

## Article

# Seizure Prediction Based on Transformer Using Scalp Electroencephalogram

Jianzhuo Yan, Jinnan Li, Hongxia Xu <sup>\*</sup>, Yongchuan Yu and Tianyu Xu

Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China; yanjianzhuo@bjut.edu.cn (J.Y.); lijinnanjn@emails.bjut.edu.cn (J.L.); yuyongchuan@bjut.edu.cn (Y.Y.); xutianyu970423@163.com (T.X.)

\* Correspondence: xhxcl@bjut.edu.cn; Tel.: +86-131-2036-2671

**Abstract:** Epilepsy is a chronic and recurrent brain dysfunction disease. An acute epileptic attack will interfere with a patient's normal behavior and consciousness, having a great impact on their life. The purpose of this study was to design a seizure prediction model to improve the quality of patients' lives and assist doctors in making diagnostic decisions. This paper presents a transformer-based seizure prediction model. Firstly, the time-frequency characteristics of electroencephalogram (EEG) signals were extracted by short-time Fourier transform (STFT). Secondly, a three transformer tower model was used to fuse and classify the features of the EEG signals. Finally, when combined with the attention mechanism of transformer networks, the EEG signal was processed as a whole, which solves the problem of length limitations in deep learning models. Experiments were conducted with a Children's Hospital Boston and the Massachusetts Institute of Technology database to evaluate the performance of the model. The experimental results show that, compared with previous EEG classification models, our model can enhance the ability to use time, frequency, and channel information from EEG signals to improve the accuracy of seizure prediction.



**Citation:** Yan, J.; Li, J.; Xu, H.; Yu, Y.; Xu, T. Seizure Prediction Based on Transformer Using Scalp Electroencephalogram. *Appl. Sci.* **2022**, *12*, 4158. <https://doi.org/10.3390/app12094158>

Academic Editors: Chunhua Su, Keping Yu, Celestine Iwendu and Thippa Reddy Gadekallu

Received: 11 March 2022

Accepted: 17 April 2022

Published: 20 April 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** transformer; STFT; epilepsy; electroencephalogram; seizure prediction

## 1. Introduction

Epilepsy, a common chronic brain disease, is caused by sudden excessive discharge of brain nerve cells. Nearly 1% of the world's population is suffering from epilepsy. During seizures, patients have transient involuntary convulsions in one part of the body (partial seizures) or the entire body (generalized seizures), sometimes accompanied by a loss of consciousness and urinary and fecal incontinence [1], greatly affecting the patients' life quality. The electroencephalogram (EEG) is a representative signal containing information on brain electrical activity, which is used as a tool for clinical diagnosis and analysis of epilepsy [2].

Since the 1970s, researchers have carried out a lot of research on seizure detection and prediction tasks [3–5]. Automatic epilepsy detection technology can help doctors improve the accuracy of epilepsy diagnosis, greatly save time, enhance the efficiency of diagnosis, and strive for more rescue time. Seizure detection comprises mainly two parts: feature extraction and classification of EEG signals.

In terms of feature extraction, this can be roughly divided into linear analysis methods and nonlinear analysis methods. Linear analysis methods mainly include time-domain analysis, frequency-domain analysis, and time-frequency domain analysis. Early EEG analysis directly extracted features from time-series signals, including time-domain features such as peak value, rhythm, duration, and sharpness [3]. After that, researchers also extracted such characteristics as spike rhythm, the relative amplitude of the EEG signal [4], fractional linear prediction error energy [6], and line length [7] to classify EEG signals. Although the time-domain waveform contains all EEG signals, this method lacks objectivity

with large errors. Studies in the literature [8,9] transformed the original EEG signal from time-domain to frequency-domain and extracted the corresponding spectral components for frequency domain analysis. The frequency domain is very important when dealing with epilepsy. In medicine, the brainwaves can be divided into five main bands in the frequency domain, namely, delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–80 Hz). After analyzing the above medical frequency bands, Perez et al. [10] further divided the beta band into four sub-bands (beta-1, 12–15 Hz; beta-2, 15–18 Hz; beta-3, 18–25 Hz; and hi-beta, 25–30 Hz), which had a positive effect on the analysis of EEG signals. Recent studies have shown that the sub-bands are also useful in the classification of epileptic EEG signals. Tsioura et al. [11] first proposed the systematic evaluation of frequency sub-bands for the classification accuracy of epileptic EEG. The results showed that additional frequency band analysis was conducive to the detection of epileptic EEG. However, EEG signals are non-stationary and time-varying, and the premise of frequency domain analysis is that the stationary random signals and the features extracted in the frequency domain do not contain time information. Therefore, a time-frequency analysis method combining the time domain and frequency domain has been gradually developed. Truong et al. [12] used short-time Fourier transform (STFT) to extract time-frequency information from EEG signals and automatically generate optimized features for each patient. In addition, wavelet transform [13–16] is also an effective tool for time-frequency feature extraction of EEG signals. Studies show that brain activity has complex dynamic characteristics, so it can be regarded as a nonlinear dynamic system. More and more researchers have become interested in nonlinear analysis methods for EEG signals. Li et al. [17] used a multiscale complexity measure to extract the nonlinear feature of the scale-related Lyapunov exponent for classification. Brari [18] proposed a novel EEG feature extraction approach for determining a correlation dimension to analyze the nonlinear characteristics of epileptic EEG signals.

Traditional machine learning algorithms and deep learning models have been successfully applied to the classification of time-series signals, especially EEG signals. In early research, the seizure detection task for EEG signals mainly used traditional machine learning algorithms such as decision tree classifier [19,20], support vector machine [21,22], k-nearest neighbor [23] (KNN), and random forest [24]. Recently, deep learning algorithms have also achieved remarkable results for EEG seizure detection. Because EEG signals are variable and long, recurrent neural networks (RNN), which are appropriate for time series information, have also been gradually applied to EEG signal processing. However, due to the short-term memory problem of RNN, the long short-term memory network (LSTM) [25,26], which combines short-term and long-term memory through a gated structure, has been more widely used in epileptic seizure tasks. In addition, the convolutional neural network (CNN) [27] has the ability of automatic feature learning, which not only simplifies the process of artificially constructing features but also significantly improves the performance of seizure detection.

Now, the frontier research on epilepsy EEG has gradually shifted from epilepsy detection to seizure prediction. Since seizure activity is unpredictable and drug-resistant epilepsy patients lack reliable treatment, there is an urgent need to develop an accurate and reliable seizure prediction model. Truong et al. [12] proposed a generalized retrospective and patient-specific seizure prediction method based on STFT and CNN. After that, Truong et al. [28] added the generative adversarial network as a feature extractor to train in an unsupervised way and connect an external classifier for classification to solve the difficulty of manual marking of seizure data and make better use of key data. Kostas et al. [29] introduced the LSTM network into EEG signal prediction of seizures, which significantly improved its performance. Jee et al. [30] proposed a patient-specific EEG channel selection optimization method based on permutation entropy, which combined the genetic algorithm and KNN to predict seizures. Alshebeili et al. [2] designed a seizure prediction framework based on statistical analysis and digital band-limiting filters, which can make robust decisions on signal activities and is suitable for long prediction horizons. Aung et al. [31]

studied the advantages of multivariate multi-scale modified distribution entropy by using artificial neural networks and proposed an epilepsy prediction system. However, these deep learning models can only achieve very limited performance in long sequences, and the attention mechanism can process the signal as a whole without being limited by the sequence length. Therefore, we consider applying it to our long-term monitoring of EEG signals. In recent years, the transformer model based on the attention mechanism proposed by Vaswani et al. [32] has achieved great success in the field of natural language processing, and it has been gradually applied to the tasks of computer vision and time series prediction and regression. Tao et al. [33] used the transformer to classify EEG data for brain vision and motor imagination to decode EEG signals from human brain activities.

In this paper, scalp EEG is used to predict the onset of epilepsy to provide enough treatment time for doctors and patients and prevent the occurrence of seizure events. The proposed seizure prediction model in this paper is based on transformer networks for extracting and fusing the three-dimensional features of EEG signals and has achieved remarkable results in seizure prediction. The main contributions of this paper are as follows:

1. Based on the transformer's self-attention coding layer and gating mechanism, a transformer network with a three-tower structure is established to extract and fuse the features of epileptic EEG signals from different dimensions, which improves the learning ability of the time, spectrum, and spatial information.
2. A feature engineering scheme for the EEG sequence prediction task is proposed that uses STFT to extract the hidden laws of EEG signals and improve the upper-performance limit of the model.
3. A three-tower transformer network is proposed to deal with the seizure prediction task. The experimental results for a Children's Hospital Boston and the Massachusetts Institute of Technology (CHB-MIT) dataset show that our method is superior to the existing ones.

## 2. Materials and Methods

### 2.1. EEG Database

In this study, a multi-channel scalp EEG database was used for the experiment. The EEG database we used is the CHB-MIT scalp EEG database collected by Children's Hospital Boston. The data, which was originally published in Shoeb's Ph.D. thesis [34] and can be accessed at <https://physionet.org/physiobank/database/chbmit> accessed on 11 April 2022, contains 24 medical records from 23 patients with intractable epilepsy. The first 23 cases were from 22 patients (17 females, aged 1.5–19 years; 5 males, aged 3–22 years. Chb01 and Chb21 were obtained from the same female subject at an interval of 1.5 years, and Chb24 had no clear gender or age recorded). Table 1 shows the basic characteristics of the database. The Children's Hospital Boston assessed the potential conditions for surgical intervention after all the epileptic patients stopped antiepileptic drugs for a period of time and monitored the patients for several days. Scalp EEG signals were sampled at 256 Hz with 16-bit resolution, and the position of electrode placement followed the international 10–20 electrode positions system. The total duration of EEG recordings was nearly 983 h, with 198 seizures.

### 2.2. Preprocessing

Epileptic EEG signals can be divided into three stages, namely, interictal, preictal, and ictal. In our seizure prediction task, we mainly focused on the preictal and interictal periods. Therefore, we discarded the ictal EEG segments and transformed seizure prediction into a binary classification problem. At the same time, we rearranged the EEG data for each patient according to the electrode order. In this dataset, the interictal and preictal data were severely unbalanced since some patients had fewer seizures during the monitoring period. In order to overcome this problem, we used overlapping sliding windows in the training

stage to obtain more preictal fragments. To ensure that the ratio of the two types of training data was close to 1, we set the window size of preictal to  $W_p$  as follows:

$$W_p = \frac{n_p}{n_i} W \quad (1)$$

where  $n_p$  and  $n_i$  represent the number of EEG segments in each patient's preictal and interictal stage, respectively, and  $W$  represents the size of the EEG window.

**Table 1.** Details of the CHB-MIT<sup>1</sup> database.

Case	Gender	Age (Years)	Number of Channels	Number of Seizures
Chb01	F	11	23	7
Chb02	M	11	23	3
Chb03	F	14	23	7
Chb04	M	22	23	4
Chb05	F	7	23	5
Chb06	F	1.5	23	10
Chb07	F	14.5	23	3
Chb08	M	3.5	23	5
Chb09	F	10	23	4
Chb10	M	3	23	7
Chb11	F	12	23	3
Chb12	F	2	18	40
Chb13	F	3	23	12
Chb14	F	9	18	8
Chb15	M	16	18	20
Chb16	F	7	18	10
Chb17	F	12	18	3
Chb18	F	18	18	6
Chb19	F	19	23	3
Chb20	F	6	23	8
Chb21	F	13	23	4
Chb22	F	9	23	3
Chb23	F	6	23	7
Chb24	-	-	23	16

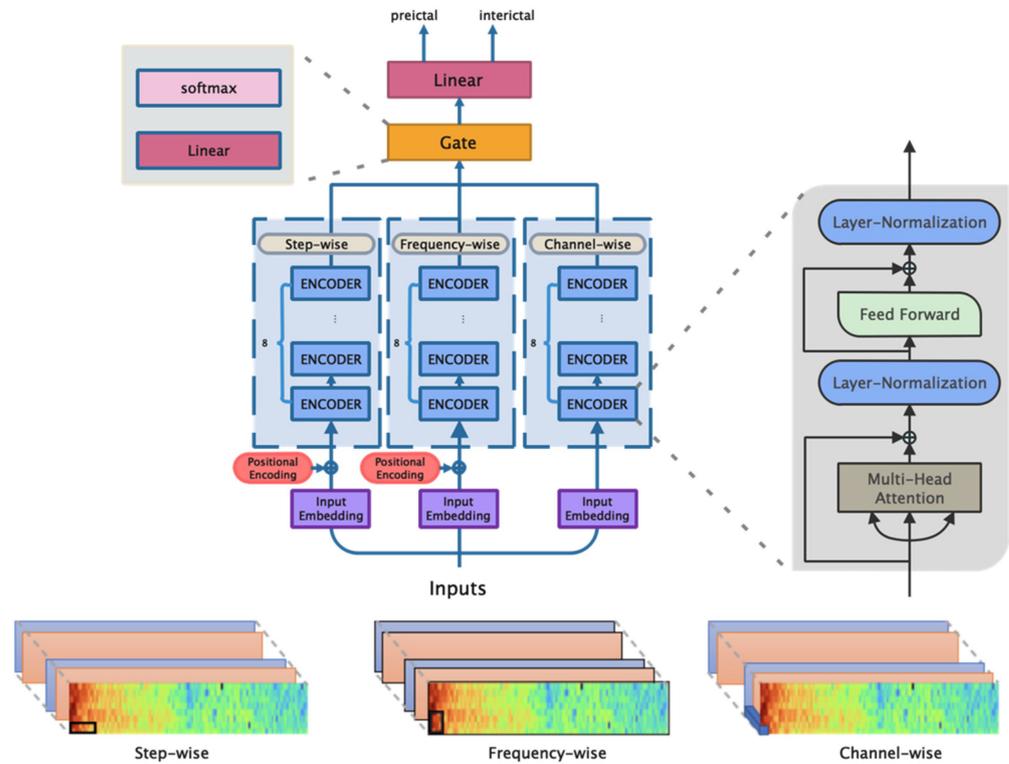
<sup>1</sup> CHB-MIT: Children's Hospital Boston and the Massachusetts Institute of Technology.

STFT has been widely used in the field of signal processing, and many studies have proved that it has advantages in the analysis of time series [35,36]. Therefore, in this paper, STFT is used to convert the EEG signal into a two-dimensional matrix composed of the time domain and frequency domain, and the EEG signal is analyzed in the time-frequency domain. We selected the cosine analysis window to perform the STFT on the 5 s sample, and then used  $\log_{10}$  to calculate the intensity value. The EEG recordings in the CHB-MIT dataset were contaminated by 60 Hz power line noise. Therefore, in the experiment, we removed the components in the frequency range of 57–63 Hz, 117–123 Hz (power line frequency is 60 Hz), and the direct current (DC) component (0 Hz) so as to conveniently and effectively remove the interference of power line noise and DC component. After STFT processing, the dimension of each 5S EEG segment was (c, 114, 9), where c is the number of EEG channels, and 114 and 9 represent the number of frequency points and time steps, respectively.

### 2.3. Classification

Because of its good performance in natural language processing, transformer has been applied to computer vision, time series classification, and prediction tasks. In recent years, it has also shown good effects in the classification of epilepsy EEG [33,37]. EEG signals contain abundant time, frequency, and spatial information. In this paper, we used transformer networks to analyze and classify the characteristics of epileptic EEG signals in

three dimensions. The overall architecture of the network model proposed in this paper is shown in Figure 1.



**Figure 1.** The model architecture of three-tower transformer model.

The traditional transformer is used in the processing of text sequence, and the text data (L, D) is generally two-dimensional, where L and D represents the length and dimension of the word vector, respectively. However, after time-frequency domain transformation, the EEG segment has a three-dimensional matrix (C, F, T), in which C, F, and T represent the channel, frequency, and time step of the EEG signal, respectively. We flatten the 3D matrix  $x \in \mathbb{R}^{C \times F \times T}$  from three different dimensions into two dimensions to obtain  $x_c \in \mathbb{R}^{C \times (F \cdot T)}$ ,  $x_f \in \mathbb{R}^{F \times (C \cdot T)}$ ,  $x_s \in \mathbb{R}^{T \times (C \cdot F)}$ . These three matrices are taken as the three inputs for the model. Because the three inputs are continuous, the embedding layer of the model is replaced by a full connection layer. To the best of our knowledge, the order of each channel in the EEG sequence has no absolute or relative correlation, so we only added positional encoding on the basis of frequency-wise and step-wise input embeddings. Then, three sets of encoders were used to capture the correlation of each dimension sequence from the aspects of step, frequency, and channel information. Each layer of the encoder has two operations, namely feed forward and multi-head attention composed of multiple self-attention. The inputs of the self-attention mechanism are Q (query), K (key), and V (values), and the calculation formula for its output matrix is shown as follows:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$

The multi-head attention mechanism further improves the self-attention layer, expands the ability of the model to focus on different positions, and gives multiple “representation subspaces” in the attention layer, which is expressed as follows:

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_m), where : head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \tag{3}$$

where  $m$  is the number of attention heads, and  $W_i^Q$ ,  $W_i^K$  and  $W_i^V$  are the learned projection matrices.

The output matrix of the multi-head attention layer is transferred into a feed-forward neural network to enhance feature extraction. Then, the model adds a gating mechanism to integrate the characteristics of time, frequency, and channel direction. We set the outputs of the three towers as  $C$ ,  $S$  and  $F$ , connected them into vectors, obtained  $H$  through the linear projection layer, and then assigned gating weights  $g_1$ ,  $g_2$ , and  $g_3$  to each output through the softmax function. Finally, the weight of each gate corresponds to the output of the corresponding tower, and the eigenvector  $y$  is obtained through the following formula.

$$H = W \cdot \text{Concat}(C, S, F) + b, \quad (4)$$

$$g_1, g_2, g_3 = \text{Softmax}(H), \quad (5)$$

$$y = \text{Concat}(C \cdot g_1, S \cdot g_2, F \cdot g_3) \quad (6)$$

Finally, the feature vector  $y$  passes through a linear full connection layer and changes it into a vector with dimensions (Batch\_size,2), and the classification results of the EEG segments are then obtained.

#### 2.4. Performance Evaluation

In the experiment, the prediction effect was achieved by learning and classifying the EEG data characteristics of preictal and interictal segments. In order to evaluate the prediction effect of the prediction system, we introduce some performance evaluation indexes. In this paper, *Accuracy*, *Sensitivity*, *Specificity*, *Precision*, *Recall*, and *F1-Score* are used as the evaluation indexes for the model. The calculation formulae are as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}, \quad (7)$$

$$\text{Sensitivity} = \text{Recall} = \frac{TP}{TP + FN}, \quad (8)$$

$$\text{Specificity} = \frac{TN}{TN + FP}, \quad (9)$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (10)$$

$$F1 - \text{score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}, \quad (11)$$

where the true positive ( $TP$ ) and true negative ( $TN$ ) are correctly classified as preictal and interictal EEG segments, respectively. False positives ( $FP$ ) and false negatives ( $FN$ ) indicated that they were incorrectly predicted as preictal and interictal EEG segments.

In real life, however, it is necessary to warn patients and doctors in advance of the impending seizure, so that the medical doctors can be prepared to properly manage the episode. In order to evaluate the performance of the seizure prediction model, we introduced seizure prediction horizon (SPH) and seizure occurrence period (SOP). SOP is defined as the time-period for predicting seizures, while SPH refers to the time-period from the alarm to the beginning of SOP, that is, the period of clinical intervention. The successful prediction of epilepsy means that seizures must occur after SPH and during SOP (Figure 2). If there is a seizure during SPH or no seizure in SOP, it is considered a false alarm. In clinical use, the SOP should not be set too long, otherwise it will increase the anxiety of patients and cause mental stress. The setting of SPH should provide doctors with enough time for clinical interventions [38]. In order to define an appropriate period for SPH and SOP, researchers have studied both. According to the survey in [39], the optimal warning time is 3–5 min. Therefore, we considered setting the SPH to 3–5 min. Nesaei [40] proposed that SPH + SOP should be more than 10 min and less than 90 min to provide treatment for

patients and avoid undesirable anxiety. Furthermore, the SOP in many seizure prediction studies [12,38,41] is generally 30 min. Based on the above considerations, SPH and SOP in this paper are set to 3 min and 30 min, respectively, that is, when the model gives a correct alarm, the patient's seizure should occur between 3 min and 33 min later. However, the SOP and SPH are usually unknown clinically, and researchers usually chose values based on assumptions [41]. Studies have shown that the electrical changes that occur in the brain before seizures are difficult to capture with the human eye [42]. Furthermore, due to the specificity of seizures, the length of the pre-onset period will vary from a few minutes to a few hours [43]. This may cause our hypothesis for pre-seizure to deviate from the ground truth, resulting in the wrong label for individual EEG segments that do not have the characteristics of typical pre-seizure, thus affecting the accuracy of the prediction results.



**Figure 2.** The definition of seizure prediction horizon (SPH) and seizure occurrence period (SOP).

The model we proposed is primarily designed to distinguish between preictal and interictal EEG segments, where sporadic false positives during interictal periods are common and false alarms will appear. The false positive rate (FPR), defined as the number of false alarms per hour, is an important index for evaluating the prediction of seizures. In order to reduce the false-positive rate, we further post-processed the classification results of the model. In this paper, we adopt the k-of-n method put forward by [12] to post-process the seizure prediction task. In the process of our experiment, we set  $n = 30$  and  $k = 24$ , that is, in the prediction of 30 consecutive EEG segments, at least 24 segments are predicted to be positive before the alarm will be sent out, and the whole process is regarded as an epilepsy prediction process.

### 3. Results

For the purpose of verifying the effectiveness of the proposed model, we conducted corresponding tests on the CHB-MIT dataset. In the experiment, the interictal phase was defined as the period 4 h before the seizure and 4 h after the end of the seizure. The division of epileptic EEG signals is shown in Figure 3. According to the annotation file for the database, in 24 cases some EEG recording electrodes added or deleted EEG channels during the measurement process. Therefore, we selected the EEG data from the following 18 electrodes that were considered in most cases for analysis: FP1–F7, F7–T7, T7–P7, P7–O1, FP1–F3, F3–C3, C3–P3, P3–O1, FP2–F4, F4–C4, C4–P4, P4–O2, FP2–F8, F8–T8, T8–P8, P8–O2, FZ–CZ, and CZ–PZ. The letter notations are as follows: FP, frontopolar; F, frontal; T, temporal; O, occipital; C, central; P, parietal (Figure 4). In addition, because the channels for cases 12 and 13 changed frequently during the recording process, the EEG recordings may have been polluted, and case 24 had frequent seizures, so we did not have enough interictal data for training. Therefore, in the experiment, we removed the EEG data for these three cases and applied the data for the remaining 21 patients to evaluate our model.

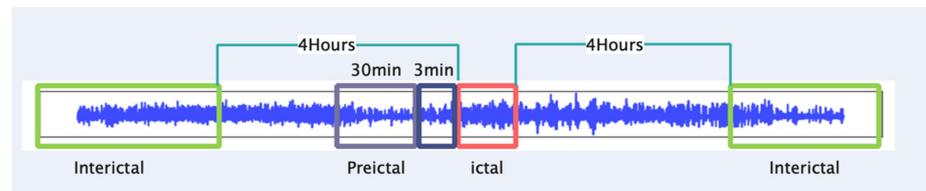


Figure 3. Division of EEG signals in the different stages of epilepsy.

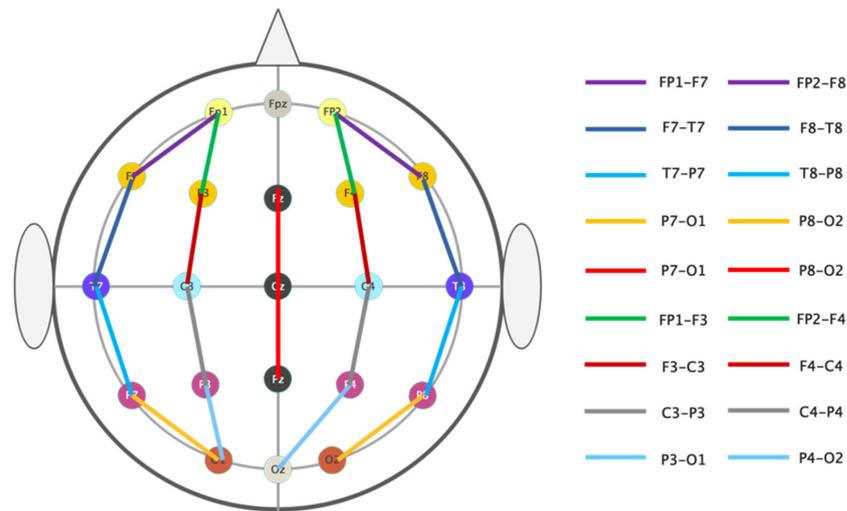


Figure 4. The electrode distribution of EEG signals.

The workflow of the seizure prediction system is shown in Figure 5. First, the EEG signals in the preictal and interictal segments were extracted and labeled. Second, overlapping sliding windows were used to balance the number of EEG segments in each patient’s preictal and interictal segments. Third, the time-frequency domain characteristic information from the above EEG segments was extracted by STFT to obtain the three-dimensional spectrogram matrixes. The number of preictal and interictal spectrograms is summarized in Table 2. Fourth, the feature matrix was inputted to the transformer network model with a three-tower structure. The characteristics of epileptic EEG signals are learned from the step-wise, frequency-wise, and channel-wise encoders, and the classification of EEG signals is realized through the gating mechanism. Finally, the seizure prediction task is finished after post-processing.

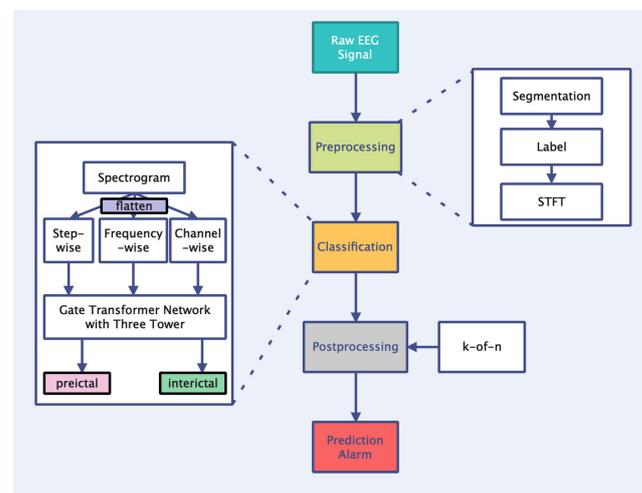


Figure 5. Workflow of the seizure prediction system.

**Table 2.** The summary of the number of spectrograms for preictal and interictal periods.

	Preictal	Interictal
Subject No.	21	21
No. of channels	18	18
No. of spectrograms	312,552	329,437

We divided the EEG segments from all patients into a training set and a test set at a ratio of 9:1. The model proposed in this paper was built in a Python 3.7 environment using PyTorch 1.9, and the code is available at <https://github.com/xutianyu540/Transformer-network-with-a-three-tower-structure> accessed on 11 April 2022. At the same time, the gated transformer networks (GTN) model [44] was established for comparative experiments. The classification results from the final experiments are shown in Table 3.

**Table 3.** The comparison between the performance of the GTN<sup>1</sup> and our model.

Patient	No. of Seizures	GTN <sup>1</sup>			Our Model		
		Sen <sup>2</sup> (%)	Spec <sup>3</sup> (%)	Pre <sup>4</sup> (%)	Sen <sup>2</sup> (%)	Spec <sup>3</sup> (%)	Pre <sup>4</sup> (%)
Chb01	7	98.62	98.99	98.96	100	99.88	99.9
Chb02	3	92.61	94.56	94.49	96.88	98.64	98.58
Chb03	7	90.65	91.01	89.87	97.82	96.87	94.31
Chb04	4	91.25	87.94	88.28	93.53	92.15	92.23
Chb05	5	92.34	90.40	90.6	94	99.45	98.95
Chb06	10	91.26	87.93	88.29	93.51	92.1	92.18
Chb07	3	89.04	93.99	93.73	99.11	97.12	97.2
Chb08	5	94.44	91.72	91.89	97.22	95.86	95.89
Chb09	4	87.59	92	91.71	94.53	97.15	97.1
Chb10	7	95.65	90.33	90.53	96.48	98.2	98.11
Chb11	3	89.5	86.25	81.14	97.48	96.39	94.69
Chb14	6	98.59	97.88	97.89	99.29	99.65	99.65
Chb15	13	91.60	88.87	86.67	93.22	97.43	96.63
Chb16	5	85.85	88	87.74	89.85	92.31	92.11
Chb17	3	93.19	95.12	95.04	98.54	97.24	97.28
Chb18	5	94.25	96.06	95.35	98.24	99.67	99.61
Chb19	3	93.56	91.59	89.05	95.56	94.21	92.35
Chb20	6	95.20	94.58	94.43	98.64	97.55	97.49
Chb21	4	89.58	88.63	88.81	93.50	91.55	91.77
Chb22	3	91.20	91.21	90.35	95.53	96.21	95.79
Chb23	5	88.25	89.01	88.81	93.33	91.21	91.29
Average	-	92.11	91.72	91.12	96.01	96.23	95.86

<sup>1</sup> GTN, gated transformer networks; <sup>2</sup> Sen, sensitivity; <sup>3</sup> Spec, specificity; <sup>4</sup> Pre, precision.

In Table 3, we compare the sensitivity, specificity, and precision of our model with GTN. It can be seen that for all patients, the average sensitivity and specificity of our model were 96.01% and 96.23%, respectively, which were significantly higher than the corresponding values of 92.11% and 91.72% for the GTN model. This indicates that our model achieved satisfactory results in the classification of both preictal and interictal states. Compared with the GTN model, the average precision of our method improved from 91.12% to 95.86%. Figure 6 shows a comparison of the classification accuracy and F1 score for each patient between the GTN model and our model. It can be seen that our model was greatly improved in accuracy and F1 score. The results show that the proposed model improves the utilization of temporal, spectral, and spatial information of EEG signals, and has a better effect on the classification of epileptic EEG signals.

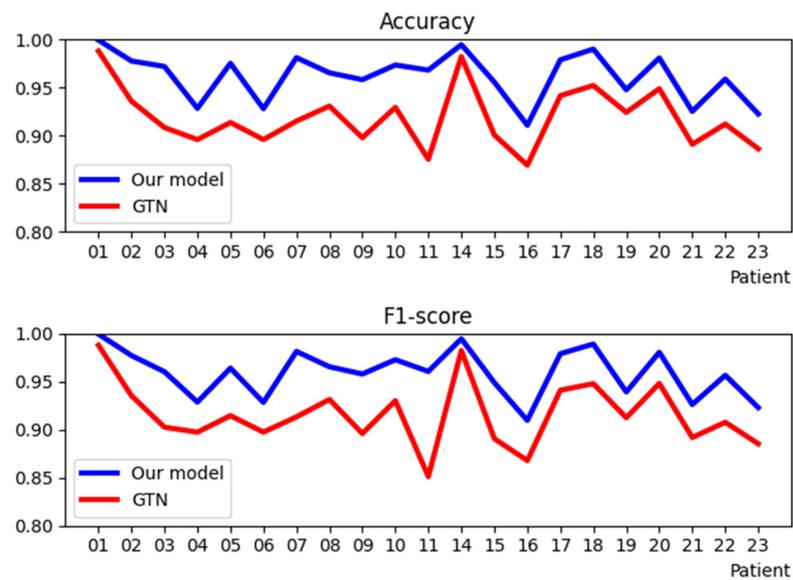


Figure 6. The accuracy and the F1 score of the GTN and our model for 21 patients.

The classification results were subsequently post-processed to obtain the prediction results from our seizure prediction model. The FPR classified for our network alone and after post-processing is shown in Table 4. It can be seen that after post-processing, our FPR decreased, on average, from 0.19/h to 0.047/h. This effectively eliminated sporadic errors from the process of model classification and improved the performance of our epileptic seizure prediction system so as to better complete the prediction task.

Table 4. The comparison between the performance of the GTN and our model.

Patient	No. of Seizures	Without Post-Processing	After Post-Processing
		FPR <sup>1</sup> (h <sup>-1</sup> )	FPR <sup>1</sup> (h <sup>-1</sup> )
Chb01	7	0	0
Chb02	3	0.07	0.01
Chb03	7	0.3	0.07
Chb04	4	0.9	0.22
Chb05	5	0.02	0
Chb06	10	0.52	0.13
Chb07	3	0.3	0.07
Chb08	5	0.02	0
Chb09	4	0.3	0.08
Chb10	7	0.1	0.02
Chb11	3	0.1	0.02
Chb14	6	0	0
Chb15	13	0.05	0
Chb16	5	0.1	0.03
Chb17	3	0.07	0.02
Chb18	5	0.02	0
Chb19	3	0.3	0.08
Chb20	6	0.1	0.03
Chb21	4	0.45	0.11
Chb22	3	0.1	0.03
Chb23	5	0.3	0.07
Average	-	0.19	0.047

<sup>1</sup> FPR: False Prediction Rate.

#### 4. Discussion

In traditional methods, researchers usually use manually constructed features to extract the time domain, frequency domain, and nonlinear features of EEG signals. We proposed a new method that can better use temporal, spectral, and spatial information for extraction. First, the spectrum of 18 channel EEG segments is obtained by STFT, and then, the time, frequency and channel information from the spectrum are analyzed by the step-wise, frequency-wise, and channel-wise encoders in the transformer tower, respectively. When combined with the gated unit, the features of time, frequency, and spatial direction are fused to realize the classification of EEG signals. The analysis results show that our model has better performance and more accurate classification results.

The CHB-MIT dataset used in this study has also been used for performance evaluation in other studies. Truong et al. [12] combined STFT and CNN to automatically extract and classify the time-frequency features of EEG signals, and in this instance, the prediction sensitivity was 81.2% and FPR was 0.16/h. Rukhsar et al. [45] extracted eight time-based features and predicted seizures through multivariate statistical process control (MSPC), and obtained 88.89% sensitivity and 0.39/h FPR. Xu et al. [46] proposed an end-to-end deep learning solution based on CNN. The overall sensitivity and FPR for scalp EEG data were 98.8% and 0.074/h, respectively. Tang et al. [47] proposed a novel framework of multi-view convolutional gated recurrent network (Mv-CGRN) and embedded the attention mechanism in Mv-CGRN, and determined the best feature combination of each patient by adaptively adjusting the weight parameters, achieving an average sensitivity of 94.50% and an average FPR of 0.118/h. Zhao et al. [48] used a binary single-dimensional convolutional neural network (BSDCNN) to predict seizures with a sensitivity of 88.89% and an FPR of 0.39 per hour. Zhang et al. [49] used a simple CNN model to classify the correlation matrix obtained by calculating the Pearson correlation coefficient to distinguish the preictal states from the interictal ones and obtained a sensitivity of 92.9%. Table 5 shows a comparison of the results for our method with the classification algorithms from the above literature. From the comparative results, our method does not require the complex process of manually constructing features and it has a lower FPR and higher sensitivity. Although the CNN network proposed by [46] had higher sensitivity, this study only evaluated EEG data of seven patients without good generalization. Therefore, from the perspective of overall performance, our model is more effective at the epilepsy prediction task.

**Table 5.** The comparison of the results of this work with other reported methods.

Year	Authors	Feature	Classifier	No. of Patients	No. of Seizures	Sen <sup>1</sup> (%)	FPR (h <sup>-1</sup> )	SPH
2018	Truong	STFT	CNN	13	64	81.2	0.16	5 min
2019	Rukhsar	Eight temporal-based features	MSPC	10	90	88.89	0.39	-
2020	Xu	-	CNN	7	45	98.8	0.074	5 min
2020	Tang	L-DFA, relative band energy, synchronization modularity	Mv-CGRNi	24	140	94.5	0.118	-
2020	Zhao	-	BSDCNN	6	-	94.69	0.095	5 min
2021	Zhang	Pearson correlation coefficient matrices.	CNN	19	-	92.9	-	15 min
-	This work	STFT	Three tower transformer	21	111	96.01	0.047	3 min

<sup>1</sup> SPH, seizure prediction horizon.

In this work, the proposed model has several limitations. (1) After the feature information is fused by the gated unit, the model only uses the full connection layer to classify it, which may lead to redundant parameters and insufficient expression of spatial structure. (2) Since different patients have specific seizures, their physiological preictal periods are different. In this paper, if we set SPH as a fixed value, the defined preictal segments and the prototypical physiological preictal signature (the ground truth) do not match completely [41], which will lead to some training samples being labeled incorrectly. In the future, we will further optimize the network structure and explore more accurate classifiers to classify the characteristic information of an epileptic EEG. In addition, we also need to analyze the EEG signals of patients in the preictal period to set a more appropriate SPH value.

## 5. Conclusions

An effective seizure prediction method will not only better help doctors diagnose and reduce the pain of patients, it can also help them avoid dangerous activities such as driving or swimming before the onset of seizures. In this paper, the time domain, frequency domain, and channel information of EEG signals were fused, combined with a gating unit and transformer model to classify EEG signals for the prediction of patients' seizures. The prediction sensitivity and FPR of our model were 96.01% and 0.047/h, respectively, which achieves effective classification of epileptic EEG signals and has good seizure prediction performance. In the future work, we will further optimize our network performance and apply it to other datasets to achieve a more effective seizure prediction system.

**Author Contributions:** Conceptualization, J.Y. and H.X.; methodology, H.X., J.L. and T.X.; software, J.L. and T.X.; validation, J.L.; formal analysis, J.Y.; investigation, J.Y.; resources, J.Y., Y.Y. and H.X.; writing—original draft preparation, J.L.; writing—review and editing, J.L., T.X. and H.X.; visualization, J.Y.; supervision, Y.Y. and H.X.; project administration, J.Y. and H.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Beijing Natural Science Foundation grant number 4222022. The APC was funded by the Beijing Natural Science Foundation.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are openly available in <https://physionet.org/physiobank/database/chbmit> (accessed on 10 September 2021).

**Acknowledgments:** The authors would like to thank the anonymous reviewers for their valuable comments and suggestions, which helped improve this paper greatly.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. WHO. WHO Report. 2018. Available online: <http://www.who.int/mediacentre/factsheets/fs999/en/> (accessed on 10 February 2022).
2. Alshebeili, S.A.; Sedik, A.; Abd El-Rahiem, B.; Alotaiby, T.N.; El Banby, G.M.; El-Khobby, H.A.; Ali, M.A.A.; Khalaf, A.A.M.; Abd El-Samie, F.E. Inspection of EEG signals for efficient seizure prediction. *Appl. Acoust.* **2020**, *166*, 107327. [CrossRef]
3. Gotman, J. Automatic recognition of epileptic seizures in the EEG. *Electroencephalogr. Clin. Neurophysiol.* **1982**, *54*, 530–540. [CrossRef]
4. Srinivasan, V.; Eswaran, C. Artificial neural network based epileptic detection using time-domain and frequency-domain features. *J. Med. Syst.* **2005**, *29*, 647–660. [CrossRef] [PubMed]
5. Winterhalder, M.; Maiwald, T.; Voss, H.; Aschenbrenner-Scheibe, R.; Timmer, J.; Schulze-Bonhage, A. The seizure prediction characteristic: A general framework to assess and compare seizure prediction methods. *Epilepsy Behav.* **2003**, *4*, 318–325. [CrossRef]
6. Joshi, V.; Pachori, R.B.; Vijesh, A. Classification of ictal and seizure-free EEG signals using fractional linear prediction. *Biomed. Signal Processing Control* **2014**, *9*, 1–5. [CrossRef]
7. Pathak, A.; Ramesh, A.; Mitra, A.; Majumdar, K. Automatic seizure detection by modified line length and Mahalanobis distance function. *Biomed. Signal Processing Control* **2018**, *44*, 279–287. [CrossRef]

8. Jahankhani, P.; Revett, K.; Kodogiannis, V. Data mining an EEG dataset with an emphasis on dimensionality reduction. In Proceedings of the 2007 IEEE Symposium on Computational Intelligence and Data Mining, Honolulu, HI, USA, 1 March–5 April 2007; pp. 405–412.
9. Kaleem, M.; Guergachi, A.; Krishnan, S. Patient-specific seizure detection in long-term EEG using wavelet decomposition. *Biomed. Signal Processing Control* **2018**, *46*, 157–165. [[CrossRef](#)]
10. Pérez-Elvira, R.; Ultra-Cucarella, J.; Carrobes, J.A.; Teodoru, M.; Bacila, C.; Neamtu, B. Individual alpha peak frequency, an important biomarker for live z-score training neurofeedback in adolescents with learning disabilities. *Brain Sci.* **2021**, *11*, 167. [[CrossRef](#)]
11. Tsipouras, M.G. Spectral information of EEG signals with respect to epilepsy classification. *EURASIP J. Adv. Signal Processing* **2019**, *2019*, 1–17. [[CrossRef](#)]
12. Truong, N.D.; Nguyen, A.D.; Kuhlmann, L.; Bonyadi, M.R.; Yang, J.; Ippolito, S.; Kavehei, O. Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram. *Neural Netw.* **2018**, *105*, 104–111. [[CrossRef](#)]
13. Ghosh-Dastidar, S.; Adeli, H.; Dadmehr, N. Mixed-band wavelet-chaos-neural network methodology for epilepsy and epileptic seizure detection. *IEEE Trans. Biomed. Eng.* **2007**, *54*, 1545–1551. [[CrossRef](#)] [[PubMed](#)]
14. Vidyaratne, L.S.; Iftekharuddin, K.M. Real-time epileptic seizure detection using EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2017**, *25*, 2146–2156. [[CrossRef](#)] [[PubMed](#)]
15. Orosco, L.; Correa, A.G.; Diez, P.; Laciari, E. Patient non-specific algorithm for seizures detection in scalp EEG. *Comput. Biol. Med.* **2016**, *71*, 128–134. [[CrossRef](#)] [[PubMed](#)]
16. Chandel, G.; Upadhyaya, P.; Farooq, O.; Khan, Y. Detection of seizure event and its onset/offset using orthonormal triadic wavelet based features. *IRBM* **2019**, *40*, 103–112. [[CrossRef](#)]
17. Li, Q.; Gao, J.; Huang, Q.; Wu, Y.; Xu, B. Distinguishing Epileptiform Discharges From Normal Electroencephalograms Using Scale-Dependent Lyapunov Exponent. *Front. Bioeng. Biotechnol.* **2020**, *8*, 1006. [[CrossRef](#)]
18. Brari, Z.; Belghith, S. A novel Machine Learning approach for epilepsy diagnosis using EEG signals based on Correlation Dimension. *IFAC-PapersOnLine* **2021**, *54*, 7–11. [[CrossRef](#)]
19. Polat, K.; Güneş, S. Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. *Appl. Math. Comput.* **2007**, *187*, 1017–1026. [[CrossRef](#)]
20. Arunkumar, N.; Ram Kumar, K.; Venkataraman, V. Automatic detection of epileptic seizures using permutation entropy, Tsallis entropy and Kolmogorov complexity. *J. Med. Imaging Health Inform.* **2016**, *6*, 526–531. [[CrossRef](#)]
21. Nicolaou, N.; Georgiou, J. Detection of epileptic electroencephalogram based on permutation entropy and support vector machines. *Expert Syst. Appl.* **2012**, *39*, 202–209. [[CrossRef](#)]
22. Li, K.; Zhang, X.; Du, Y. A SVM based classification of EEG for predicting the movement intent of human body. In Proceedings of the 2013 10th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), Jeju, Korea, 30 October–2 November 2013; pp. 402–406.
23. Wang, S.; Chaovalitwongse, W.A.; Wong, S. Online seizure prediction using an adaptive learning approach. *IEEE Trans. Knowl. Data Eng.* **2013**, *25*, 2854–2866. [[CrossRef](#)]
24. Wang, G.; Deng, Z.; Choi, K.-S. Detection of epileptic seizures in EEG signals with rule-based interpretation by random forest approach. In *International Conference on Intelligent Computing*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 738–744.
25. Hussein, R.; Palangi, H.; Ward, R.K.; Wang, Z.J. Optimized deep neural network architecture for robust detection of epileptic seizures using EEG signals. *Clin. Neurophysiol.* **2019**, *130*, 25–37. [[CrossRef](#)] [[PubMed](#)]
26. Hu, X.; Yuan, S.; Xu, F.; Leng, Y.; Yuan, K.; Yuan, Q. Scalp EEG classification using deep Bi-LSTM network for seizure detection. *Comput. Biol. Med.* **2020**, *124*, 103919. [[CrossRef](#)] [[PubMed](#)]
27. Wei, Z.; Zou, J.; Zhang, J.; Xu, J. Automatic epileptic EEG detection using convolutional neural network with improvements in time-domain. *Biomed. Signal Processing Control* **2019**, *53*, 101551. [[CrossRef](#)]
28. Truong, N.D.; Kuhlmann, L.; Bonyadi, M.R.; Querlioz, D.; Zhou, L.; Kavehei, O. Epileptic seizure forecasting with generative adversarial networks. *IEEE Access* **2019**, *7*, 143999–144009. [[CrossRef](#)]
29. Tsiouris, K.M.; Pezoulas, V.C.; Zervakis, M.; Konitsiotis, S.; Koutsouris, D.D.; Fotiadis, D.I. A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals. *Comput. Biol. Med.* **2018**, *99*, 24–37. [[CrossRef](#)]
30. Ra, J.S.; Li, T.; Li, Y. A Novel Permutation Entropy-Based EEG Channel Selection for Improving Epileptic Seizure Prediction. *Sensors* **2021**, *21*, 7972. [[CrossRef](#)]
31. Aung, S.T.; Wongsawat, Y. Prediction of epileptic seizures based on multivariate multiscale modified-distribution entropy. *PeerJ Comput. Sci.* **2021**, *7*, e744. [[CrossRef](#)]
32. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Processing Syst.* **2017**, *30*, 5999–6009.
33. Tao, Y.; Sun, T.; Muhamed, A.; Genc, S.; Jackson, D.; Arsanjani, A.; Yaddanapudi, S.; Li, L.; Kumar, P. Gated Transformer for Decoding Human Brain EEG Signals. In Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Guadalajara, Mexico, 1–5 November 2021; pp. 125–130.
34. Shoeb, A.H. *Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment*; Massachusetts Institute of Technology: Cambridge, MA, USA, 2009.

35. Kovacs, P.; Samiee, K.; Gabbouj, M. On application of rational discrete short time fourier transform in epileptic seizure classification. In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014; pp. 5839–5843.
36. Tsai, A.-C.; Luh, J.-J.; Lin, T.-T. A novel STFT-ranking feature of multi-channel EMG for motion pattern recognition. *Expert Syst. Appl.* **2015**, *42*, 3327–3341. [[CrossRef](#)]
37. Chen, R.; Parhi, K.K. Seizure Prediction using Convolutional Neural Networks and Sequence Transformer Networks. In Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Guadalajara, Mexico, 1–5 November 2021; pp. 6483–6486.
38. Sun, B.A.; Lv, J.J.; Rui, L.G.; Yang, Y.X.; Chen, Y.G.; Ma, C.; Gao, Z.K. Seizure prediction in scalp EEG based channel attention dual-input convolutional neural network. *Phys. A-Stat. Mech. Its Appl.* **2021**, *584*, 126376. [[CrossRef](#)]
39. Fourth International Workshop on Seizure Prediction. Available online: <http://www.iwsp4.org/patients.htm> (accessed on 5 April 2022).
40. Nesaei, S.; Sharafat, A.R. Real-time mining of epileptic seizure precursors via nonlinear mapping and dissimilarity features. *IET Signal Processing* **2015**, *9*, 193–200. [[CrossRef](#)]
41. 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](#)] [[PubMed](#)]
42. Li, Z.; Fields, M.; Panov, F.; Ghatan, S.; Yener, B.; Marcuse, L. Deep Learning of Simultaneous Intracranial and Scalp EEG for Prediction, Detection, and Lateralization of Mesial Temporal Lobe Seizures. *Front. Neurol.* **2021**, *12*, 705119. [[CrossRef](#)] [[PubMed](#)]
43. Bandarabadi, M.; Rasekhi, J.; Teixeira, C.A.; Karami, M.R.; Dourado, A. On the proper selection of preictal period for seizure prediction. *Epilepsy Behav.* **2015**, *46*, 158–166. [[CrossRef](#)]
44. Liu, M.; Ren, S.; Ma, S.; Jiao, J.; Chen, Y.; Wang, Z.; Song, W. Gated Transformer Networks for Multivariate Time Series Classification. *arXiv* **2021**, arXiv:2103.14438.
45. Rukhsar, S.; Khan, Y.; Farooq, O.; Sarfraz, M.; Khan, A. Patient-specific epileptic seizure prediction in long-term scalp EEG signal using multivariate statistical process control. *IRBM* **2019**, *40*, 320–331. [[CrossRef](#)]
46. Xu, Y.; Yang, J.; Zhao, S.; Wu, H.; Sawan, M. An end-to-end deep learning approach for epileptic seizure prediction. In Proceedings of the 2020 2nd IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS), Genova, Italy, 31 August–2 September 2020; pp. 266–270.
47. Tang, L.; Xie, N.; Zhao, M.; Wu, X. Seizure Prediction Using Multi-View Features and Improved Convolutional Gated Recurrent Network. *IEEE Access* **2020**, *8*, 172352–172361. [[CrossRef](#)]
48. Zhao, S.; Yang, J.; Xu, Y.; Sawan, M. Binary single-dimensional convolutional neural network for seizure prediction. In Proceedings of the 2020 IEEE International Symposium on Circuits and Systems (ISCAS), Seville, Spain, 12–14 October 2020; pp. 1–5.
49. Zhang, S.; Chen, D.; Ranjan, R.; Ke, H.; Tang, Y.; Zomaya, A.Y. A lightweight solution to epileptic seizure prediction based on EEG synchronization measurement. *J. Supercomput.* **2021**, *77*, 3914–3932. [[CrossRef](#)]