

Article

Deep-Learning-Based Automated Anomaly Detection of EEGs in Intensive Care Units

Jacky Chung-Hao Wu ^{1,†}, Nien-Chen Liao ^{2,3,4,†}, Ta-Hsin Yang ¹, Chen-Cheng Hsieh ¹, Jin-An Huang ^{3,5}, Yen-Wei Pai ^{3,6}, Yi-Jhen Huang ², Chieh-Liang Wu ^{2,6,*} and Henry Horng-Shing Lu ^{1,7,*}

- ¹ Institute of Statistics, National Yang Ming Chiao Tung University, Hsinchu 300093, Taiwan; chwu@nycu.edu.tw (J.C.-H.W.); z85021212@gmail.com (T.-H.Y.); astanley414@gmail.com (C.-C.H.)
- ² Department of Critical Care Medicine, Taichung Veterans General Hospital, Taichung 407219, Taiwan; rufus0822@gmail.com (N.-C.L.); encourage1009@vghtc.gov.tw (Y.-J.H.)
- ³ Department of Neurology, Neurological Institute, Taichung Veterans General Hospital, Taichung 407219, Taiwan; jahuang2639@gmail.com (J.-A.H.); jasminezxcv@vghtc.gov.tw (Y.-W.P.)
- ⁴ Institute of Clinical Medicine, National Yang Ming Chiao Tung University, Taipei 112304, Taiwan
- ⁵ Department of Health Business Administration, Hungkuang University, Taichung 433304, Taiwan
- ⁶ Department of Post-Baccalaureate Medicine, College of Medicine, National Chung Hsing University, Taichung 402202, Taiwan
- ⁷ Department of Statistics and Data Science, Cornell University, Ithaca, NY 14853, USA
- * Correspondence: cljeff.wu@gmail.com (C.-L.W.); henryhslu@nycu.edu.tw (H.H.-S.L.)
- † These authors contributed equally to this work.

Abstract: An intensive care unit (ICU) is a special ward in the hospital for patients who require intensive care. It is equipped with many instruments monitoring patients' vital signs and supported by the medical staff. However, continuous monitoring demands a massive workload of medical care. To ease the burden, we aim to develop an automatic detection model to monitor when brain anomalies occur. In this study, we focus on electroencephalography (EEG), which monitors the brain electroactivity of patients continuously. It is mainly for the diagnosis of brain malfunction. We propose the gated-recurrent-unit-based (GRU-based) model for detecting brain anomalies; it predicts whether the spike or sharp wave happens within a short time window. Based on the banana montage setting, the proposed model exploits characteristics of multiple channels simultaneously to detect anomalies. It is trained, validated, and tested on separated EEG data and achieves more than 90% testing performance on sensitivity, specificity, and balanced accuracy. The proposed anomaly detection model detects the existence of a spike or sharp wave precisely; it will notify the ICU medical staff, who can provide immediate follow-up treatment. Consequently, it can reduce the medical workload in the ICU significantly.

Keywords: anomaly detection; EEG; GRU; ICU; intensive care unit; spike



Citation: Wu, J.C.-H.; Liao, N.-C.; Yang, T.-H.; Hsieh, C.-C.; Huang, J.-A.; Pai, Y.-W.; Huang, Y.-J.; Wu, C.-L.; Lu, H.H.-S. Deep-Learning-Based Automated Anomaly Detection of EEGs in Intensive Care Units. *Bioengineering* **2024**, *11*, 421. <https://doi.org/10.3390/bioengineering11050421>

Academic Editors: Gianluca Zaza and Crescenzo Gallo

Received: 27 March 2024

Revised: 20 April 2024

Accepted: 23 April 2024

Published: 25 April 2024



Copyright: © 2024 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/>).

1. Introduction

There are many causes of unconscious patients in the intensive care unit (ICU). Figuring out the reason behind it has always been a tricky process. Whether it is blood testing, brain computed tomography, or even MRI, they are all tools that are often used for differential diagnosis. Still, these tools can only represent the current situation, the situation of a point. Continuous electroencephalography (EEG) monitoring is essential for more in-depth tracking of constant changes. Early detection and early treatment are very important milestones in the medical field. We try to use different methods of framing data to achieve the purpose of anomaly detection. According to research statistics for the intensive care unit, 8–37% of patients have had a non-convulsive seizure [1]. Delayed diagnosis or treatment of non-convulsive seizures is associated with a high death rate. Between 10 and 67% of non-convulsive seizures may go undetected without continuous EEG monitoring,

and 56% of non-convulsive seizures will be detected within the first hour with continuous EEG monitoring. It has been monitored that 88% of non-convulsive seizures are seen within 24 h. In particular, continuous EEG monitors can detect non-convulsive seizures in patients early, assist physicians in the timely detection of the brain or neurological changes, and provide patients with immediate treatment to prevent permanent damage, which is an essential tool for clinicians in diagnosing disease. Interpreting a large number of EEGs is a very labor-intensive task. The importance of continuous EEG monitoring lies in the early diagnosis of epilepsy and can significantly reduce complications and mortality. About 30% of ICU patients with impaired consciousness have epilepsy, and 90% of them have non-convulsive epilepsy, and only EEGs can make a diagnosis. When the burden of epilepsy is heavier, it means that the damage to the brain will continue to increase over time, which will lead to aggravating epilepsy changes and form a vicious circle. Therefore, immediate treatment becomes very important. In the case of a large amount of continuous EEG data, the clinical side cannot load such a large amount of EEG interpretation. Therefore, the aided performance of artificial intelligence and deep learning is a good choice and development goal.

In this study, we focus on the EEGs characters that are highly related to epilepsy. Using the patients' EEG data in the past and the doctors' interpretation and marking can teach the machine to quickly identify abnormal EEGs in subsequent continuous EEG monitoring, thereby improving the efficiency of diagnosis and reducing the burden on the clinical side. The novelty and contributions of this study can be summarized as follows.

- The automated anomaly detection targets patients who are heavily ill and taken care of intensively in the intensive care unit of the Taichung Veterans General Hospital (TCVGH), a national-level medical center, not patients taking some routine and/or physical examinations.
- We attempt to detect anomaly brainwaves before their occurrence so that we can consider possible follow-up treatments in advance. The developed early detection models have promising performance and show great potential in clinical applications.

2. Related Works

EEGs have been used to conduct various kinds of research works [2]. The problem of sleep stage classification is studied to help the diagnosis of sleep disorders [3,4] and to measure sleep quality [5]. Some researchers study how to perform automatic emotion recognition [6,7]. The investigations of EEG motor imagery signals have proliferated due to the great potential in brain–computer interface applications [8,9]. The evaluation of mental workload is studied for maintaining working performance and preventing accidents [10,11]. Some researchers have attempted to solve the problem of automatic detection of epileptic seizures, which can be used to improve the patient's life quality [12], and some have focused on the task of event-related potential detection [13,14].

There has been a large amount of spike detection methods published in the literature [15,16]. The released methods are mainly divided into the following: mimetic analysis [17,18], template matching [19], power spectral analysis [18,20], wavelet analysis [21], and artificial neural networks (ANNs) [22–24]. The features obtained from the above methods are seen as input of the methods. In some methods mentioned above, they use their data to fit a classifier. In clinical application, spikes and sharp waves have the same clinical performance when these events happen. As a result, they may be seen as the same class in some papers. Due to the quick development in Graphics Processing Units (GPUs) and Compute Unified Device Architecture (CUDA), a software layer gives direct access to the GPU's virtual instruction set and parallel computational elements, for the execution of compute kernels [25], and most of the current methods are mainly based on deep learning models [26–29]. Beyond the task of spike detection for EEGs, the framework of deep learning and transfer learning is now dominating the domain of healthcare in the diagnosis of various diseases and for solving many biomedical problems [30,31]. Applications include, but are not limited to, the automated detection of mycobacterium tuberculosis [32], per-

sonalized medicine with electronic health record (EHR) data [33], diagnosis of ophthalmic diseases [34], drug discovery [35], and gene expression classification [36].

Finally, we compare our work with existing work. Due to the differences in the data acquisition and experimental setting, such as the sampling rate and the filter band, it is hard to provide a fair comparison. However, we manage to provide the qualitative comparison in Table 1. In a nutshell, our work targets patients who are heavily ill and taken care of intensively in the ICU of the national-level medical center, TCVGH, not patients taking some routine and/or physical examinations, and it achieves comparable performance with the finest time resolution (i.e., the window size). In addition, we develop early detection models that have promising performance and show great potential in clinical applications, while cross-institutional validation should be conducted in the future to further support and expand the impact.

Table 1. Comparison of existing works.

Objective	Method	ICU Patients	Montage	Window Size	Performance
transient event classification [18]	mimetic analysis, power spectral analysis	No	bipolar	355 ms	87.38% accuracy
spike detection [19]	template matching	No	average reference, bipolar	5.12 s	92.6% selectivity
IED detection [21]	wavelet analysis	No	average reference	3 s	90.5% accuracy
epileptic activity classification [23]	artificial neural network	No	bipolar	355 ms	84.48% accuracy
spike detection [28]	deep learning	No	average reference	0.5 s	0.947 AUC
spike detection (ours)	deep learning	Yes	bipolar	160 ms	94.66% balanced accuracy

3. Materials and Methods

3.1. Working Flow

In this study, the experimental subjects are ICU patients. We focus on patients’ EEG characteristics. Since patients were admitted to ICUs for various reasons, close and intensive care was performed. Many of them suffered from persistent conscious disturbance even though some common causes such as hemodynamic instability and electrolyte imbalance were ruled out. After consulting neurologists, EEG examinations were performed. The working process of this study is shown in Figure 1. Taichung Veterans General Hospital (TCVGH) collected and provided retrospective EEG data recorded from the ICU patients, which are used to train the deep learning models. Medical doctors in TCVGH marked the patients’ EEGs with their or their family’s permission.

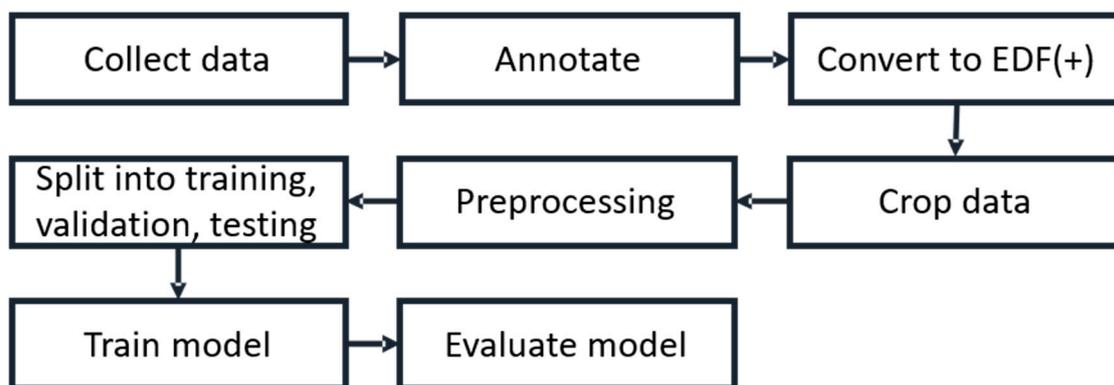


Figure 1. The working process performed in this study.

3.2. ICU Data

Electroencephalography (EEG) is a method used to record an electrogram of the discharge of electrodes attached to the scalp. It has become widely accepted for recording activity below the surface of the brain. Because the electrodes are placed along the scalp according to certain methods, such as the International 10–20 system, it is not invasive typically. EEG measures voltage fluctuations resulting from ionic currents within the neurons of the brain. Clinically, EEG refers to the recording of the brain's spontaneous electrical activity over a period of time, as recorded from multiple electrodes placed on the scalp [37]. Clinical applications usually focus on either event-related potentials or the spectral content of EEG. The former investigates potential fluctuation changes at the moment of the event, such as “eyes open” or “stimulus onset”. The latter analyzes the type of neural oscillations, which is popularly referred to as “brainwaves”, that can be observed in EEG signals in the frequency domain. EEG is widely used in many clinical applications, including sleep disorders, depth of anesthesia, coma, encephalopathies, and brain death. But it is most often used to diagnose epilepsy, which causes abnormalities in EEG readings [38]. Nowadays, EEG mainly uses disc electrodes in clinical practice, and the electrodes are placed according to the International 10–20 system, including 19 recording electrodes and 3 reference electrodes, as shown in Figure 2. Among them, 10 and 20 refer to 10% and 20% of the distance from the nasion to the inion. Fp represents pre-frontal, F represents frontal, C represents central, O represents occipital, and T represents temporal. Also, the mastoid process after ears, A1 and A2, is defined as reference electrodes. The main advantage of using the International 10–20 system is that it can identify the same relative position on the scalp regardless of the size of the head.

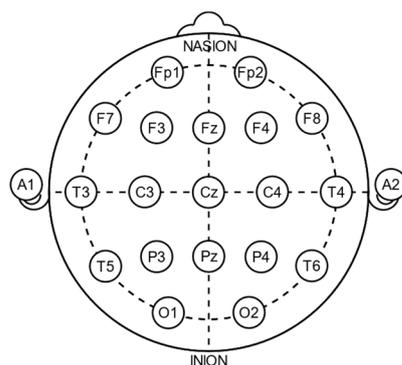


Figure 2. The International 10–20 system for EEG recording. トマトン124, Public domain, via Wikimedia Commons [39].

EEG data in this study were collected and provided by TCVGH. We use European Data Format (EDF) as the data format to store the EEG recording of ICU patients. EDF is a standard file format designed for the exchange and storage of medical time series [40]. Being an open and non-proprietary format, EDF(+) is commonly used to archive, exchange, and analyze data from commercial devices in a format that is independent of the acquisition system. In this way, the data can be retrieved and analyzed by independent software. EDF(+) software (browsers, checkers, etc.) and example files are freely available [41]. Neurologists put annotations in EDF(+) files. We use the Python language and the MNE package to read EDF(+) and retrieve the information, such as when and where anomaly brainwaves occur. In this study, a total of 8 ICU patients' records are used to conduct the experiments.

As a general rule, modern montages allow for easy visualization of comparable scalp areas, so they may be assessed for symmetry [42]. There are two primary montages: bipolar montage and monopolar montage. For epilepsy brainwaves, bipolar montage is a better choice to observe its anomaly situation. Medical doctors of TCVGH provided EEG data with bipolar montages which consist of chains of electrodes, each one connected to

two neighboring electrodes. The bipolar montage is also called the “banana montage”. Its transverse montage links adjacent electrodes in a chain like two bananas, as shown in Figure 3.

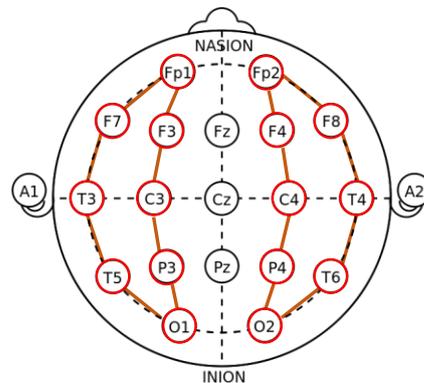


Figure 3. The chains of the bipolar montage. Red circles denote the electrodes and orange lines represent the neighboring relationship between electrodes in clinical practice.

In this study, the main detection targets are spikes and sharp waves, which are typical epilepsy abnormal brainwaves. In the bipolar montage, these two types of abnormal brainwaves have similar characteristics. Surges occur in adjacent channels. The peaks of the surges point tip to tip, and thus it can be identified at a glance. The only difference is that the two occur for different durations. Typically, spikes occur in 20 to 70 milliseconds and sharp waves occur in 70 to 200 milliseconds, as shown in Figure 4 [43].

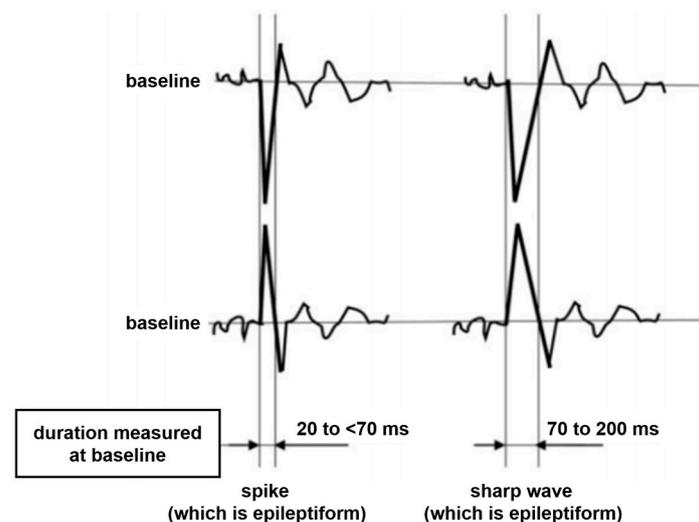


Figure 4. The patterns of the spike and the sharp wave in the bipolar montage.

3.3. Preprocessing

The sampling rate of all EDF(+) files provided by TCVGH medical doctors is not consistent. To make sure that every EDF(+) is at the same sampling rate, we check if the EDF(+) is at the most common sampling rate, which corresponds to 125 Hz in this study. If not, the sampling rate of EDF(+) data is downsampled to 125 Hz. The downsampling process consists of low-pass finite impulse response filtering followed by a sub-selecting mechanism. The difference between two adjacent timesteps is 8 milliseconds. Due to the short period of spikes and sharp waves, we divide the original EDF(+) data into 20 timesteps per sample which correspond to 160 milliseconds in a time window.

When a sample contains the annotation provided by the TCVGH medical doctors, it is treated as an anomaly brainwave. Otherwise, it is marked as a normal brainwave. In

the left figure of Figure 5, because the medical doctors did not annotate any labels in this window, we treat the samples gained from this period as normal brainwaves. In contrast, in the right figure of Figure 5, the medical doctors thought there was a spike in this period, so they marked the time point with the red box. Now, we get samples with two different classes. Next, we split all samples into three datasets: the training set, the validation set, and the testing set.

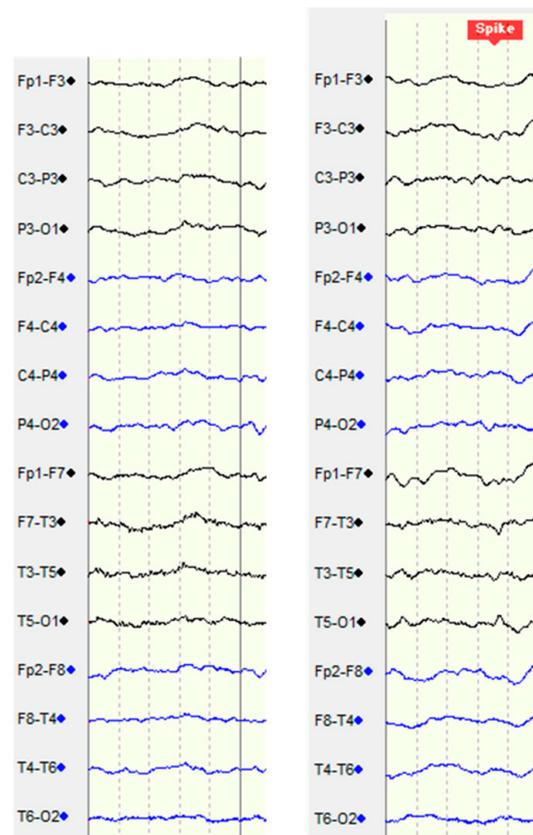


Figure 5. Example brainwaves with and without annotations provided by the TCVGH medical doctors. Black curves are the brainwaves collected from the left brain, whereas blue curves are those collected from the right brain.

Some periods of EEG data may vibrate violently, and data from previous and subsequent time periods will not be on the same scale. We deal with this problem by normalizing every channel data individually to ensure that the values of each channel are on the same scale without losing the information about numerical level differences. We re-scale linearly the values of all channels to the range between zero and one by shifting the minimum and maximum values to zero and one, respectively. Consequently, the values of every sample data are on the same scale.

To maintain the same proportion as the original proportion of the number of different classes before dividing the data into training, validation, and testing sets, we divide each class into three sets in the same proportion individually and then merge them together, as shown in Figure 6. We first divide the original data into pseudo-training and testing sets according to the ratio of 8:2. Then, we further divide the pseudo-training set corresponding to 80% of the total data into the training and validation sets according to the ratio of 8:2. As a result, the training set accounts for 64%, the validation set accounts for 16%, and the testing set accounts for 20%.

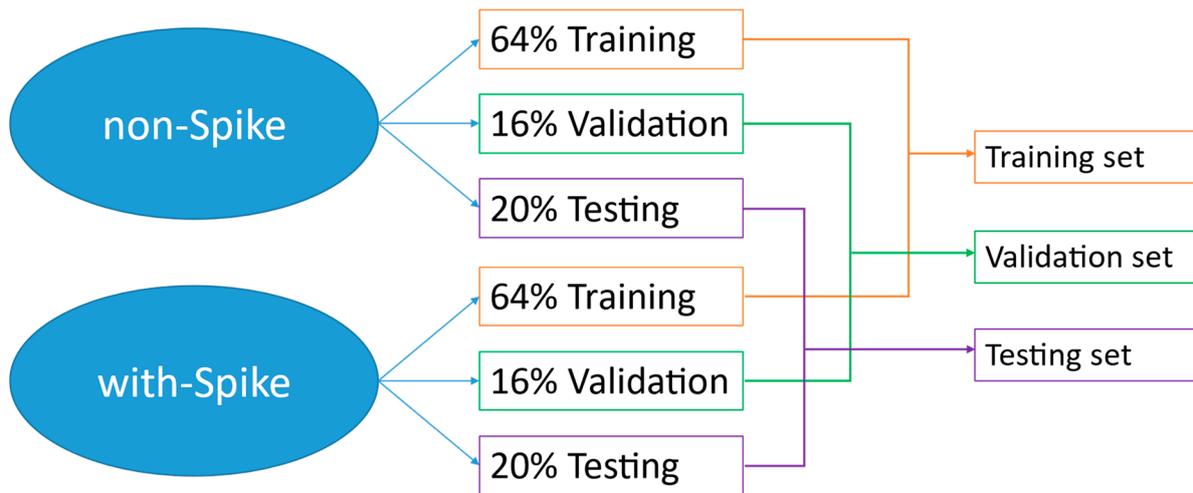


Figure 6. The process of dividing the data into training, validation, and testing sets to maintain the same proportion.

3.4. Sampling Method

We crop the samples from the EEG recordings. There are 20 timesteps in one window, which corresponds to 160 milliseconds. If we discard some samples, the data cannot reflect the true situation of patients in ICUs. Thus, we crop and keep all the samples from 0 to the last second, as shown in Figure 7.

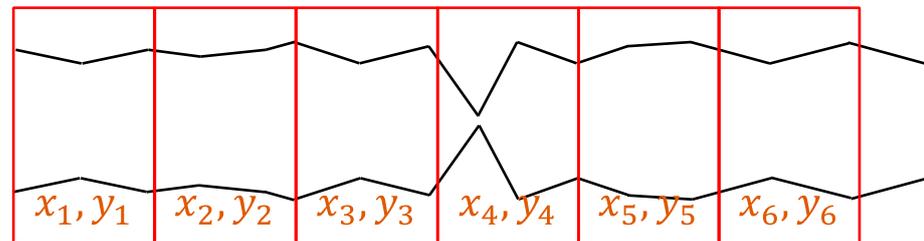


Figure 7. The cropping method. The EEG recording is partitioned consecutively into samples of 160 milliseconds without overlapping from the start to the end.

3.5. GRU-Based Model Architecture

Because of the time dependency of EEG data, we adopt gated recurrent units (GRUs) in the proposed model to detect anomaly brainwaves in EEG recordings. GRU-based models are a kind of recurrent neural network and are particularly suited for performing predictions for time series data. The proposed model architecture is depicted in Figure 8 and will be the same for all experiments conducted in this study.

To be more specific, the operations of the proposed GRU-based model in the inference stage are unrolled and shown in Figure 9, where x_t is the normalized 16-channel EEG recordings of the t -th timestep in one window. $h_{t-1}^{(k)}$ corresponds to the hidden states of the GRU layer k for the input x_t that store the sequence information of the EEG recordings up to the $(t - 1)$ -th timestep. Each GRU layer consists of 64 GRU units involving the update and reset gating mechanisms that capture long- and short-term dependencies in EEG recordings. At the end of the 20th timestep, which corresponds to the last timestep in one window, the hidden states of the GRU layer 2, $h_{20}^{(2)}$, are used as the input of the fully connected layers with the ReLU activation functions to produce the final output prediction.

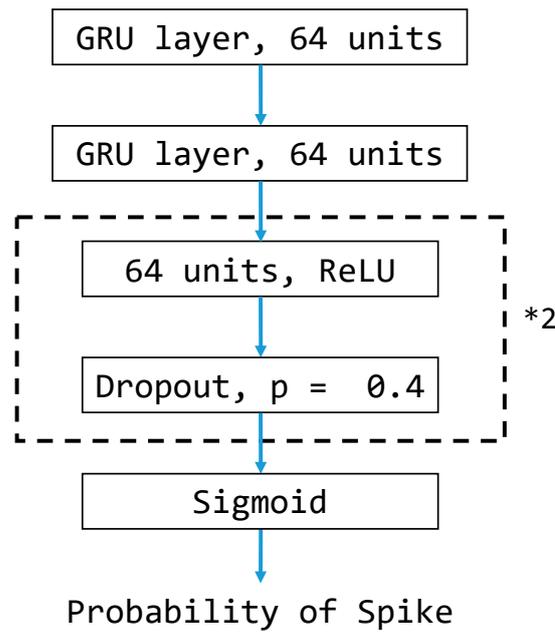


Figure 8. Proposed GRU-based model architecture. *2 means that the layers in the black dotted box are repeated twice in a cascade fashion in the model structure.

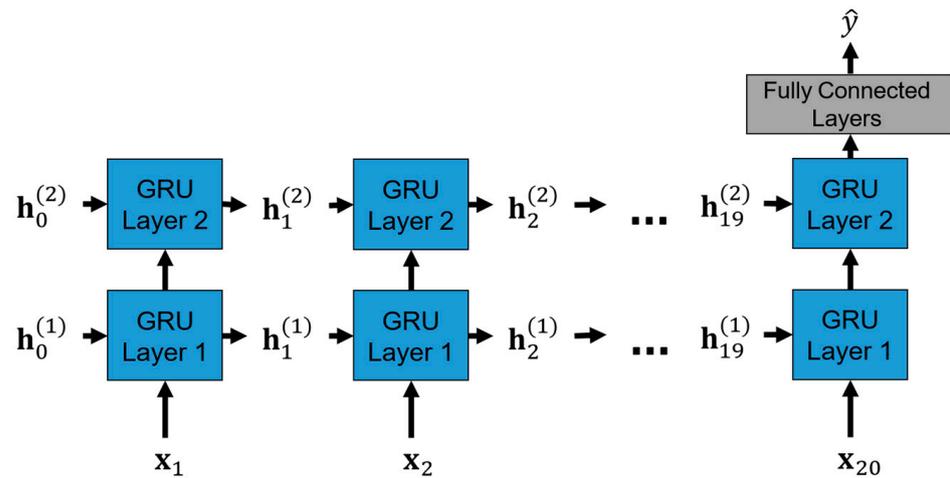


Figure 9. The unrolled operations of the proposed GRU-based model in the inference stage.

3.6. CNN-Based Model Architecture

Because of convolutional neural networks’ (CNNs’) powerful ability to extract features from every sample, we also test CNN-based models by treating the cropped brainwave samples as images. Thus, we construct the CNN-based model for the classification task. The architecture of the CNN-based models is depicted in Figure 10 and will be the same for all experiments conducted in this study. To be more specific, we treat each cropped brainwave sample as a 16×20 grayscale image fed into the CNN-based model. Each CNN layer consists of 64 convolutional filters that perform feature extraction to capture the local correlation within small patches and is followed by a batch normalization layer to provide suitable rescaling. The resulting features of the second CNN layer are used as the input of the fully connected layers with the ReLU activation functions to produce the final output prediction.

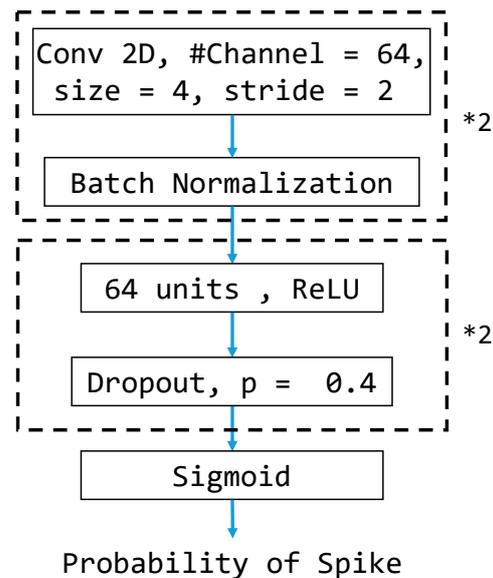


Figure 10. CNN-based model architecture for comparisons. *2 means that the layers in the corresponding black dotted box are repeated twice in a cascade fashion in the model structure.

3.7. Class Weight

Adjusting the class weight in the training stage is a critical step in reducing the influence of the imbalance of the data. If the data are imbalanced, the models focus on the class with a larger amount. Models pay less attention to the class with a smaller amount. To reduce the influence, we adjust the class weight in the training stage. The class weight of each class is disproportionate to its amount so that models can pay attention to the pattern of both classes equally,

$$\text{weight}_i = \frac{\# \text{ of total training data}}{\# \text{ of training data from class } i}$$

where weight_i is the class weight assigned to class i during the model training.

3.8. Performance Metrics

In the stage of model training, we will use the validation set to choose the best model by monitoring the prediction performance after every training epoch. Since the detection of anomaly brainwaves is treated as a binary classification task, we may pick the epoch with the highest validation accuracy and retrieve the corresponding model as the final model. Due to the type of task, the accuracy of the model is the most important metric in our study. In the following section, we will use some metrics to quantify the performance of the deep learning models we build.

The confusion matrix, as shown in Table 2, can be used to provide the details of prediction results by the model. (1) defines the accuracy, which is the proportion of samples that are predicted correctly by the model. In medical applications, it shows the proportion of patients who are diagnosed with correct health status. (2) defines the sensitivity, which is the proportion of positive samples that are predicted as positive by the model and is an indicator to avoid a false negative. In medical applications, it shows the proportion of sick patients who are diagnosed with the disease. (3) defines the specificity, which is the proportion of negative samples that are predicted as negative by the model and is an indicator to avoid a false positive. In medical applications, it shows the proportion of people without the disease who are not diagnosed with it. (4) defines the balanced accuracy, which is the arithmetic mean of the sensitivity and specificity. Since the data are highly imbalanced

between classes in this study, sensitivity, specificity, and balanced accuracy (BA) are more representative than accuracy for the performance evaluation of models.

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \tag{1}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{Specificity} = \frac{TN}{FP + TN} \tag{3}$$

$$\text{Balanced Accuracy (BA)} = \frac{\text{Sensitivity} + \text{Specificity}}{2} \tag{4}$$

Table 2. Confusion matrix for performance evaluation of models.

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

4. Experiments and Results

This section illustrates experiments under different situations. We will use two different kinds of models with the same model complexity. After training, we pick the models corresponding to the highest validation accuracy or the highest validation balanced accuracy of the epoch and perform with the testing set to validate the model performance. We also attempt to achieve early detection in this study.

For all experiments, we use the same setting. Adam is adopted as the optimizer with a learning rate of 10^{-4} . The batch size is 512, and the maximum number of epochs is 500. The cross-entropy loss with the adjusting class weights is used to guide the model training. All experiments are conducted within the TensorFlow framework.

4.1. Experiment 1

There are two EEG recordings containing a few spike annotations with a timeline error. Thus, we treat all samples from the two with a timeline error as negative in this experiment. Figures 11 and 12 show the training curves of GRU-based models without and with adjusted class weights, respectively. Figures 13 and 14 show the training curves of CNN-based models without and with adjusted class weights, respectively. All four figures show that the respective model learns well. The models are selected at different epochs indicated by the red points by monitoring the validation accuracy or the validation balanced accuracy.

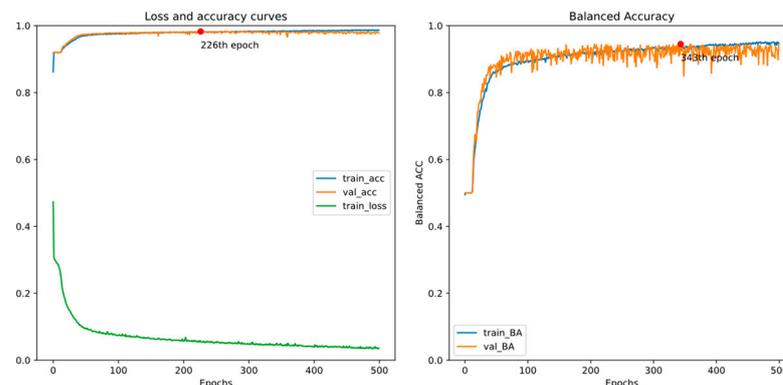


Figure 11. Training curves of the GRU-based model without adjusted class weights.

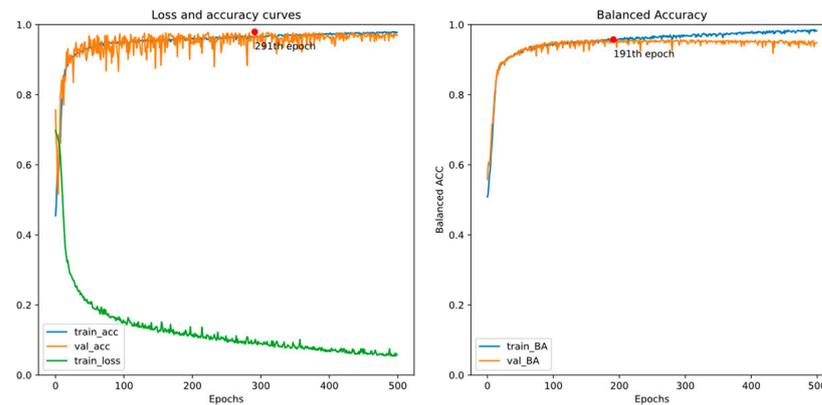


Figure 12. Training curves of the GRU-based model with adjusted class weights.

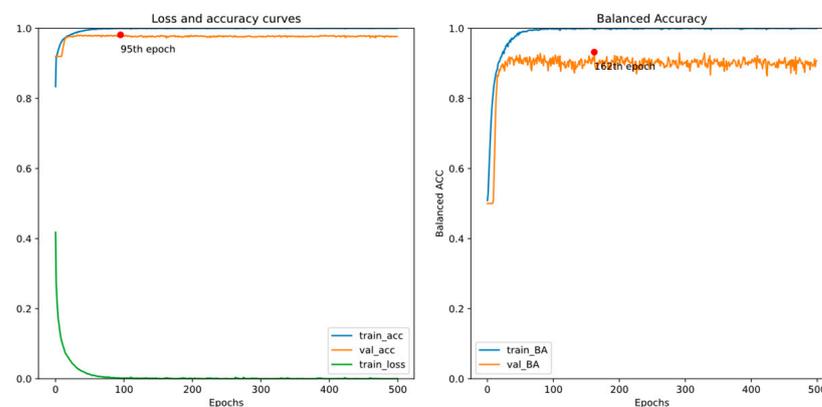


Figure 13. Training curves of the CNN-based model without adjusted class weights.

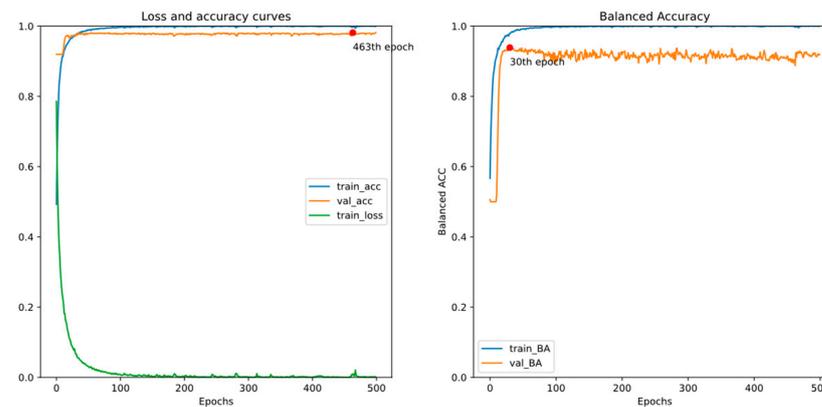


Figure 14. Training curve of the CNN-based model with adjusted class weights.

Table 3 illustrates the performance of all models in this experiment. If we choose the final model by monitoring the validation accuracy, the sensitivity value is less than 90%. In addition, by monitoring validation balanced accuracy (BA), accuracy and specificity reduce a little, but sensitivity and BA, the metrics of interest in our study, may increase. Also, we can see that adjusting class weights can improve models' performances and make them perform more stably in every metric.

Table 3. Comparison of the testing performance (%) in Experiment 1.

	Model	By	Accuracy	Sensitivity	Specificity	BA
No Class Weight	GRU	Acc	98.19	85.24	99.33	92.28
		BA	97.32	91.54	97.83	94.68
	CNN	Acc	97.93	81.30	99.38	90.34
		BA	97.21	87.80	98.04	92.92
With Class Weight	GRU	Acc	97.83	87.99	98.69	93.34
		BA	95.95	93.11	96.19	94.65
	CNN	Acc	97.61	81.10	99.05	90.08
		BA	97.15	90.16	97.76	93.96
Zero-rule Baseline			91.95	0.00	100.00	50.00

4.2. Experiment 2

In Experiment 1, we treat all the samples cropped from the two recordings with time-line errors as negative. There are still some annotations that are not problematic in these two recordings, so we correct the samples corresponding to these annotations manually to the positive class. The sizes of training, validation, and testing set samples may change slightly. Eleven samples are fixed. The comparison of resulting model performances is shown in Table 4. We observe a similar result. The GRU-based model chosen by the validation BA with adjusted class weight performs the best in terms of the BA (94.66%) and sensitivity (93.12%) on the testing set.

Table 4. Comparison of the testing performance (%) in Experiment 2.

	Model	By	Accuracy	Sensitivity	Specificity	BA
No Class Weight	GRU	Acc	98.02	82.12	99.41	90.77
		BA	97.58	90.18	98.23	94.20
	CNN	Acc	97.80	85.27	98.90	92.08
		BA	97.13	88.21	97.92	93.06
With Class Weight	GRU	Acc	97.93	85.27	99.04	92.15
		BA	95.95	93.12	96.19	94.66
	CNN	Acc	97.94	83.50	99.21	91.35
		BA	97.61	89.98	98.28	94.13
Zero-rule Baseline			91.94	0.00	100.00	50.00

4.3. Early Detection

Following Experiment 2, in which the annotation errors are corrected, we attempt to perform early detection. To achieve early detection, we crop data and label them in different windows. Figure 15 shows the cropping and labeling example for one-window-early detection. We try early detection up to nine windows by monitoring validation BA to pick the final model. The setting we adopt for training the model uses the GRU-based model and adjusts the class weight during the training stage. Table 5 shows the testing result of early detection. The case with zero window early corresponds to the result of Experiment 2. We can see that the performance is maintained in one window early. As the number of windows increases, the model's performance gradually differs from the original. Again, our primary concern is still balanced accuracy because of the imbalance of the data. We can find that as the number of windows increases, the value of this indicator has a gradual downward trend but is still in good shape (almost always above 90%).

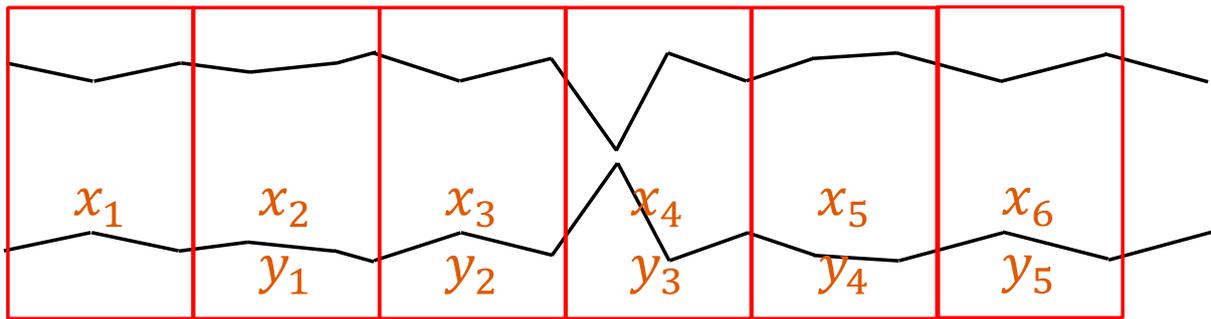


Figure 15. Cropping and labeling for one-window-early detection.

Table 5. Comparison of the testing performance for early detection.

	# of Windows Early									
	0	1	2	3	4	5	6	7	8	9
Accuracy	95.95	93.38	93.65	94.26	93.15	94.18	94.96	93.65	94.97	96.57
Sensitivity	93.12	93.59	90.82	93.22	92.37	91.56	89.36	92.33	85.54	82.04
Specificity	96.19	93.36	93.83	94.34	93.21	94.35	95.31	93.74	95.61	97.49
Balanced Accuracy	94.66	93.48	92.33	93.78	92.79	92.95	92.34	93.03	90.57	89.76
Zero-rule Accuracy	91.94	93.82	93.79	93.69	93.77	93.99	94.04	93.80	93.64	94.08
Zero-rule Baseline	Sensitivity: 0.00			Specificity: 100.00			Balanced Accuracy: 50.00			

5. Conclusions

We propose the GRU-based model with adjusted class weights to accomplish anomaly brainwave detection, which detects the existence of spikes and sharp waves in the EEGs of ICU patients. Unlike most other research, we adopt the bipolar montage in which medical doctors can easily find the anomaly events. The proposed GRU-based model can be used to monitor the brain activity of ICU patients more efficiently and cost-effectively. In clinical applications, the proposed GRU-based model can lighten the workload of medical staff in ICUs. Despite the data imbalance, our models’ sensitivity, specificity, and balanced accuracy are still all above 90%. These three metrics can represent the actual model performance. Although there is room for better sensitivity, our models can ease the burden on the medical staff in ICUs.

In addition to in-time detection, we also attempted early detection. In units of one window, we tried from one to nine windows and observed what pattern the models learned. As the number of windows increases, the balanced accuracy has a gradual downward trend but is still almost always above 90%. This justifies that the proposed GRU-based model with adjusted class weights has great potential in clinical applications.

In this study, we also include the CNN-based model architecture for comparisons. The methods of CNN-based models and GRU-based models can both be used for offline detection. Hence, we conduct the performance comparison for these two methods for offline detection. It turns out that the performance of the GRU-based model is better than that of the CNN-based model for offline detection in this empirical study. Furthermore, the GRU-based model can be used for real-time detection, but the CNN-based model cannot be used for such an application. Hence, the method of the GRU-based model is the suitable approach for offline and real-time detection.

The current models for automated anomaly detection are developed for patients’ data collected only in a medical center, i.e., TCVGH. Their detection performance in clinical applications may be degraded when they are applied to different medical centers and/or regional hospitals. To ensure satisfying detection performance across medical institutions, we may need to collaborate with them and collect their ICU patients’ data to fine-tune or

re-train the detection models. Another way is to develop federated learning-based models so that the detection models can be jointly trained in a decentralized way while preventing the violation of patient privacy [44].

Author Contributions: Conceptualization, all authors; methodology, J.C.-H.W. and T.-H.Y.; software, J.C.-H.W., T.-H.Y. and C.-C.H.; validation, N.-C.L. and Y.-W.P.; formal analysis, J.C.-H.W. and T.-H.Y.; investigation, all authors; resources, J.-A.H., C.-L.W. and H.H.-S.L.; data curation, N.-C.L., T.-H.Y., C.-C.H. and Y.-W.P.; writing—original draft preparation, J.C.-H.W. and T.-H.Y.; writing—review and editing, all authors; visualization, J.C.-H.W. and T.-H.Y.; supervision, C.-L.W. and H.H.-S.L.; project administration, Y.-J.H.; funding acquisition, J.C.-H.W., C.-L.W. and H.H.-S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work received funding from various sources, including the National Science and Technology Council (Grants: 110-2811-M-A49-550-MY2, 112-2811-M-A49-557-, 110-2118-M-A49-002-MY3, 111-2634-F-A49-014-, 112-2634-F-A49-003-, 113-2923-M-A49-004-MY3, 112-2634-F-A49-003-1), the Taichung Veterans General Hospital (Grants: TCVGH-YMCT1109105, TCVGH-1114404C), the Higher Education Sprout Project of the National Yang Ming Chiao Tung University from the Ministry of Education, and the Yushan Scholar Program of the Ministry of Education, Taiwan.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the Taichung Veterans General Hospital Ethics Review Committee (TCVGH:SE21316A, TCVGH:CG21307B).

Informed Consent Statement: Patient consent was waived due to the analyzed data being deidentified.

Data Availability Statement: The data in the present study are available upon request from the corresponding author.

Acknowledgments: We thank Wan-Yi Tai for her valuable assistance and acknowledge the National Center for High-performance Computing for providing computing resources.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Kennedy, J.D.; Gerard, E.E. Continuous EEG Monitoring in the Intensive Care Unit. *Curr. Neurol. Neurosci. Rep.* **2012**, *12*, 419–428. [[CrossRef](#)]
- Craik, A.; He, Y.; Contreras-Vidal, J.L. Deep learning for electroencephalogram (EEG) classification tasks: A review. *J. Neural Eng.* **2019**, *16*, 031001. [[CrossRef](#)]
- Chambon, S.; Galtier, M.N.; Arnal, P.J.; Wainrib, G.; Gramfort, A. A Deep Learning Architecture for Temporal Sleep Stage Classification Using Multivariate and Multimodal Time Series. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2018**, *26*, 758–769. [[CrossRef](#)]
- Biswal, S.; Kulas, J.; Sun, H.; Goparaju, B.; Westover, M.B.; Bianchi, M.T.; Sun, J. SLEEPNET: Automated Sleep Staging System via Deep Learning. *arXiv* **2017**, arXiv:1707.08262. [[CrossRef](#)]
- Eldele, E.; Chen, Z.; Liu, C.; Wu, M.; Kwok, C.K.; Li, X.; Guan, C. An Attention-Based Deep Learning Approach for Sleep Stage Classification With Single-Channel EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2021**, *29*, 809–818. [[CrossRef](#)] [[PubMed](#)]
- Jirayucharoensak, S.; Pan-Ngum, S.; Israsena, P. EEG-Based Emotion Recognition Using Deep Learning Network with Principal Component Based Covariate Shift Adaptation. *Sci. World J.* **2014**, *2014*, 627892. [[CrossRef](#)]
- Chen, J.X.; Zhang, P.W.; Mao, Z.J.; Huang, Y.F.; Jiang, D.M.; Zhang, Y.N. Accurate EEG-Based Emotion Recognition on Combined Features Using Deep Convolutional Neural Networks. *IEEE Access* **2019**, *7*, 44317–44328. [[CrossRef](#)]
- Tabar, Y.R.; Halici, U. A novel deep learning approach for classification of EEG motor imagery signals. *J. Neural Eng.* **2017**, *14*, 016003. [[CrossRef](#)] [[PubMed](#)]
- Amin, S.U.; Alsulaiman, M.; Muhammad, G.; Mekhtiche, M.A.; Hossain, M.S. Deep Learning for EEG motor imagery classification based on multi-layer CNNs feature fusion. *Future Gener. Comput. Syst.* **2019**, *101*, 542–554. [[CrossRef](#)]
- Yin, Z.; Zhang, J. Cross-session classification of mental workload levels using EEG and an adaptive deep learning model. *Biomed. Signal Process. Control.* **2017**, *33*, 30–47. [[CrossRef](#)]
- Zhang, P.; Wang, X.; Zhang, W.; Chen, J. Learning Spatial–Spectral–Temporal EEG Features With Recurrent 3D Convolutional Neural Networks for Cross-Task Mental Workload Assessment. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 31–42. [[CrossRef](#)] [[PubMed](#)]
- Hussein, R.; Palangi, H.; Ward, R.; Wang, Z.J. Epileptic Seizure Detection: A Deep Learning Approach. *arXiv* **2018**, arXiv:1803.09848. [[CrossRef](#)]

13. Cecotti, H.; Eckstein, M.P.; Giesbrecht, B. Single-Trial Classification of Event-Related Potentials in Rapid Serial Visual Presentation Tasks Using Supervised Spatial Filtering. *IEEE Trans. Neural Netw. Learn. Syst.* **2014**, *25*, 2030–2042. [[CrossRef](#)] [[PubMed](#)]
14. Vahid, A.; Bluschke, A.; Roessner, V.; Stober, S.; Beste, C. Deep Learning Based on Event-Related EEG Differentiates Children with ADHD from Healthy Controls. *J. Clin. Med.* **2019**, *8*, 1055. [[CrossRef](#)] [[PubMed](#)]
15. Wilson, S.B.; Emerson, R. Spike detection: A review and comparison of algorithms. *Clin. Neurophysiol.* **2002**, *113*, 1873–1881. [[CrossRef](#)] [[PubMed](#)]
16. Halford, J.J. Computerized epileptiform transient detection in the scalp electroencephalogram: Obstacles to progress and the example of computerized ECG interpretation. *Clin. Neurophysiol.* **2009**, *120*, 1909–1915. [[CrossRef](#)] [[PubMed](#)]
17. Gotman, J.; Gloor, P. Automatic recognition and quantification of interictal epileptic activity in the human scalp EEG. *Electroencephalogr. Clin. Neurophysiol.* **1976**, *41*, 513–529. [[CrossRef](#)]
18. Exarchos, T.P.; Tzallas, A.T.; Fotiadis, D.I.; Konitsiotis, S.; Giannopoulos, S. EEG transient event detection and classification using association rules. *IEEE Trans. Inf. Technol. Biomed.* **2006**, *10*, 451–457. [[CrossRef](#)] [[PubMed](#)]
19. Ji, Z.; Sugi, T.; Goto, S.; Wang, X.; Ikeda, A.; Nagamine, T.; Shibasaki, H.; Nakamura, M. An automatic spike detection system based on elimination of false positives using the large-area context in the scalp EEG. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 2478–2488. [[CrossRef](#)]
20. Adjouadi, M.; Cabrerizo, M.; Ayala, M.; Sanchez, D.; Yaylali, I.; Jayakar, P.; Barreto, A. A new mathematical approach based on orthogonal operators for the detection of interictal spikes in epileptogenic data. *Biomed. Sci. Instrum.* **2004**, *40*, 175–180.
21. Indiradevi, K.P.; Elias, E.; Sathidevi, P.S.; Dinesh Nayak, S.; Radhakrishnan, K. A multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram. *Comput. Biol. Med.* **2008**, *38*, 805–816. [[CrossRef](#)]
22. Acir, N.; Oztura, I.; Kuntalp, M.; Baklan, B.; Guzelis, C. Automatic detection of epileptiform events in EEG by a three-stage procedure based on artificial neural networks. *IEEE Trans. Biomed. Eng.* **2005**, *52*, 30–40. [[CrossRef](#)]
23. Tzallas, A.T.; Karvelis, P.S.; Katsis, C.D.; Fotiadis, D.I.; Giannopoulos, S.; Konitsiotis, S. A method for classification of transient events in EEG recordings: Application to epilepsy diagnosis. *Methods Inf. Med.* **2006**, *45*, 610–621. [[CrossRef](#)]
24. Güler, I.; Übeyli, E.D. Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients. *J. Neurosci. Methods* **2005**, *148*, 113–121. [[CrossRef](#)] [[PubMed](#)]
25. Abi-Chahla, F. Nvidia’s CUDA: The End of the CPU? Tom’s Hardware. 2008. Available online: <https://www.tomshardware.com/reviews/nvidia-cuda-gpu,1954.html> (accessed on 26 March 2024).
26. Rácz, M.; Liber, C.; Németh, E.; Fiáth, R.; Rokai, J.; Harmati, I.; Ulbert, I.; Márton, G. Spike detection and sorting with deep learning. *J. Neural Eng.* **2020**, *17*, 016038. [[CrossRef](#)]
27. Fukumori, K.; Nguyen, H.T.T.; Yoshida, N.; Tanaka, T. Fully Data-driven Convolutional Filters with Deep Learning Models for Epileptic Spike Detection. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 2772–2776.
28. Johansen, A.R.; Jin, J.; Maszczyk, T.; Dauwels, J.; Cash, S.S.; Westover, M.B. Epileptiform spike detection via convolutional neural networks. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 754–758. [[CrossRef](#)]
29. Saif-ur-Rehman, M.; Lienkämper, R.; Parpaley, Y.; Wellmer, J.; Liu, C.; Lee, B.; Kellis, S.; Andersen, R.; Iossifidis, I.; Glasmachers, T.; et al. SpikeDeepTector: A deep-learning based method for detection of neural spiking activity. *J. Neural Eng.* **2019**, *16*, 056003. [[CrossRef](#)]
30. Thirunavukarasu, R.; Gnanasambandan, R.; Gopikrishnan, M.; Palanisamy, V. Towards computational solutions for precision medicine based big data healthcare system using deep learning models: A review. *Comput. Biol. Med.* **2022**, *149*, 106020. [[CrossRef](#)] [[PubMed](#)]
31. Miotto, R.; Wang, F.; Wang, S.; Jiang, X.; Dudley, J.T. Deep learning for healthcare: Review, opportunities and challenges. *Brief. Bioinform.* **2017**, *19*, 1236–1246. [[CrossRef](#)]
32. Kotei, E.; Thirunavukarasu, R. Computational techniques for the automated detection of mycobacterium tuberculosis from digitized sputum smear microscopic images: A systematic review. *Prog. Biophys. Mol. Biol.* **2022**, *171*, 4–16. [[CrossRef](#)] [[PubMed](#)]
33. Rajkomar, A.; Oren, E.; Chen, K.; Dai, A.M.; Hajaj, N.; Hardt, M.; Liu, P.J.; Liu, X.; Marcus, J.; Sun, M.; et al. Scalable and accurate deep learning with electronic health records. *npj Digit. Med.* **2018**, *1*, 18. [[CrossRef](#)]
34. Ting, D.S.W.; Peng, L.; Varadarajan, A.V.; Keane, P.A.; Burlina, P.M.; Chiang, M.F.; Schmetterer, L.; Pasquale, L.R.; Bressler, N.M.; Webster, D.R.; et al. Deep learning in ophthalmology: The technical and clinical considerations. *Prog. Retin. Eye Res.* **2019**, *72*, 100759. [[CrossRef](#)] [[PubMed](#)]
35. Zhang, L.; Tan, J.; Han, D.; Zhu, H. From machine learning to deep learning: Progress in machine intelligence for rational drug discovery. *Drug Discov. Today* **2017**, *22*, 1680–1685. [[CrossRef](#)]
36. Singh, R.; Lanchantin, J.; Robins, G.; Qi, Y. DeepChrome: Deep-learning for predicting gene expression from histone modifications. *Bioinformatics* **2016**, *32*, i639–i648. [[CrossRef](#)]
37. Niedermeyer, E.; Lopes da Silva, F.H. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*; Lippincott Williams & Wilkins: Baltimore, MD, USA, 2004.
38. Tatum, W.O. *Handbook of EEG Interpretation*; Demos Medical Publishing: New York, NY, USA, 2014.

39. トマトン124. Electrode Locations of International 10-20 System for EEG (Electroencephalography) Recording. 2010. Available online: https://commons.wikimedia.org/wiki/File:21_electrodes_of_International_10-20_system_for_EEG.svg (accessed on 26 March 2024).
40. Kemp, B.; Värri, A.; Rosa, A.C.; Nielsen, K.D.; Gade, J. A simple format for exchange of digitized polygraphic recordings. *Electroencephalogr. Clin. Neurophysiol.* **1992**, *82*, 391–393. [[CrossRef](#)] [[PubMed](#)]
41. Kemp, B.; Oliván, J. European data format ‘plus’ (EDF+), an EDF alike standard format for the exchange of physiological data. *Clin. Neurophysiol.* **2003**, *114*, 1755–1761. [[CrossRef](#)] [[PubMed](#)]
42. Louis, E.K.S.; Frey, L.C.; Britton, J.W.; Hopp, J.L.; Korb, P.; Koubeissi, M.Z.; Lievens, W.E.; Pestana-Knight, E.M.; Foundation, C.C.C. *Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants*; Louis, E.K.S., Frey, L.C., Eds.; American Epilepsy Society: Chicago, IL, USA, 2016.
43. Chatrian, G.E. A glossary of terms most commonly used by clinical electroencephalographers. *Electroenceph. Clin. Neurophysiol.* **1974**, *37*, 538–548.
44. Wu, J.C.; Yu, H.W.; Tsai, T.H.; Lu, H.H. Dynamically Synthetic Images for Federated Learning of medical images. *Comput. Methods Programs Biomed.* **2023**, *242*, 107845. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.