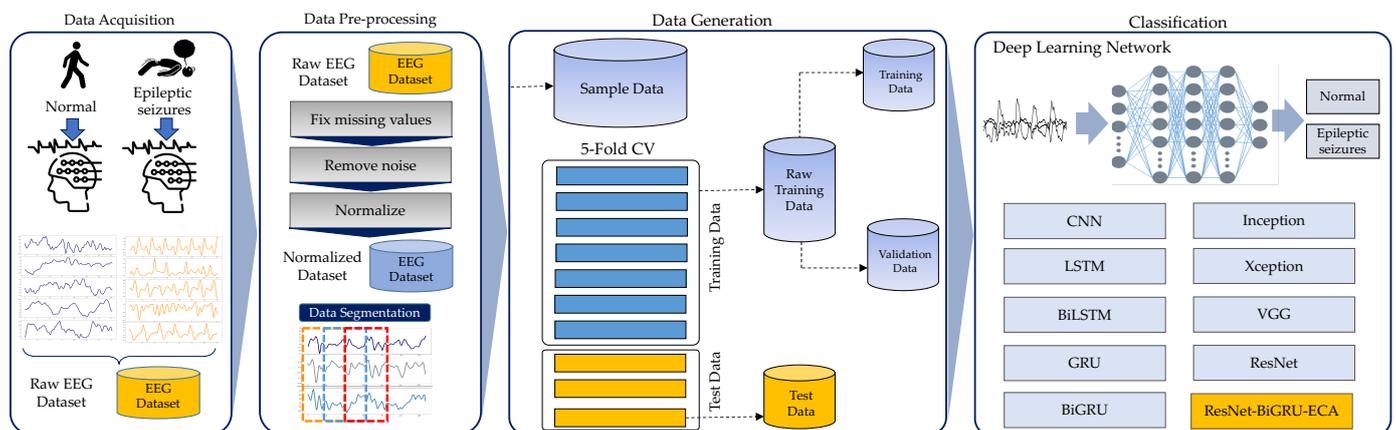
Figure 1. The epileptic seizure detection (ESD) framework proposed in this study.

3.1. Data Acquisition

Data acquisition refers to the systematic process of collecting and preserving digital or numerical data to incorporate it into our mathematical framework. This can be done either in its original, unprocessed state or after undergoing preliminary processing, depending on data accessibility. Therefore, it is crucial to have a comprehensive understanding of the dataset before obtaining the data.

To develop and test our model, we employed the epileptic seizures recognition dataset (ESRD), a publicly accessible benchmark widely utilized for detecting seizures [27]. This compilation contains EEG readings from multiple patients, acquired through a standardized 128-channel system and averaged reference methodology. The analogue signals were digitized via 12-bit analog-to-digital conversion and continuously saved onto a computer at 173.61 Hz sampling frequency. Understanding such data collection and processing details is crucial for appropriate methodology. The ESRD dataset comprises data from 500 individuals, including 11,500 instances of time-series EEG signals specifically designed for investigating EEG signal alterations during seizure events. Before being made available online, the initial dataset undergoes preliminary processing by the university of California Irvine (UCI). Each dataset sample in our study consists of 4097 data points, further divided into 23 parts, with each part comprising 178 data points representing one second of data. These 23 components are shuffled in a random manner, yielding a collective count of 11,500 time-series EEG signal samples sourced from the 500 individuals in the study group. The UCI dataset for recognizing epileptic seizures encompasses five distinct medical

conditions, with one being explicitly linked to epileptic seizures, while the remaining four conditions pertain to individuals displaying no signs of epilepsy. More information about these five categories and their respective samples can be found in Table 1.

Table 1. Summarized details of the epileptic seizures recognition dataset (ESRD).

Class	No. of Samples	ESR Classes	Details
1	2300	Ictal	Recordings made during epileptic episodes
2	2300	Pre-ictal	Recording obtained prior to the onset of a single-site seizure
3	2300	Inter-ictal	Recording obtained prior to the onset of the consecutive seizure
4	2300	Healthy (close eye)	An individual in good health with their eyes shut
5	2300	Healthy (open eye)	An individual in good health with their eyes open

Numerous analyses have recommended simplifying the multi-class categorization challenge embodied in the UCI seizure dataset, which segregates EEG data into five unique groups as shown in Figure 2. Rather than attempting to draw boundaries between all categories, they propose concentrating specifically on discriminating class one—representing active epileptic seizures—from aggregate signals of the other classes signifying ordinary non-seizure brain states. This binary framework focuses model resources on the most vital differentiation between pathological and healthy function. Though it sacrifices granularity within normal activity subtypes, enhanced accuracy and reliability in identifying the central seizure condition demonstrate the sensibility of this targeted approach.
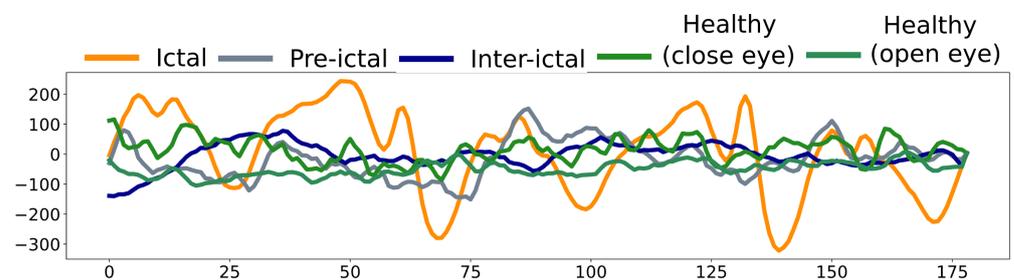
Figure 2. Electroencephalography (EEG) waveforms of five epileptic seizures recognition (ESR) classes from the epileptic seizures recognition dataset (ESRD).

3.2. Data Pre-Processing

During this phase, the EEG data undergo filtering and normalization procedures. The goal is to generate a dataset that is uniform and appropriate for training a detection model. In this approach, any incomplete or anomalous data values are removed from the dataset according to the following procedure:

- We used the linear interpolation procedure for imputation to address the issue of incomplete values in sensor data. This helped eliminate any existing noises. To reduce noise that could obscure relevant EEG patterns, we implemented sequential low and median pass filtering. Initially, a third-order Butterworth filter eliminated high frequency artifacts over 20 Hz. Next, a third-order median filter replaced each point with the regional median value, calculated from surrounding data. By emphasizing the median, sporadic outliers and anomalies get rejected in favor of the less skewed central tendency. This substitutes misleading erratic spikes with smoothed waveforms reflecting underlying dynamics. The process removes irrelevant deviations so key shape features needed for model training remain, without distortion from sporadic spikes that misrepresent normal or pathological brain activity.

- In addition, we standardized each individual segment of EEG data using a normalization method that involved calculating the mean and standard deviation. This step was crucial in ensuring consistency and comparability across the data.

The normalization process entails using the min–max method to adjust the unprocessed EEG data linearly. After the data are cleaned and normalized, they become the input for the subsequent data preparation and classification procedures. To facilitate the training of the classifier, the data are partitioned based on the chosen methodology. The second group is subsequently employed as a test set to assess the performance of the classifier that has been trained.

3.3. The Proposed ResNet-BiGRU-ECA Model

This study introduces the ResNet-BiGRU-ECA model that utilizes EEG signals to identify epileptic seizures accurately. The model demonstrated in this study can autonomously extract distinctive features from the input sensor data. Figure 3 provides a depiction of the ResNet-BiGRU-ECA architecture, comprising a convolutional block (ConvB) and eight hybrid residual blocks designed for extracting spatial features with unique characteristics. These features are further processed through a global average pooling (GAP) layer and a fully connected (FC) layer.
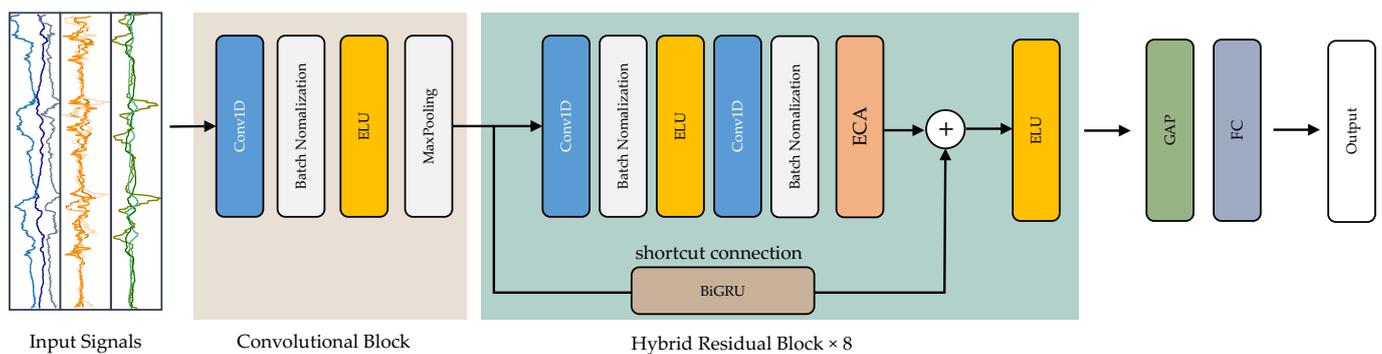
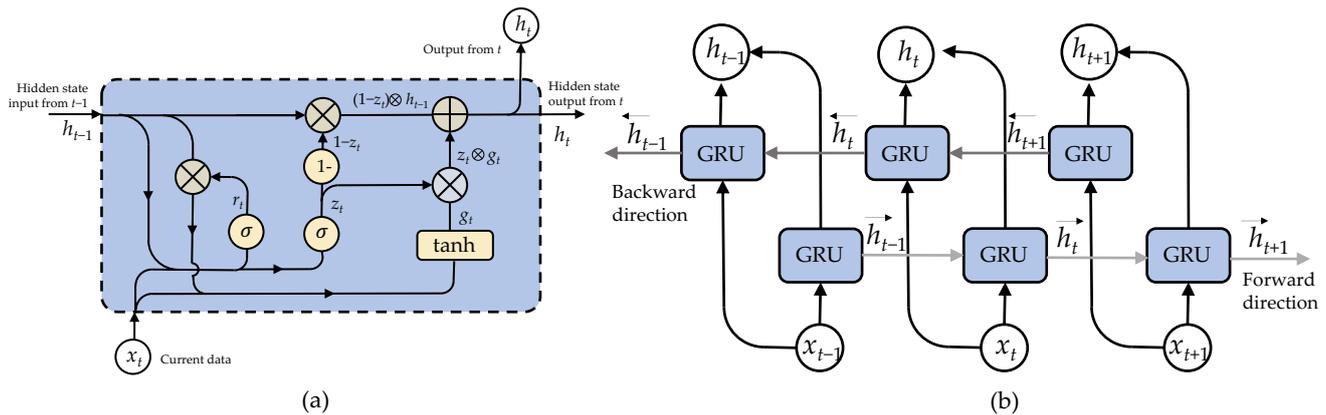**Figure 3.** The structure of the ResNet-BiGRU-ECA model introduced in this study.

3.3.1. Convolutional Block

CNNs are a type of neural network commonly used in supervised learning. CNNs have a specific architecture in which each neuron in one layer is connected to every neuron in the next layer. The input to each neuron is transformed into an output via an activation function [28]. Two key properties of the activation function that impact CNN performance are sparsity (having many zeros) and ability to propagate gradients to lower layers during backpropagation. Overall, CNNs leverage their specialized architecture with sparse activation functions to effectively learn representations of data like images for classification and other supervised learning tasks [29]. In CNNs, pooling operations are frequently employed to reduce dimensionality. Two frequently employed pooling functions include max-pooling and average-pooling, both of which aid in extracting the most significant features from the data.

In this research, ConvB is employed to extract foundational features from unprocessed sensor data. The ConvB structure comprises four components: 1D-convolutional (Conv1D), batch normalization (BN), exponential linear unit (ELU), and max-pooling (MP) layers, as visualized in Figure 3. The Conv1D layer utilizes multiple trainable convolutional kernels to detect distinct features, generating a feature map for each kernel. The BN layer is applied to improve stability and accelerate the training process. Additionally, the ELU layer is integrated to enhance the model's expressive capacity. Furthermore, the MP layer is used to downsize the feature map's dimensions while retaining the most salient features. This amalgamation of layers within the ConvB architecture assists in extracting valuable foundational features from the raw sensor data.

3.3.2. Bidirectional Gated Recurrent Unit

GRU was introduced to address the issue of gradients either growing too large or vanishing in RNNs. However, the incorporation of memory cells in the GRU design elevates the memory resource requirements [30]. Unlike the LSTM model, the GRU is a more streamlined version, lacking a separate memory cell in its architecture [31]. Instead, the GRU network integrates update and reset gates, which play a role in determining the adjustments made to each hidden state. These gates control the information that should be propagated to the following state and the data deemed unnecessary, as illustrated in Figure 4a.

(a)   (b)

Figure 4. The structure of the ResNet-BiGRU-ECA model introduced in this study: (**a**) GRU cell and (**b**) unroll BiGRU.

The calculation of the hidden state h_t at time t in a GRU model involves utilizing the update gate z_t, reset gate r_t, current input x_t, and the prior hidden state h_{t-1}. These components work together to manage the flow of information in the GRU network efficiently.

$$z_t = s(W_z x_t \oplus U_z h_{t-1}) \tag{1}$$

$$r_t = s(W_r x_t \oplus U_r h_{t-1}) \tag{2}$$

$$g_t = \tanh(W_g x_t \oplus U_g(r_t \otimes h_{t-1})) \tag{3}$$

$$h_t = ((1 - z_t) \otimes h_{t-1}) \oplus (z_t \otimes g_t) \tag{4}$$

where s denotes a sigmoid function, $\oplus$ denotes an elementary addition operation, and $\otimes$ denotes an elementary multiplication operation.

In 1997, Schuster and Paliwal [32] introduced the bidirectional RNN (BiRNN) to address limitations of traditional unidirectional RNNs. The key innovation was to have the outputs at each time step incorporate contextual information not just from preceding time steps but also future time steps. This is achieved by training two separate hidden states, one processing the inputs in a forward direction and the other processing in reverse. The outputs at each time step are then computed based on the hidden states from both directions. By leveraging future context as well as past, BiRNNs can better model sequential data like text or speech compared to unidirectional RNNs. This bidirectional architecture has become widely adopted for sequence modeling tasks.

Within this BiRNN framework, the neurons in a standard RNN are segregated into two separate components: one dedicated to processing information in the forward direction and the other focused on information from the opposite direction. Notably, the output of positive neurons is independent of negative neurons, and vice versa. This arrangement

gives rise to a general structure illustrated in Figure 4b. The underlying computational processes are outlined by the equations provided below.

$$\overrightarrow{h_t} = GRU(x_t, \overrightarrow{h_{t-1}}) \tag{5}$$

$$\overleftarrow{h_t} = GRU(x_t, \overleftarrow{h_{t+1}}) \tag{6}$$

$$h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}] \tag{7}$$

3.3.3. ECA Mechanism

The ECA strategy holds great promise for boosting the effectiveness of deep CNNs. However, most current approaches focus on developing intricate attention components to enhance effectiveness, inadvertently leading to more complex models and increased processing demands. A solution known as ECA has been proposed to address the challenges of overfitting and high computational needs [33]. This ECA module determines weights for individual channels and captures interrelationships among different channels.

In time series data, the norm involves assigning higher weights to crucial features while assigning lower weights to less relevant ones. Herein, the ECA method takes on the responsibility of prioritizing pertinent information, thus bolstering the network's capacity to discern and respond to pivotal characteristics. The configuration of the ECA module is visually represented in Figure 5. The channel weights in ECA are generated by applying a 1D convolution with a creative choice of kernel size (k) to the aggregated data obtained through GAP. The choice of the value for k is decided by mapping the channel dimensions (C).
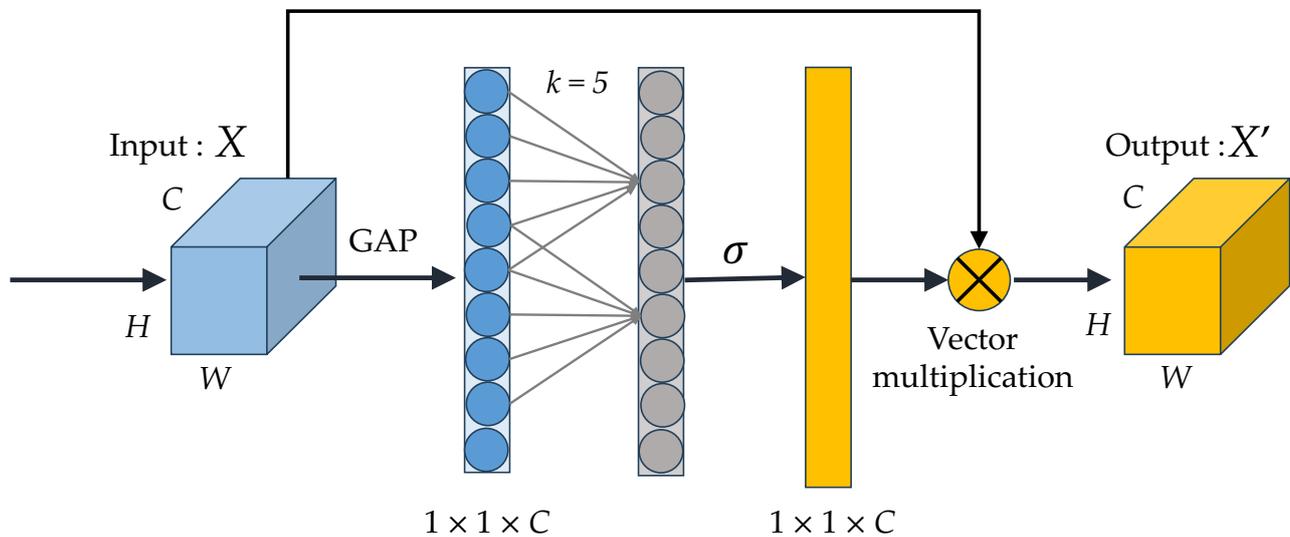
Figure 5. The structure of the efficient channel attention (ECA) block.

3.3.4. Training and Hyperparameters

Model performance hinges critically on sufficient volumes of representative, varied data for training along with properly tuned architectural design parameters called hyperparameters, like iteration counts, learning rates, batch size, activation logic, etc. We applied a standardized approach separating data into training (for hyperparameter optimization) and holdout validation sets (for independent comparative testing). Guided by trial and error, the following settings maximized eventual accuracy: 128 batch size, 1×10^{-3} learning rate, 200 epoch count, plus adaptive logic to cut the learning rate by 25% after 10 stagnant epochs lacking improvement. Beyond tuning, we enabled data shuffling before each epoch for robustness. For model specification, we utilized an Adam optimizer for weight adjust-

ments and cross-entropy error quantification. Refer to Table 2 for full architectural details, including all set hyperparameters for our customized ResNet-BiGRU-ECA framework.

Table 2. The summary of hyperparameters of the ResNet-BiGRU-ECA used in this work.

Stage	Hyperparameters		Values
Architecture	Convolutional Block		
	1D-Convolution	Kernel Size	5
		Stride	1
		Filters	64
	Batch Normalization		-
	Activation		ELU
	Max Pooling		2
	Hybrid Residual Block × 8 Main Path		
	1D-Convolution	Kernel Size	5
		Stride	1
		Filters	32
	Batch Normalization		-
	Activation		ELU
	1D-Convolution	Kernel Size	5
		Stride	1
		Filters	64
	Batch Normalization		-
	ECA Module		-
	Parallel Path		
	BiGRU Unit		128
	Global Average Pooling		-
	Flatten		-
	Dense		128
Training	Loss Function		Cross-entropy
	Optimizer		Adam
	Batch Size		64
	Number of Epochs		200

3.4. Cross Validation

The k-fold cross-validation (k-CV) procedure is a technique employed to assess the performance of a classification model [34]. This method entails partitioning a dataset, collected from one or more sources, into approximately equal-sized, distinct, and non-overlapping subsets of k. Each of these subsets is then utilized to evaluate the model, which is trained on the remaining $k - 1$ subsets. The overall performance of the model is determined by computing the mean of various performance metrics, such as accuracy, precision, recall, and F-measure, all derived from the k-CV process [35].

Nonetheless, it is important to acknowledge that this method can pose significant computational demands, particularly when working with extensive datasets or when k is set to a high value. In our study, we have employed the k-CV procedure with a chosen value of $k = 5$, as depicted in Figure 6, to assess the effectiveness of our models.
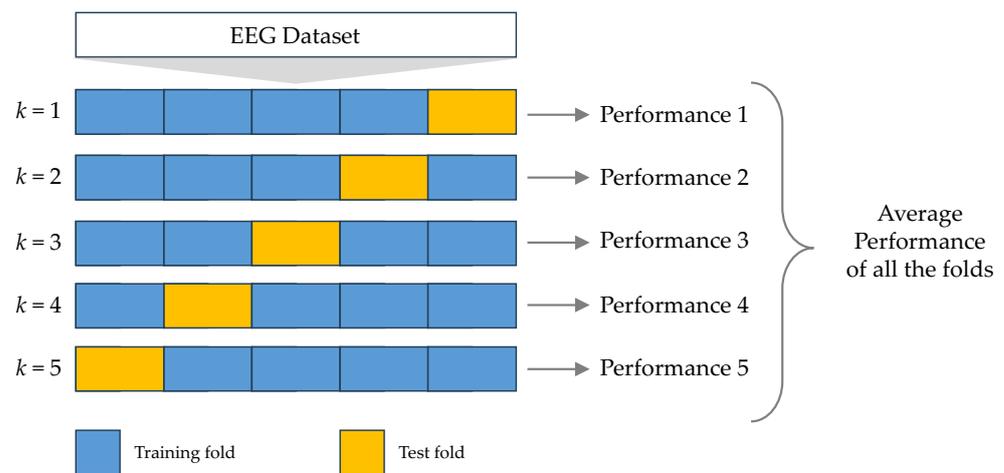
Figure 6. The 5-fold cross validation.

4. Experimental Results and Discussion

In this section, we will detail the experimental setup and showcase the outcomes acquired through our assessment of DL models for the purpose of identifying epileptic seizures.

4.1. Experiments

The experiments in this study were conducted using Google Colab Pro, which provides access to Tesla V100 GPUs. The implementation was done in Python 3.6.9, utilizing key libraries, including TensorFlow 2.2.0 for building the neural network models, Keras 2.3.1 for the high-level API, Scikit-Learn for machine learning utilities, Numpy 1.18.5 for numerical processing, and Pandas 1.0.5 for data manipulation. By leveraging Google Colab and these state-of-the-art libraries, the experiments could be efficiently run on powerful hardware, enabling the exploration of deep neural network architectures for the research questions under investigation.

This empirical study compared four main DL architectures for the application of ESD on the ESRD dataset. The models tested were CNNs, LSTMs, bidirectional LSTMs (BiLSTMs), GRUs, and BiGRUs. These represented the state-of-the-art approaches. Our proposed model, called ResNet-BiGRU-ECA, was benchmarked against these models to assess its performance at identifying seizures relative to conventional architectures. By evaluating both unidirectional and bidirectional variants of LSTM and GRU models, we aimed to thoroughly compare our novel approach to existing methods using these epileptic seizure data.

Additionally, we conducted an in-depth investigation into various CNN backbone models. Specifically, we examined VGG16 [36], ResNet18 [37], Pyramid-Net18 [38], Inception [39], and Xception [40] to perform a comprehensive experimental comparison analysis. These models were considered as potential solutions for addressing the challenge of time-series classification. Consequently, we restructured each model to suit the context of ESR.

In this study, we conducted research on two distinct scenarios involving EEG data for ESRD. In each of these scenarios, we utilized separate datasets for both training and testing DL models, as detailed in Table 3.

Table 3. A compilation of experiments applied in this study.

Scenario	Description
I	Subjects in category 1 exclusively experience epileptic seizures, whereas those in category 3 are strictly categorized as non-epileptic.
II	Subjects in category 1 are the sole ones with epileptic seizures, whereas subjects in the remaining classes are designated as non-epileptic.

4.2. Experimental Results

In each of our experiments, we used the ESRD dataset for training DL models. To assess these models, we employed a 5-CV approach. Our study centered on assessing the effectiveness of five core DL models (CNN, LSTM, BiLSTM, GRU, and BiGRU), in addition to cutting-edge DL models, within the framework of the two situations outlined in Tables 4 and 5.

Table 4. Experimental results of the deep learning (DL) networks trained and tested using electroencephalography (EEG) from scenario I.

Model	Recognition Performance		
	Accuracy	Loss	F1-Score
CNN	0.882 (±0.1911)	0.245 (±0.2322)	0.849 (±0.0258)
LSTM	0.982 (±0.0019)	0.142 (±0.0487)	0.982 (±0.0019)
BiLSTM	0.982 (±0.0016)	0.138 (±0.0467)	0.982 (±0.0016)
GRU	0.988 (±0.0041)	0.081 (±0.0465)	0.988 (±0.0041)
BiGRU	0.985 (±0.0016)	0.116 (±0.0240)	0.985 (±0.0016)
ResNet-BiGRU-ECA	0.998 (±0.0007)	0.008 (±0.0021)	0.998 (±0.0008)

The DL models underwent training and testing using EEG data from scenario I, as detailed in Table 4. Our experiments revealed that the ResNet-BiGRU-ECA model, as proposed, demonstrated exceptional efficiency, boasting an average accuracy of 0.998 and an average F1-score of 0.998.

Table 5. Experimental results of the deep learning (DL) networks trained and tested using electroencephalography (EEG) from scenario II.

Model	Recognition Performance		
	Accuracy	Loss	F1-Score
CNN	0.973 (±0.0075)	0.140 (±0.0284)	0.957 (±0.0132)
LSTM	0.985 (±0.0023)	0.095 (±0.0280)	0.976 (±0.0039)
BiLSTM	0.983 (±0.0032)	0.132 (±0.0436)	0.973 (±0.0051)
GRU	0.985 (±0.0027)	0.104 (±0.0388)	0.977 (±0.0042)
BiGRU	0.985 (±0.0042)	0.113 (±0.0367)	0.977 (±0.0065)
ResNet-BiGRU-ECA	0.996 (±0.0010)	0.020 (±0.0059)	0.994 (±0.0016)

The DL models underwent training and testing with EEG data sourced from scenario II, as specified in Table 5. Upon analyzing the experiment results, it became evident that the ResNet-BiGRU-ECA model, as initially proposed, displayed the highest level of effectiveness. Its impressive performance supports this conclusion, boasting an average accuracy rate of 0.996 and an average F1-score of 0.994.

4.3. Comparative Results with ML Models

Guided by prior analyses [5], we selected leading ML classifiers for comparative benchmarking, including k-nearest neighbors (KNN), naive Bayes (NB), logistic regression (LR), random forest (RF), decision trees (DT), stochastic gradient boosting (SGDC), and gradient boosting (GB). Recent studies confirm the utility of these algorithms paired with neurologists in accurately detecting seizures and characterizing epileptiform EEG dynamics [41,42]. Our experiments leveraged the standard ESRD dataset under equivalent scenarios to examine model performance variability when using raw EEG readings versus extracted feature sets as inputs. Table 6 presents accuracy outcomes with our proposed deep ResNet-BiGRU-ECA architecture versus these widely adopted shallow ML approaches. By evaluating on equal inputs, we aimed to isolate the performance gains stemming solely from algorithmic and architectural optimizations rather than data pre-processing.

Table 6. Comparative results of the proposed model and machine learning (ML) models using electroencephalography (EEG) data from the epileptic seizures recognition dataset (ESRD).

Model	Scenario I		Scenario II	
	Accuracy	F1-Score	Accuracy	F1-Score
KNN	0.870	0.836	0.890	0.851
NB	0.950	0.878	0.950	0.878
LR	0.790	0.878	0.790	0.878
RF	0.922	0.892	0.937	0.901
DT	0.910	0.787	0.910	0.787
SGDC	0.810	0.719	0.810	0.726
GB	0.800	0.799	0.834	0.819
ResNet-BiGRU-ECA	0.998	0.998	0.996	0.994

Results showed NB reached 95% accuracy, the sole ML technique with comparable proficiency. However, our deep ResNet-BiGRU-ECA architecture significantly exceeded all benchmark methods under both data scenarios. Surpassing 99% accuracy and 99% F1, the optimized network architecture demonstrated superior feature extraction and pattern recognition even from raw EEG readings. This substantial performance gap despite shared inputs suggests deep networks intrinsically outperform shallow ML at derives nuanced relationships within complex physiological signals. Rather than relying on predefined assumptions in simplified models, DL constructs intricate representations via hierarchical data transformations. Though some conventional methods approach sufficiency for classification tasks, deep neural networks attain state-of-the-art performance by learning intricate embeddings uniquely tuned to the intricacies of the data through backpropagation and gradient descent optimization.

4.4. Comparative Results with DL Models

The ResNet-BiGRU-ECA model under examination is subjected to a comparative analysis alongside previously developed models using the same dataset, specifically the ESDR dataset. Prior research studies [36–40] have consistently shown that leveraging CNN-based DL architectures yields remarkable results in the field of time-series classification. The existing literature introduced the 5-CV methodology, which we also employed in our study. The summary of comparative results can be found in Tables 7 and 8. These outcomes reveal that the ResNet-BiGRU-ECA model, as presented here, exhibits superior accuracy compared to the earlier models across the majority of actions.

Table 7. A side-by-side comparison of the outcomes achieved by the proposed model and state-of-the-art deep learning (DL) networks, all trained and tested using electroencephalography (EEG) data from scenario I.

Model	Recognition Effectiveness		
	Accuracy	Loss	F1-Score
Inception	0.995 (±0.0021)	0.03 (±0.01)	0.995 (±0.0021)
Xception	0.984 (±0.0213)	0.10 (±0.14)	0.984 (±0.0021)
VGG11	0.975 (±0.0065)	0.35 (±0.18)	0.975 (±0.0065)
VGG13	0.980 (±0.0059)	0.27 (±0.23)	0.980 (±0.0059)
VGG16	0.985 (±0.0039)	0.15 (±0.07)	0.985 (±0.0039)
VGG19	0.989 (±0.0048)	0.08 (±0.05)	0.989 (±0.0048)
ResNet16	0.995 (±0.0017)	0.02 (±0.01)	0.995 (±0.0017)
ResNet18	0.994 (±0.0018)	0.03 (±0.01)	0.994 (±0.0018)
ResNet34	0.994 (±0.0029)	0.04 (±0.01)	0.994 (±0.0029)
ResNet-BiGRU-ECA	0.998 (±0.0007)	0.01 (±0.01)	0.998 (±0.0008)

Table 8. A side-by-side comparison of the outcomes achieved by the proposed model and state-of-the-art deep learning (DL) networks, all trained and tested using electroencephalography (EEG) data from scenario II.

Model	Recognition Effectiveness		
	Accuracy	Loss	F1-Score
Inception	0.994 (±0.0018)	0.03 (±0.01)	0.991 (±0.0028)
Xception	0.962 (±0.0663)	0.22 (±0.39)	0.952 (±0.0808)
VGG11	0.985 (±0.0019)	0.20 (±0.11)	0.975 (±0.0031)
VGG13	0.984 (±0.0023)	0.23 (±0.10)	0.975 (±0.0038)
VGG16	0.989 (±0.0035)	0.12 (±0.05)	0.982 (±0.0056)
VGG19	0.991 (±0.0030)	0.11 (±0.06)	0.986 (±0.0047)
ResNet16	0.995 (±0.0023)	0.03 (±0.01)	0.991 (±0.0037)
ResNet18	0.883 (±0.2065)	1.39 (±2.60)	0.877 (±0.2052)
ResNet34	0.990 (±0.0071)	0.21 (±0.36)	0.984 (±0.0117)
ResNet-BiGRU-ECA	0.996 (±0.0010)	0.02 (±0.01)	0.994 (±0.0016)

4.5. Effect of BiGRU and ECA Modules

Additional investigations were conducted to provide a more comprehensive assessment of the BiGRU and ECA modules within the proposed ResNet-BiGRU-ECA architecture. As depicted in Table 9, the findings obviously indicate that both the BiGRU module and the ECA module contribute significantly to enhancing the model's efficiency in identifying datasets from two distinct situations.

Table 9. A comparative analysis of ResNet-based models, with and without the bidirectional gated recurrent unit (BiGRU) and efficient channel attention (ECA) modules.

BiGRU	ECA	Accuracy	
		Scenario I	Scenario II
-	-	0.9904 (±0.0116)	0.9924 (±0.0011)
✓	-	0.9965 (±0.0017)	0.9948 (±0.0007)
-	✓	0.9926 (±0.0080)	0.9935 (±0.0045)
✓	✓	0.9980 (±0.0007)	0.9960 (±0.0010)

A checkmark (✓) indicates that the module is included in the given variation of the ResNet model and a dash (-) indicates that it is not included in that ResNet model variation.

5. Conclusions and Future Work

Epilepsy, a neurological disorder, can be significantly mitigated if detected early. This study introduces a novel hybrid DL model named ResNet-BiGRU-ECA, designed to accurately identify epileptic seizures using EEG signals. This model combines residual blocks, an ECA module, and a BiGRU to recognize epileptic seizures in pre-processed multichannel EEG recordings. To gauge the efficiency of our suggested model, we carried out a thorough assessment by contrasting it with five fundamental DL models and leading time-series classification models, all applied to the same publicly available dataset. The DL models underwent training and evaluation using a 5-CV approach. We analyzed model performance using standard evaluation metrics, and the ResNet-BiGRU-ECA model consistently outperformed other models, achieving an average accuracy of 0.998 and an F1-score of 0.998. Furthermore, our model surpassed most existing systems in terms of effectiveness when compared using the same dataset. The objective of this research is to make a meaningful contribution to the field of neurology by investigating the potential benefits of employing EEG data in the context of epilepsy detection and related matters. Our primary goal is to investigate the possibility of reducing examination duration while simultaneously improving diagnostic efficiency and effectiveness.

In future research, we intend to employ the ResNet-BiGRU-ECA model to address other detection issues relying on EEG signals. Additionally, we plan to delve into the

explainability of our model to gain insights into the mechanisms and reasoning behind its accurate decision-making process.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
BiGRU	Bidirectional gated recurrent unit
BiLSTMs	Bidirectional long short-term memory
BiRNN	Bidirectional recurrent neural network
BN	Batch normalization
CNN	Convolutional neural network
Conv1D	1D-convolutional
ConvB	Convolutional block
DL	Deep learning
DT	Decision trees
ECA	Efficient channel attention
ECG	Electrocardiogram
EEG	Electroencephalography
ELU	Exponential linear unit
EMG	Electromyography
ESD	Epileptic seizure detection
ESR	Epileptic seizures recognition
ESRD	Epileptic seizures recognition dataset
FC	Fully connected
GAP	Global average pooling
GB	Gradient boosting
GRU	Gated recurrent unit
HAR	Human activity recognition
(k-CV)	k-fold cross-validation
KNN	k-nearest neighbors
LR	Logistic regression
LSTM	Long short-term memory
ML	Machine learning
MRI	Magnetic resonance imaging
MP	Max-pooling
NB	Naive Bayes
RF	Random forest
RNN	Recurrent neural network
SGDC	Stochastic gradient boosting
UCI	The university of California Irvine

References

1. Siddiqui, M.K.; Morales-Menendez, R.; Huang, X.; Hussain, N. A review of epileptic seizure detection using machine learning classifiers. *Brain Inform.* **2020**, *7*, 1–18. [CrossRef] [PubMed]
2. Sharma, R.; Pachori, R.B.; Sircar, P. Seizures classification based on higher order statistics and deep neural network. *Biomed. Signal Process. Control* **2020**, *59*, 101921. [CrossRef]
3. Fisher, R.S.; Boas, W.V.E.; Blume, W.; Elger, C.; Genton, P.; Lee, P.; Engel, J., Jr. Epileptic Seizures and Epilepsy: Definitions Proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). *Epilepsia* **2005**, *46*, 470–472. [CrossRef] [PubMed]
4. Kim, T.; Nguyen, P.; Pham, N.; Bui, N.; Truong, H.; Ha, S.; Vu, T. Epileptic Seizure Detection and Experimental Treatment: A Review. *Front. Neurol.* **2020**, *11*, 701. [CrossRef]
5. Ayman, U.; Zia, M.S.; Okon, O.D.; Rehman, N.U.; Meraj, T.; Ragab, A.E.; Rauf, H.T. Epileptic Patient Activity Recognition System Using Extreme Learning Machine Method. *Biomedicines* **2023**, *11*, 816. [CrossRef]
6. Mekruksavanich, S.; Phaphan, W.; Hnoohom, N.; Jitpattanakul, A. Attention-Based Hybrid Deep Learning Network for Human Activity Recognition Using WiFi Channel State Information. *Appl. Sci.* **2023**, *13*, 8884. [CrossRef]
7. Shoeibi, A.; Khodatars, M.; Ghassemi, N.; Jafari, M.; Moridian, P.; Alizadehsani, R.; Panahiazar, M.; Khozeimeh, F.; Zare, A.; Hosseini-Nejad, H.; et al. Epileptic Seizures Detection Using Deep Learning Techniques: A Review. *Int. J. Environ. Res. Public Health* **2021**, *18*, 5780. [CrossRef]
8. Shoka, A.; Dessouky, M.; El-Sayed, A.; Hemdan, E.E.D. EEG seizure detection: Concepts, techniques, challenges, and future trends. *Multimed. Tools Appl.* **2023**, *82*, 42021–42051. [CrossRef]
9. Sharma, R.; Pachori, R.B. Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions. *Expert Syst. Appl.* **2015**, *42*, 1106–1117. [CrossRef]
10. Wang, Z.; Yang, J.; Wu, H.; Zhu, J.; Sawan, M. Power efficient refined seizure prediction algorithm based on an enhanced benchmarking. *Sci. Rep.* **2021**, *11*, 23498. [CrossRef]
11. Sheoran, M.; Kumar, S.; Chawla, S. Methods of denoising of electroencephalogram signal: A review. *Int. J. Biomed. Eng. Technol.* **2015**, *18*, 385–395. [CrossRef]
12. Romaine, J.B.; Pereira Martín, M.; Salvador Ortiz, J.R.; Manzano Crespo, J.M. EEG—Single-Channel Envelope Synchronisation and Classification for Seizure Detection and Prediction. *Brain Sci.* **2021**, *11*, 516. [CrossRef] [PubMed]
13. Tang, X.; Zhang, X. Conditional Adversarial Domain Adaptation Neural Network for Motor Imagery EEG Decoding. *Entropy* **2020**, *22*, 96. [CrossRef] [PubMed]
14. Alickovic, E.; Kevric, J.; Subasi, A. Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction. *Biomed. Signal Process. Control* **2018**, *39*, 94–102. [CrossRef]
15. Bou Assi, E.; Nguyen, D.K.; Rihana, S.; Sawan, M. Towards accurate prediction of epileptic seizures: A review. *Biomed. Signal Process. Control* **2017**, *34*, 144–157. [CrossRef]
16. Varone, G.; Boulila, W.; Lo Giudice, M.; Benjdira, B.; Mammone, N.; Ieracitano, C.; Dashtipour, K.; Neri, S.; Gasparini, S.; Morabito, F.C.; et al. A Machine Learning Approach Involving Functional Connectivity Features to Classify Rest-EEG Psychogenic Non-Epileptic Seizures from Healthy Controls. *Sensors* **2022**, *22*, 129. [CrossRef] [PubMed]
17. Craik, A.; He, Y.; Contreras-Vidal, J.L. Deep learning for electroencephalogram (EEG) classification tasks: A review. *J. Neural Eng.* **2019**, *16*, 031001. [CrossRef] [PubMed]
18. Ullah, I.; Hussain, M.; ul Haq Qazi, E.; Aboalsamh, H. An automated system for epilepsy detection using EEG brain signals based on deep learning approach. *Expert Syst. Appl.* **2018**, *107*, 61–71. [CrossRef]
19. Goodfellow, I.J.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016
20. Bengio, Y.; Simard, P.; Frasconi, P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw.* **1994**, *5*, 157–166. [CrossRef]
21. Yu, D.; Deng, L. Deep Learning and Its Applications to Signal and Information Processing [Exploratory DSP]. *IEEE Signal Process. Mag.* **2011**, *28*, 145–154. [CrossRef]
22. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
23. Golmohammadi, M.; Ziyabari, S.; Shah, V.; de Diego, S.L.; Obeid, I.; Picone, J. Deep Architectures for Automated Seizure Detection in Scalp EEGs. *arXiv* **2017**, arXiv:1712.09776.
24. Chen, X.; Ji, J.; Ji, T.; Li, P. Cost-sensitive deep active learning for epileptic seizure detection. In Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, BCB '18, New York, NY, USA, 29 August–1 September 2018; pp. 226–235. [CrossRef]
25. Fukumori, K.; Thu Nguyen, H.T.; Yoshida, N.; Tanaka, T. Fully Data-driven Convolutional Filters with Deep Learning Models for Epileptic Spike Detection. In Proceedings of the ICASSP 2019—IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 2772–2776. [CrossRef]
26. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent advances in convolutional neural networks. *Pattern Recognit.* **2018**, *77*, 354–377. [CrossRef]

27. Andrzejak, R.G.; Lehnertz, K.; Mormann, F.; Rieke, C.; David, P.; Elger, C.E. 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* **2001**, *64*, 061907. [CrossRef] [PubMed]

28. Varone, G.; Boulila, W.; Driss, M.; Kumari, S.; Khan, M.K.; Gadekallu, T.R.; Hussain, A. Finger pinching and imagination classification: A fusion of CNN architectures for IoMT-enabled BCI applications. *Inf. Fusion* **2023**, *101*, 102006. [CrossRef]

29. Van Doorn, J. Analysis of deep convolutional neural Network architectures. In Proceedings of the 21st Twente Student Conference on IT (TScIT 21), Enschede, The Netherlands, 21–24 June 2014.

30. Cho, K.; van Merriënboer, B.; Bahdanau, D.; Bengio, Y. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. In Proceedings of the SSST-8, 8th Workshop on Syntax, Semantics and Structure in Statistical Translation, Doha, Qatar, 25 October 2014; pp. 103–111. [CrossRef]

31. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. In Proceedings of the NIPS 2014 Workshop on Deep Learning, Montreal, QC, Canada, 8–13 December 2014.

32. Schuster, M.; Paliwal, K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681. [CrossRef]

33. Yang, Z.; He, X.; Gao, J.; Deng, L.; Smola, A. Stacked attention networks for image question answering. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Los Alamitos, CA, USA, 27–30 June 2016; pp. 21–29. [CrossRef]

34. Wong, T.T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognit.* **2015**, *48*, 2839–2846. [CrossRef]

35. Bragança, H.; Colonna, J.G.; Oliveira, H.A.B.F.; Souto, E. How Validation Methodology Influences Human Activity Recognition Mobile Systems. *Sensors* **2022**, *22*, 2360. [CrossRef]

36. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2015**, arXiv:cs.CV/1409.1556.

37. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:cs.CV/1512.03385.

38. Han, D.; Kim, J.; Kim, J. Deep Pyramidal Residual Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6307–6315. [CrossRef]

39. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826. [CrossRef]

40. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Los Alamitos, CA, USA, 21–26 July 2017; pp. 1800–1807. [CrossRef]

41. Elghamry, H.; Ghoneim, M.; Abo Haggag, A.; Saeed Darweesh, M.; Ismail, T. Comparative Analysis of Various Machine Learning Techniques for Epileptic Seizures Detection and Prediction Using EEG Data. In Proceedings of the 2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES), Giza, Egypt, 24–26 October 2020; pp. 175–180. [CrossRef]

42. Torse, D.A.; Khanai, R.; Desai, V.V. Classification of Epileptic Seizures using Recurrence Plots and Machine Learning Techniques. In Proceedings of the 2019 International Conference on Communication and Signal Processing (ICCSP), Tamil Nadu, India, 4–6 April 2019; pp. 611–615. [CrossRef]