Reliable and Fast Automatic Artifact Rejection of Long-Term EEG Recordings Based on Isolation Forest

Runkai Zhang [a], Rong Rong [b,+], John Q. Gan [c], Yun Xu [b], Haixian Wang [a,*], Xiaoyun Wang [b,*]

[a] Key Laboratory of Child Development and Learning Science of Ministry of Education, School of Biological Science & Medical Engineering, Southeast University, Nanjing 210096, Jiangsu, PR China

[b] Department of Neurology, Nanjing Drum Tower Hospital, Nanjing 210008, Jiangsu, PR China

[c] School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, UK

+ Joint first author.

* Corresponding authors. E-mail address: hxwang@seu.edu.cn (H. Wang); xysypy@126.com (X. Wang)

Abstract

Long-Term Electroencephalogram (Long-Term EEG) has the capacity to monitor over a long period, making it a valuable tool in medical institutions. However, due to the large volume of patient data, selecting clean data segments from raw Long-Term EEG for further analysis is an extremely time-consuming and labor-intensive task. Furthermore, the various actions of patients during recording make it difficult to use algorithms to denoise part of the EEG data, and thus lead to the rejection of these data. Therefore, tools for the quick rejection of heavily corrupted epochs in Long-Term EEG records are highly beneficial. In this paper, a new reliable and fast automatic artifact rejection method for Long-Term EEG based on Isolation Forest (IF) is proposed. Specifically, the IF algorithm is repetitively applied to detect outliers in the EEG data, and the boundary of inliers is promptly adjusted by using a statistical indicator to make the algorithm proceed in an iterative manner. The iteration is terminated when the distance metric between clean epochs and artifact-corrupted epochs remains unchanged. Six statistical indicators (i.e., min, max, median, mean, kurtosis, and skewness) are evaluated by setting them as centroid to adjust the boundary during iteration, and the proposed method is compared with several state-of-the-art methods on a retrospectively collected dataset. The experimental results indicate that utilizing the min value of data as the centroid yields the most optimal performance,

and the proposed method is highly efficacious and reliable in the automatic artifact rejection of Long-Term EEG, as it significantly improves the overall data quality. Furthermore, the proposed method surpasses compared methods on most data segments with poor data quality, demonstrating its superior capacity to enhance the data quality of the heavily-corrupted data. Besides, owing to the linear time complexity of IF, the proposed method is much faster than other methods, thus providing an advantage when dealing with extensive datasets.

Keywords: Long-Term EEG; Automatic Rejection; Isolation Forest; Outlier Detection; Machine Learning

1. Introduction

Long-Term Electroencephalogram (Long-Term EEG) is a type of EEG that records over a long period of time, rather than a specific duration [1]. It is used primarily for epilepsy monitoring, but is also used in Intensive Care Units (ICU), Operating Rooms, and Emergency Rooms [2-4]. As Long-Term EEG is used to monitor seizures, cerebrovascular diseases, and psychiatric conditions, it typically lasts from hours to days [5-11]. During the recording process, the EEG may contain multiple signals from both neuronal and non-neuronal sources, with the latter often referred to as artifacts, which interfere with neural signals [12]. Artifacts are usually manually identified and removed from the data before EEG signals are further analyzed [13]. However, this manual annotation procedure is both time-consuming and subjective [14], making an efficient and reliable automatic artifact removal tool highly desirable.

Several advanced algorithms have been developed for the automated preprocessing of artifacts, which can be divided into two categories: identifying or detecting artifacts in EEG, and processing artifacts that have been discovered [15-19]. In recent years, many methods based on Deep Learning (DL) or Machine Learning (ML) have been proposed for the former category [20-29]. For example, in 2022, the Convolutional Neural Network with Transformer (CNN-Transformer) was proposed to detect artifacts at single channel level and segment level, which was validated on the TUH Artifact dataset (TUH-ART) [30]. For the latter category, despite the famous Independent Component Analysis (ICA), Artifact Subspace Reconstruction (ASR), and Signal Space Projection (SSP) methods that have been applied to the correction and reconstruction of EEG, novel DL-based denoising methods continue to emerge [31-40].

However, the aforementioned methods have certain limitations. Most researchers who employ DL validate their methods on one or a few datasets, resulting in poor generalization performance and scalability [41]. Additionally, algorithms based on signal reconstruction theory encounter challenges to process epochs of signals that have been heavily corrupted by artifacts. In such cases, employing certain signal completion techniques can be beneficial. However, in the absence of recordings regarding these completion signals in a clinical setting, repairing artifacts becomes particularly challenging. Given these circumstances, rejecting these signals may be regarded as a viable and favorable choice.

Most existing tools reject epochs based on the Peak-to-Peak Amplitude (PTP), and the mainstream EEG analysis software integrates the PTP-based method [42, 43]. However, the selection of threshold in PTP value is data-specific and requires the expertise of practitioners, making an automated specification of threshold preferable. In this regard, the automated artifact rejection for MEG and EEG data (Autoreject) has achieved remarkable success and has been utilized in various types of research [44]. Nevertheless, Autoreject has some drawbacks. Firstly, due to its implementation of an interpolation algorithm and Bayesian Optimization, it runs slowly when processing intensive data. Secondly, its performance is not satisfactory when the overall data quality of the signal is poor. These shortcomings make it unsuitable for Long-Term EEG.

In order to address the aforementioned issues, this paper proposes a novel automatic method for the artifact rejection of Long-Term EEG. The proposed method can significantly improve the overall data quality in a relatively short running time, making the method both reliable and fast. The superior performance of the proposed method is mainly due to the ability of Isolation Forest (IF) to accurately partition the feature space, as well as the linear time complexity of IF. The contributions of this paper can be summarized as follows:

(a) A reliable and fast automatic rejection method for clinical Long-Term EEG is proposed, which is based on iterative application of the IF, to avoid manually selecting the peak-to-peak amplitude threshold for artifact rejection of Long-Term EEG.

(b) A metric utilized to measure the class distance between epochs that should be dropped and epochs that should be retained is designed to promptly terminate the iteration of the IF. To evaluate the performance of artifacts removal by using the proposed method, six different statistical indicators are utilized and considered accordingly by setting them as the metric during iteration.

(c) The proposed method is evaluated across Long-Term EEG recordings of six retrospectively collected patients, in comparison with four state-of-the-art methods, including two unsupervised methods and two supervised methods.

2. Materials and methods

In this section, the description of patients and a series of basic preprocessing operations are introduced first. Subsequently, the proposed automatic artifact rejection method based on IF is detailed comprehensively.

The aim of the design of this method is to provide a reliable and efficient tool for swiftly removing corrupted segments from the data when conducting non-continuous epoch-level EEG analyses in research or clinical settings, thus avoiding the need for extensive manual data rejection. The motivation for constructing the proposed method is based on the aim of providing a reliable and efficient tool for capturing artifact introduced by subjects during Long-Term EEG recording. In this regard, the PTP metric remains an effective and practical choice. Furthermore, the IF algorithm was initially designed for efficient and precise outlier detection, making it well-suited for this purpose. When both are appropriately combined, it becomes effortless for the algorithm to detect outlier epochs in a Long-Term EEG recording. However, due to the complexity of clinical environments, it is often challenging for subjects to maintain a fully relaxed resting state, which is often desirable for research and medical purposes. Instead, subjects typically exhibit some degree of mental activity or minor movements. As a result, there are relatively few data segments with low noise levels, and the entire dataset is characterized by moderate levels of noise, often accompanied by notable artifacts. In this scenario, the proposed method considers an iterative use of the IF algorithm for identifying outliers in the data while retaining epochs with relatively low PTP values in each iteration, until all epochs distorted by artifacts are correctly classified. Following the aforementioned design philosophy, the flowchart of the proposed method is shown in Fig. 1. The main procedures are presented in Fig. 1 and will be demonstrated in this section, respectively.

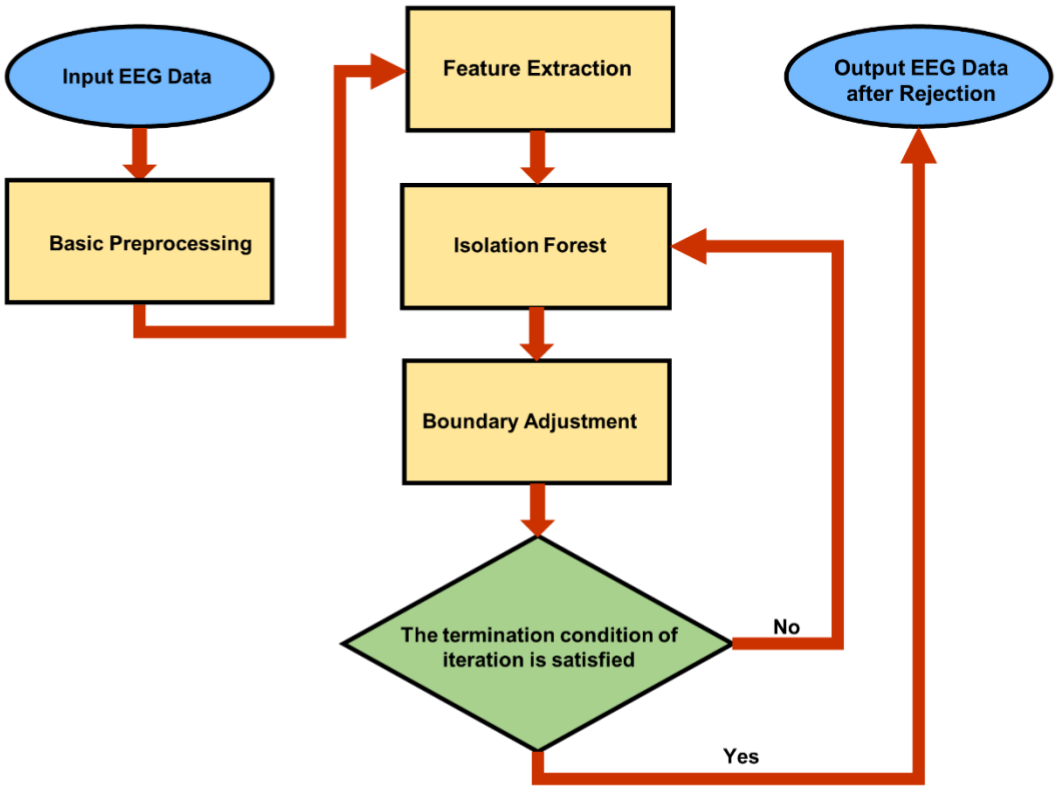


Fig. 1. Flowchart of the proposed method.

2.1 Data Preparation and Basic Preprocessing

The present study enrolled 6 subjects in total, of which 2 subjects had epilepsy and 4 subjects were healthy. Data from all subjects were retrospectively collected from the database of the Department of Neurology of Nanjing Drum Tower Hospital from 2021 to 2022. All the subjects provided informed consent and underwent Long-Term Video-EEG for up to 20 hours. The inclusion criteria for epilepsy were: (1) subject age ≥ 18 years old; (2) EEG recordings of subjects showed obvious epileptic discharge, and subjects had previously experienced seizures. The inclusion criteria for healthy subjects were: (1) subject age ≥ 18 years old; (2) The subjects presented for consultation due to heatstroke or syncope, however, no visible abnormalities were observed in the EEG. This study was approved by the Ethics Committee of the Department of Neurology, Nanjing Drum Tower Hospital, Nanjing, China.

As shown in Fig. 2, a 19-channel montage based on the 10-20 International System was used to collect the EEG signals. In order to maintain the configuration of each subject consistent, the recordings were referenced to the average of earlobes, namely A1 and A2. Since the sampling rates of the EEG recordings are different, the data were resampled at 500 Hz for the convenience of subsequent analysis. The high-frequency muscle artifacts resulting from movements such as chewing and head motion are common artifacts in EEG recordings, and eye movement signals often interfere with the subsequent analysis of EEG signals. Therefore, in order to verify the effectiveness of the proposed methods, bandpass filtering and independent component analysis were intentionally omitted for reducing or suppressing these artifacts. Instead, notch filters were exclusively used to remove global noise associated with the power source, ensuring proper operation of the methods. To remove power-line interference, an FIR notch filter with zero phase and hamming window is designed by using MNE [43], and is applied at 50 Hz for each subject's EEG recording.

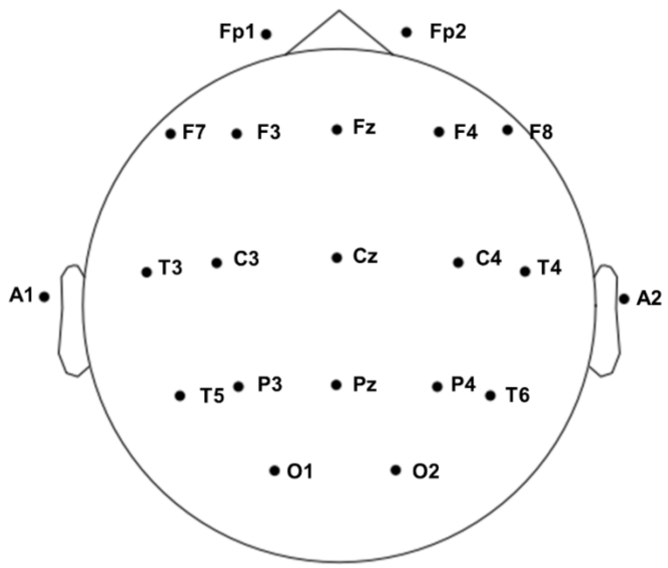


Fig. 2. Channel configuration of selected Long-Term EEG recordings.

2.2 Feature Extraction

Initially, as shown by the first block in the middle column of Fig. 1, the features of epochs in the original signals are extracted. PTP is widely used in the analysis of EEG. For the purpose of simplicity of notation, a formal definition is established in this subsection. In consideration of EEG research with a large amount of data, the raw data is always divided into epochs of equal time length, so the PTP values for each channel in one epoch are defined by the following equation:

$$E_n^c = \max x_n^c(k) - \min x_n^c(k) \tag{1}$$

where E_n^c is the PTP value for cth channel in nth epoch, $x(k)$ is in discretization form of time course of single channel within an epoch, with k being the sampling point, where $0 \leq k \leq K - 1$ and K is the sampling rate. By computing the PTP value E_n^c for each channel in each epoch, the raw EEG time series can be reformulated as a new feature matrix:

$$X_{feature} = \begin{bmatrix} E_1^1 & E_2^1 & \cdots & E_N^1 \\ E_1^2 & E_2^2 & \cdots & E_N^2 \\ \vdots & \vdots & \ddots & \vdots \\ E_1^C & E_2^C & \cdots & E_N^C \end{bmatrix} \tag{2}$$

Thus, $X_{feature}$ is a dense matrix of size $C \times N$.

2.3 Isolation Forest

Following the extraction of epoch-level features, the subsequent step is to identify the corrupted epochs. Since the ratio of corrupted epochs to clean epochs in a given Long-Term EEG recording is indeterminate, it is suitable to define the process of corrupted epochs rejection as a two-class imbalanced classification task. Hence, methods formulated for anomaly detection can be employed to resolve this problem.

This paper proposes to use the IF to distinguish between corrupted and clean epochs. The reason to choose the IF is that it has linear time complexity with a low constant and a low memory requirement [45]. The IF assumes that, when randomly partitioning the feature space of specified samples, samples with distinguishable features are more likely to be separated in early partitioning [46]. Thus, nodes with shorter path lengths in the isolation trees are highly likely to be anomalies.

Overall, anomaly detection using the IF is a two-stage procedure. The first stage involves constructing isolation trees through the application of the specified samples and the partitioning process, while the second stage entails passing the samples through the isolation trees to generate an anomaly score for each sample.

An isolation tree built from a dataset $X = \{x_1, x_2, \ldots, x_n\}$ is a complete binary tree, where each node in the tree has exactly zero or two children. In the first stage of the IF algorithm, a collection of t isolation trees are generated by initially sub-sampling the provided set X to size ψ, followed by a recursive partition of the set X_{ψ}. The partition operation selects a feature dimension of the samples q and a split value p, such that the comparison $q < p$ divides samples into left subtree and right subtree. The termination conditions of constructing an isolation tree are:

(1) The depth of the tree reaches the limit l, where $l = ceiling(log2\ \psi)$.

(2) $|X_{\psi}| = 1$.

(3) All values in q are the same.

Upon completion of the aforementioned operations for M times, an isolation forest consisting of M isolation trees is constructed.

In the second stage of the IF algorithm, an anomaly score s is derived from the expected path length $E(h(x))$ for each sample presented in X, where $h(x)$ is calculated by counting the number of edges from the root node to a terminating node as instance x traverses through an isolation tree and $E(h(x))$ is the average of $h(x)$ from a collection of isolation trees. As described in [46], the trees constructed by the IF algorithm have an equivalent structure to a Binary Search Tree (BST). In a BST, an unsuccessful search is defined as the inability to find a specified element within the tree. Therefore, the path length of termination due to an external node in the IF algorithm can be estimated using the theory of BST. The average path length of an unsuccessful search in a binary search tree is defined as follows:

$$c(n) = 2H(n-1) - \left(\frac{2(n-1)}{n}\right) \quad (3)$$

where $H(i)$ is the harmonic number and it can be estimated by $ln(i) + 0.5772156649$ (Euler's constant). Since $c(n)$ can be used to normalize $h(x)$, the anomaly score s of a sample is formulated as:

$$s = 2^{-\frac{E(h(x))}{c(n)}} \quad (4)$$

The anomaly score s can be used to identify anomalies within the data. The algorithm arranges the data in descending order according to s, and the first m instances are the top m anomalies. As illustrated in Fig. 1, the IF algorithm is utilized repeatedly. Following each iteration, the most probable outliers are identified as potentially corrupted epochs.

2.4 Boundary Adjustment

Since the IF detects outliers on the basis of anomaly score, it is capable of discovering corrupted epochs in the iteration. According to the definition of PTP, clean epochs are supposed to have a lower value of PTP compared to heavily corrupted epochs. However, in clinical Long-Term EEG recordings, the proportion of time for patients to maintain resting-state and perform daily activities is unknown. In other words, patients only need to wear the signal collector during clinical EEG monitoring, and they can engage in activities such as eating and drinking within the monitoring area. However, the duration of patients lying down and keeping their mind empty is uncertain. As a result, the collected signals contain both resting-state and other daily activity EEG signals in a coupling manner. Consequently, the ratio of time spent in resting-state and other daily activities cannot be determined explicitly. Under such circumstances, in a clinical Long-Term recording that needs to be preprocessed, the aggregation trend of data is also non-deterministic. In the feature space, when epochs with higher values of PTP tend to cluster, those with lower values of PTP (representing data when the patient is in a resting state) are significantly distanced from the cluster center, often resulting in their identification as outliers by the IF algorithm and subsequent elimination, despite being the epochs of interest.

To address this issue, six methods are proposed to adjust the boundary between outliers and inliers. Since it is unambiguous that clean epochs always have a lower value of PTP, those epochs considered as outliers by the IF but with a low value of PTP should be retained as well. After each decision of the IF, the data can be divided into two sets, one as inliers and the other as outliers, which can be expressed as follows:

$$S_{inliers} = \{s^1, s^2, \dots, s^p\} \tag{5}$$

and

$$S_{outliers} = \{s^1, s^2, \dots, s^q\} \tag{6}$$

where p and q are the numbers of epochs determined as inliers and outliers in the current iteration. Due to the multi-dimensional nature of EEG signals, segmental feature extraction using peak-to-peak values allows for temporal separation and reduction of the data. However, the spatial dimensions are still preserved. The main design motivation of the proposed method is to facilitate comparison and iteration on the most salient features of the data, enabling the progressive identification of epochs heavily contaminated by signals. Principal component analysis (PCA) is a classical data analysis technique that linearly combines the original variables to generate new composite variables while preserving the maximum amount of information. With the aim of identifying the primary components of variance in the data, PCA is employed in the boundary adjustment stage to reduce dimensionality and eliminate redundant information. To mathematically represent the boundary adjustment method, the PTP value at the epoch level after PCA reduction is defined as:

$$E_{epo}(s) = PCA\left(s, n_{compoents} = 1\right) \tag{7}$$

where $s \in S_{inliers}$ or $S_{outliers}$, and $n_{components} = 1$ indicates that original feature matrix is reduced to one dimension. Thus, the maximum value of retained epochs is used to adjust the boundary:

$$E_{max}^{inliers} = \max\left(E_{epo}(s)\right) \tag{8}$$

where $s \in S_{inliers}$, and $E_{max}^{inliers}$ represent the maximum PTP value of retained epochs. If $E_{epo}(s) < E_{max}^{inliers}$ when $s \in S_{outliers}$, then this epoch should be regarded as an inlier and be retained with other inliers. This boundary method is denoted as the Max Method. In the same way, the min value of retained epochs can be defined as:

$$E_{min}^{inliers} = \min\left(E_{epo}(s)\right) \tag{9}$$

and it is used as a comparator to retain epochs with $E_{epo}(s) < E_{min}^{inliers}$ when $s \in S_{outliers}$. This boundary method is denoted as the Min Method. Similarly, mean and median values of the retained epochs can also be used to adjust the boundary of outliers and inliers. They can be defined as:

$$E_{mean}^{inliers} = \text{mean}\left(E_{epo}(s)\right) \tag{10}$$

and

$$E_{median}^{inliers} = \text{median}\left(E_{epo}(s)\right) \tag{11}$$

respectively. Then the epochs decided as outliers by the IF are supposed to be retained if $E_{epo}(s) < E_{mean}^{inliers}$ or $E_{epo}(s) < E_{median}^{inliers}$ when $s \in S_{outliers}$. These two methods are called the Mean Method and the Median Method, respectively. Moreover, we have also taken into consideration the metrics of the distribution. Therefore, kurtosis and skewness of the distribution are used to adjust the boundary of outliers and inliers, and they are denoted as

Kurtosis Method and Skewness Method. Thus, they are defined to be:

$$E_{kurtosis}^{inliers} = \text{kurtosis}\left(E_{epo}(s)\right) \tag{12}$$

and

$$E_{skewness}^{inliers} = \text{skewness}\left(E_{epo}(s)\right) \tag{13}$$

respectively. These six methods are all effective for adjusting the boundary but have different characteristics, which will be explored in the next section.

2.5 Termination Condition of the IF Iteration

In order to implement automatic rejection of artifacts, it is necessary to specify a termination condition for the IF iteration. For this reason, two classes should be defined:

(1) Ω_{retain}: all the epochs that are confirmed to be clean by the algorithm.

(2) Ω_{drop}: all the epochs that are confirmed to be artifact-corrupted by the algorithm.

After adjusting the boundary of separation results of the IF, the sets $S_{inliers}$ and $S_{outliers}$ can be rearranged and merged to Ω_{retain} and Ω_{drop}. By using L2 norm, the distance between Ω_{retain} and Ω_{drop} is defined as:

$$Dist = \left|\max\left(E_{epo}(s^i)\right) - \min\left(E_{epo}(s^j)\right)\right|_2 \tag{14}$$

where $s^i \in \Omega_{\text{retain}}$ and $s^j \in \Omega_{drop}$. As the number of iterations increases, the distance defined above will be subject to alteration. This is due to the gap in the PTP values between clean epochs and artifact-corrupted epochs being varied in the successive iteration. Finally, when all the artifact-corrupted epochs are correctly classified, the IF algorithm will detect no outlier present or incorrectly classify minor clean epochs as outliers. In either of the scenarios, the distance between the two classes will remain unchanged. Consequently, the iteration should be terminated and the automatic rejection is completed.

The source code for this research is available on GitHub at the following URL: https://github.com/RunKZhang/Isolation_Forest_Automatic_Rejection.

3. Experiments and results

In this section, the proposed method is verified and compared with the state-of-the-art. For each patient described in the previous section, the first two-hour data of the recordings is extracted. For each two-hour data, the data is split from the middle to build two one-hour datasets. Next, non-overlapping one-second epochs were made for each one-hour data. Thus, twelve one-hour data segments with 3600 epochs were built. The two-hour data is extracted because, upon registration and commencement of Long-Term EEG recording, patients are typically unable to remain still and at rest. The segmentation of EEG recordings is executed by using MNE.

Considering that the length of data processed by different algorithms may not be consistent, the evaluation metrics for data quality should be independent of data length. Thus, the metrics used to evaluate the data quality before and after rejection were Overall Data Quality (ODQ) and Overall Data Quality Rating (DQR) introduced by Zhao et al. in [47]. The ODQ value is expressed as:

$$ODQ = \frac{M_{good\ windows}}{M_{total\ windows}} \tag{15}$$

where $M_{good\ windows}$ is the number of good windows in a data segment, while $M_{total\ windows}$ is the number of total windows in a data segment. A window refers to the time course of a single channel within an epoch and a good window indicates that the single time course is with an acceptable level of noise. Since the quantitative signal quality evaluation method for EEG proposed in [47] is threshold-based, it requires no ground-truth labels for good windows and automatically identifies the good windows based on a set of parameters. Considering its nature of not requiring manual labeling, this metric is particularly suitable for evaluating data quality in situations where the dataset lacks true labels and the size of the dataset is tremendous. The number of total windows $M_{total\ windows}$ is denoted as:

$$M_{total\ windows} = C \times N \tag{16}$$

where C is the number of channels in a data segment and N is the number of epochs in a data segment. From the definition of ODQ, it can be inferred that this metric is suitable for evaluating the data quality of the same data segment after being processed by different methods. In can be summarized that the higher the ODQ value, the better the quality of the data. The authors in [47] manually partition the ratings of DQR, which correspond to "perfect," "good," "poor," and "bad." The rating of DQR is determined based on the numerical value of ODQ. When ODQ is less than 60, an EEG recording's DQR is classified as D, indicating bad data quality. When

ODQ is greater than or equal to 60 but less than 80, the DQR of a recording is classified as C, indicating poor data quality. When ODQ is greater than or equal to 80 but less than 90, the DQR of a recording is classified as B, indicating good data quality. When ODQ is greater than or equal to 90, the DQR is classified as A, indicating perfect data quality. The ODQ values and their corresponding DQR ratings are illustrated in Table 1.

Table 1. The ODQ value and corresponding DQR.

DQR	ODQ Value
A	$ODQ \geq 90$
B	$90 > ODQ \geq 80$
C	$80 > ODQ \geq 60$
D	$60 > ODQ$

The IF is implemented using scikit-learn and the parameters are set to default, which are shown in Table 2. Since the IF algorithm is a type of ensemble method, the number of base estimators actually represents the number of random trees in the forest, and it is defaulted to be 100 in scikit-learn. The number of samples denotes the amount of data drawn from the dataset to construct a random tree, and this parameter value set to 'Auto' indicates that scikit-learn will automatically select the minimum value between the size of the dataset and 256. The contamination represents the proportion of outliers in the original data, and this parameter value set to 'Auto' means that the threshold is determined as in the original paper [46][47]. The number of features denotes the percentage of the dimension of the feature vector drawn from the original data to train random trees, and 1.0 means that all the features are used for training. Bootstrap is a parameter used to control the sampling method, where 'True' means that training data were sampled with replacement, and 'False' indicates sampling without replacement is performed. The ODQ value and the DQR are calculated using the WeBrain platform and Python scripts [48].

Table 2. The parameters of the IF in scikit-learn.

Parameter Names	Parameter Values
The number of base estimators	100
The number of samples	Auto
contamination	Auto
The number of features	1.0
bootstrap	False

The data quality prior to artifact rejection is shown in Fig. 3. It is evident that the ODQ varies in a large range from 0 to 88.89 and the DQR is between D and B. The average quality value of these segments is (53.24 ± 27.87) and the average rating is D.

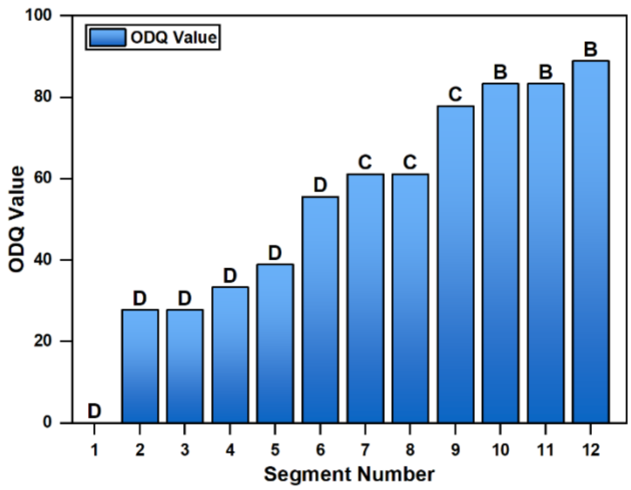


Fig. 3. Overall Data Quality and corresponding ratings.

3.1 Evaluation of Six Boundary Adjustment Methods

The six methods for boundary adjustment were compared in terms of ODQ values in our experiment to illustrate which adjustment method is the most reasonable. The ODQ values after rejection by the six boundary adjustment methods are illustrated in Fig. 4.

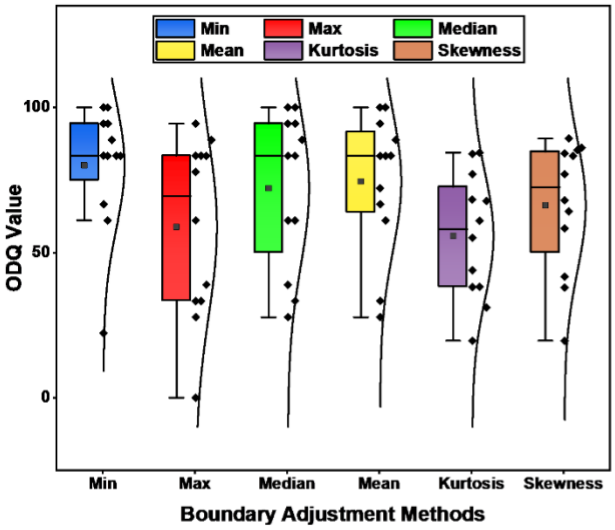


Fig. 4. Boxplot of the performance of six boundary adjustment methods.

From Fig. 4, it can be noted that the lowest ODQ value of the Min Method is 22.22, while the lowest ODQ values of the other five methods are 0, 27.77, 27.77, 19.64 and 19.64, respectively. Meanwhile, it is evident from the figure that the highest ODQ values among the boundary adjustment methods based on simple statistics exceed 90, while the two methods based on distribution metrics only exceed 80. This indicates that these six boundary adjustment methods are effective in eliminating artifact-corrupted epochs. However, the box size of the Min Method is the smallest among the six methods, implying that the overall data quality after rejection by the proposed method with the Min Method as the boundary adjustment method can reach a superior level. As observed from the distribution curve on the right side of each box, the ODQ values of the other five methods are more dispersed and the average ODQ value of the Min Method is higher than those of other methods.

The number of iterations required for convergence for each data segment is illustrated in Fig. 5. The distance between two classes is calculated after each iteration, and the iteration stops when the distance values of two adjacent iterations are the same.

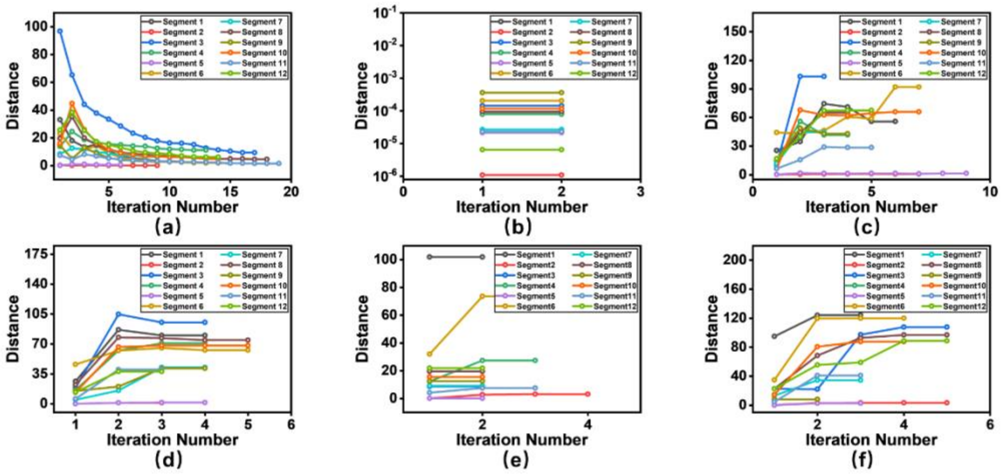


Fig. 5. Numbers of iterations required for convergence by the six methods that use different statistical indicators as the centroid. (a) Min Method. (b) Max Method. (c) Median Method. (d) Mean Method. (e) Kurtosis Method. (f) Skewness Method.

The distance values recorded in Fig. 5 were calculated after each iteration. However, prior to any iterations or before the first iteration, the number of elements in set Ω_{drop} is zero, indicating that the set containing epochs confirmed to be artifact-corrupted is empty. In equation (14), the minimum value of elements in the set Ω_{drop} is required, but the minimum value of an empty set is undefined. Hence, Fig. 5 omits the distance values before any iterations and starts recording from the first iteration onwards. After the first iteration, both sets Ω_{retain} and Ω_{drop} contain elements. However, due to different data segments evaluated in the experiments, the elements of the two sets are distinct, resulting in different calculated distance values. Fig. 5 shows a discernible trend that the Max Method requires the least number of iterations to achieve convergence and only needs two iterations in most data segments. Similarly, the Kurtosis Method converges after a few iterations. On the contrary, using the min value as the centroid to adjust the boundary needs more iterations to achieve convergence. The performances of the other three methods tend to be analogous, which indicates that they need a similar running time. In general, although the Min Method requires the maximum number of iterations to achieve convergence compared with the other five boundary adjustment methods, this method is capable of reaching the best data quality after rejecting artifact-corrupted epochs.

Fig. 6 provides an overview of the number of epochs of each data segment after corrupted epochs being rejected by the proposed method. It is noted from the figure that using the min value as the centroid always resulted in more iterations and fewer retained epochs.

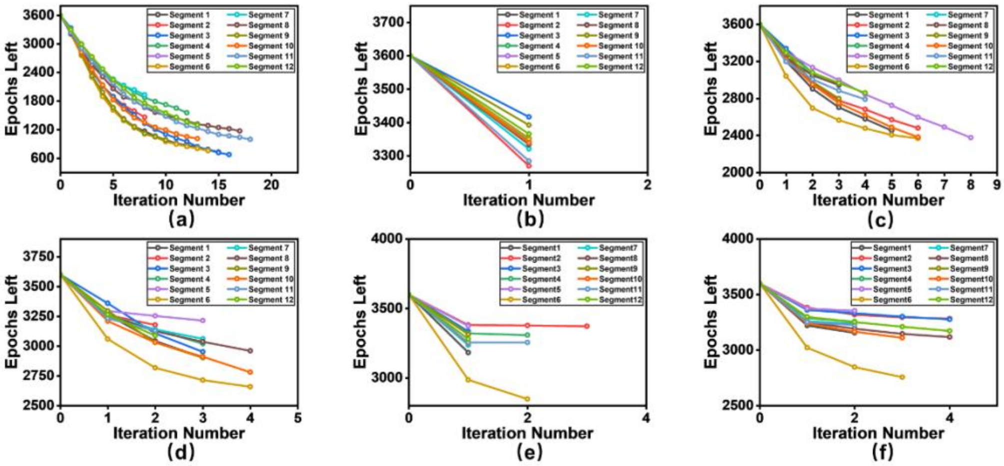


Fig. 6. The number of epochs of each data segment after corrupted epochs being rejected by the proposed method that use different statistical indicators as the centroid. (a) Min Method. (b) Max Method. (c) Median Method. (d) Mean Method. (e) Kurtosis Method. (f) Skewness Method.

3.2 Comparison with the State-of-the-Art

In this subsection, the proposed method is compared with several state-of-the-art automatic artifact removal methods, including two unsupervised methods and two supervised methods. The first method is Autoreject, which adaptively selects thresholds for discriminating artifacts from clean segments using cross-validation and employs Bayesian Optimization for optimal threshold [44]. The second method, referred to as AUTO in this paper, takes a series of complex handcrafted features as input to an autoencoder for unsupervised learning and outlier removal [49]. The third and fourth methods are EEGdenoiseNet and Interpretable CNN, well-known convolutional neural network-based approaches for artifact handling in EEG signals [35] [22]. In this paper, the AUTO is trained for 100 epochs to discriminate artifact-corrupted epochs, and contamination is set to 0.1 by default according to [49]. For EEGdenoiseNet, a classifier was added to the end as stated in [22] and trained for 100 epochs on the EEGdenoiseNet dataset, which is a semi-synthetic EEG dataset. According to [22], Interpretable CNN is also trained for 100 epochs on the EEGdenoiseNet dataset to achieve a high accuracy of classifying clean EEG and artifacts. After completing the training of both networks, they were validated on the collected dataset used in this paper. Each epoch's data was inputted to the network as a batch for validation purposes. The training and testing of deep learning models in this paper were executed using NVIDIA GeForce GTX 1080 Ti. The comparison results are shown in Table 3. It is noteworthy that the epochs left by these methods differ as shown in Table 3, and this can be attributed to the normal inconsistency in the length of the remaining data after being processed by different algorithms. This is because during the running process, different algorithms will reject epochs that they deem to be highly contaminated by artifacts, so it is challenging to ensure that processing the same data segment with different algorithms will yield data of the same length. Thus, the discrepancy in data length after automatic rejection renders it sensible to compare the quality of the artifact-rejected data using ODQ values.

Table 3. Results from comparison between the proposed method and other methods.

Method	Running Time	Number of Retained Epochs	ODQ Value
Proposed Method	$5.48 \pm 1.04\text{s}$	1226.25 ± 468.42	80.09 ± 21.77
Autoreject	$582.28 \pm 34.29\text{s}$	1888.58 ± 1212.27	58.33 ± 37.57
AUTO	$100.09 \pm 11.32s$	3240 ± 0	58.66 ± 26.27
EEGdenoiseNet	$24.96 \pm 2.51s$	3048.91 ± 705.22	62.57 ± 23.07
Interpretable CNN	$7.19 \pm 0.89s$	2443.83 ± 1242.99	63.68 ± 23.70

The ODQ values before and after artifact removal by using the proposed method and other methods are shown in Fig. 7(a). The proposed method yields superior or equivalent results in most data segments compared to other methods. However, not all these methods can be used to improve the quality of the data. After being processed by AUTO and EEGdenoiseNet, segment twelve even exhibits a lower ODQ value compared to the original data, and Autoreject shows a similar performance on segment four. As shown in Fig. 3, the bars that illustrate data segments are sorted by ODQ values, and segments one to six are rated as D. The proposed method yields the highest improvement in ODQ value through artifact removal when applied to these data segments. This indicates that the proposed method resulted in a higher proportion of good windows in the artifact-rejected data compared to other methods, implying that the proposed method is able to perform better than other methods on poor-quality data segments. Fig. 8(a) is a scatter plot, with the y-axis showing the ODQ values of the proposed method and the x-axis showing the ODQ values of the other methods. Evidently, the ODQ values of segments after artifact rejection by the proposed method are higher than those of the other methods, as most data points are above the dashed line. On the contrary, the scatter plot illustrates that the proposed method is capable of improving the data quality to some extent, as there are few data points located in the bottom left corner.

Subsequently, the number of epochs left by the proposed method was compared to those left by other methods, and the results are depicted in Fig. 7(b). The figure suggests that the proposed method tends to reject more epochs than other methods on poor-quality data segments, thus indicating that it is more proficient in detecting artifact-corrupted epochs. Notably, the two methods based on supervised learning refused to drop any epochs on segments 4, 7, and 10, leading to the failure of improving the ODQ value. From the perspective of ML/DL, the quality of data is more important than the quantity. Figure 8(b) shows the scatter plot of the number of retained epochs after rejection by the proposed method (y-axis) and other methods (x-axis). More data points are below the dashed line in Fig. 8(b), which demonstrates that the data segments have fewer epochs left after rejection by the proposed method than after rejection by other methods. From Fig. 7 and Fig. 8, it can be seen that, when using these methods to reject heavily-corrupted epochs, the proposed method results in a higher proportion of good windows in the artifact-rejected data, even when fewer epochs are retained. In other words, the ratio of clean epochs to contaminated epochs in the remaining data segments has been increased.

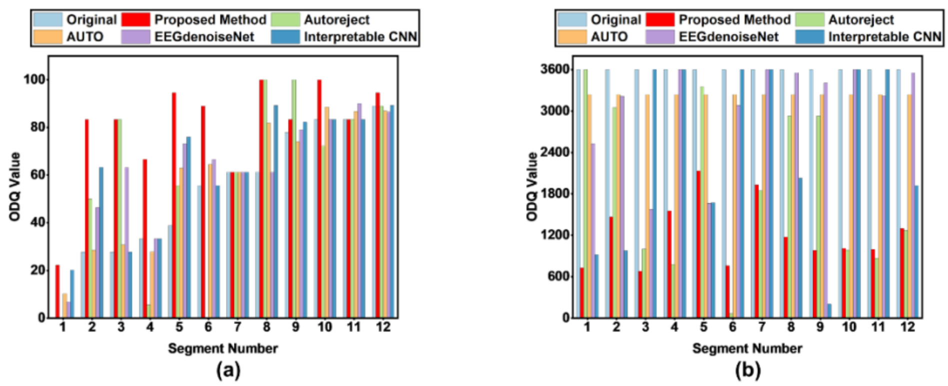


Fig. 7. Comparisons of ODQ values and number of epochs between original data and data after rejection by the proposed method and other methods. (a) ODQ values of original data and data after rejection by the proposed method and other methods. (b) Number of epochs

of original data and data after rejection by the proposed method and other methods.

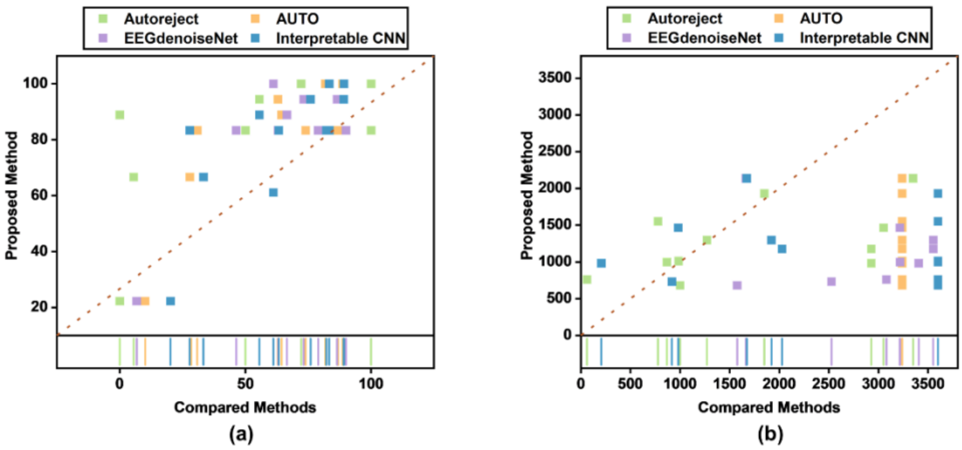


Fig. 8. Scatter plots for comparing the performances of the proposed method and other methods in terms of the ODQ value and the number of retained epochs after artifact rejection. (a) Scatter plot for comparison between the proposed method and other methods in terms of ODQ value. (b) Scatter plot for comparison between the proposed method and other methods in terms of number of retained epochs.

Furthermore, the relationship between the running time and the length of the data segment of the aforementioned rejection methods was compared. To this end, data segments with lengths equal to one hour, three hours, and five hours were extracted from the patients and used for comparison. The five methods were executed on these new data segments, and their corresponding running times were recorded. As a result, Fig. 9 shows the logarithmic average running time across segments versus the corresponding epoch length with points in square. The straight lines through these points represent the linear fit curve between the running time and the corresponding length of the data segment. The parameters after fitting are presented in Table 4. As demonstrated in Fig. 9 and Table 4, the linear fit result of the proposed method has a lower intercept than that of other methods, implying that the running time of the mentioned methods is comparable when the length of the data segment is short. Moreover, although the slopes of the two supervised learning methods is slightly higher than the proposed method, the time cost for pretraining the two methods should also be taken into consideration, implying a potential higher time complexity.

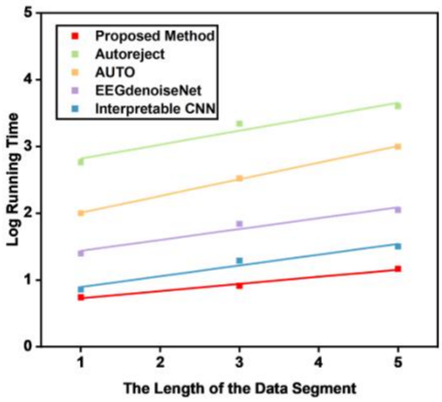


Fig. 9. Relationships between running time and length of the data segment of the proposed method and other methods.

Table 4. Results of linear fitting between the proposed method and other methods.

Method	Slope	Intercept	R square	Pearson's r
Proposed method	0.10 ± 0.01	0.62 ± 0.04	0.98	0.99
Autoreject	0.21 ± 0.05	2.61 ± 0.16	0.95	0.97
AUTO	0.24 ± 0.01	1.76 ± 0.02	0.99	0.99
EEGdenoiseNet	0.16 ± 0.03	1.27 ± 0.12	0.96	0.98
Interpretable CNN	0.16 ± 0.03	0.73 ± 0.10	0.96	0.98

4. Discussion

In this study, a novel automatic rejection method for enhancing the data quality of clinical Long-Term EEG recordings at the very initial stage of preprocessing is proposed and the performance of the proposed method is evaluated by comparison with four state-of-the-art methods. This section discusses the results depicted in the previous section, and the distinctions and similarities between the proposed method and other methods are elucidated in certain contexts.

In the previous section, it was initially demonstrated that using the minimum value of the data as the centroid to adjust the boundary is more effective than other statistical indicators. The rationale behind this phenomenon is that the boundary adjustment technique based on the minimum value can consistently target the minimum value in the data and increase the iteration times to discover more outliers. It has been observed that while using median, mean, maximum, kurtosis, and skewness can be beneficial in improving the quality of the data, they also have the potential to terminate the iteration process prematurely, thus preventing from achieving the optimum result. In theory, using skewness and kurtosis would yield better results compared to simple statistical metrics, as they consider the overall shape of the distribution. However, their performance is inferior when compared to the Min Method. This can be attributed to the relatively small variation in skewness and kurtosis across the entire data, making it easy for them to remain unchanged after removing a portion of data segments, thereby ending the removal process. Nevertheless, using the minimum value as the centroid will result in fewer epochs compared to other methods, since using the minimum value as the centroid requires more iterations. Therefore, when aiming to keep fewer epochs than the other five methods, those epochs with a high level of data quality will be preserved.

The results from the previous sections demonstrate that the proposed method exhibits certain advantages over the state-of-the-art methods in terms of reliability. From a principle-based perspective, Autoreject employs a cross-validation framework and utilizes the Frobenius norm of the mean value of good trials in the training set, as well as the median value in the validation set. Therefore, it assumes that the mean and median values of the PTP of the dataset are sufficient to discriminate artifacts from the clean data. However, as stated in Section II, the proportion of time for a patient to stay in a resting-state or perform activities is unknown, and there may be instances where the artifact-corrupted data exceeds the clean data. Mean and median values used in Autoreject may not be able to effectively differentiate between artifact-contaminated epochs and clean epochs since they potentially assume that epochs with a higher PTP value than mean and median are likely to contain artifacts, and epochs with a lower PTP value are likely to be clean. However, this assumption may not always hold true in clinical settings. The reason for the superiority of the proposed method over Autoreject lies in its divergence from the approach employed by Autoreject. Unlike Autoreject, the proposed method utilizes the minimum value of retained epochs as a criterion to distinguish between clean epochs and contaminated epochs during the iterative process. Although this may result in a reduction in the number of retained epochs, the minimum value of the retained epochs consistently ensures an improvement in data quality. For the comparison between the proposed method and other methods, it is imperative to acknowledge the significant contributions that DL-based methods have made to the preprocessing stage of EEG. In terms of feature extraction, AUTO applies manipulation to multiple features and metrics extracted from the EEG, compressing them into a more compact space. Subsequently, a decision boundary is curved to separate the outliers from the inliers. Compared to the proposed method, the major drawback of AUTO lies in its assumption of a fixed proportion of outliers in the data. Although this proportion can be treated as a hyperparameter of the model, the proposed method, on the other hand, continuously discovers contaminated epochs in the data through an iterative process. EEGdenoiseNet and Interpretable CNN employ distinct convolution kernels to extract features from the original EEG signals and classify clean EEG signals from artifacts based on these extracted features, and they necessitate pre-training on the dataset prior to being transferred to other scenarios. Nonetheless, the training source data often lacks diversity, leading to suboptimal transfer performance. The proposed approach, in contrast, can be regarded as an unsupervised method that achieves satisfactory results without the need for pre-training.

On the other hand, the proposed method needs the shortest execution time compared to other methods. From a theoretical standpoint, the proposed method constructs isolation trees by partitioning the feature space to form sub-trees, and the feature space shrinks with each iteration step. Contrarily, Autoreject requires the adoption of Bayesian Optimization to select the optimal threshold from a set of candidate thresholds. The optimization-based design often yields satisfactory results within an acceptable runtime for datasets of small sizes. However, for larger datasets, the time required to find the global optimum solution is considerably longer compared to directly partitioning the feature space. For the DL-based methods, EEGdenoiseNet and Interpretable CNN, they exhibit short execution times on large-scale datasets with the support of GPUs. However, the time required for pre-training them should also be taken into consideration. Although their pre-training time in this study is in the order of thousands of seconds, this time may increase exponentially as the size of the source dataset increases. The low runtime efficiency of AUTO can be attributed to its extraction of numerous features from the EEG as inputs to the network, which consumes more time compared to using only PTP.

However, this study still has some limitations. Firstly, a limited number of patients were used to validate the proposed method in this study. Due to the specificity of EEG, it is reasonable to consider a larger number of participants for validation. Secondly, this paper only addresses the removal of contaminated data without considering the use of methods for data restoration, which serves as a potential direction for future research.

5. Conclusion

In summary, a novel reliable and fast method for automatic rejection of clinical EEG recordings is proposed in this paper. The results illustrated in Section III indicate that the proposed method with min value as the centroid greatly improved the data quality, which implies that the proposed method is reliable in the automatic artifact rejection of Long-Term EEG recordings. Furthermore, the proposed method is also compared with the current state-of-the-art methods for preprocessing clinical EEG data. The comparison results suggest that the proposed method is competitive in most circumstances and it performs better than other methods especially when data quality is poor. Meanwhile, the comparison of running time indicates that the proposed method has a lower time complexity and is much faster than other methods. This is the consequence of the repeated application of an advanced data-driven outlier detection algorithm, accompanied by the establishment of an appropriate centroid, which ultimately led to the fulfillment of the termination condition. By building a tool to help researchers clean up data automatically, researchers can reduce the time required to inspect data, thus allowing them to focus on scientific research instead of parameter tuning for preprocessing.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 62176054.

Declarations

Conflict of interest: The authors declare that they have no conflict of interest.

References

[1] T. William O. IV, Long-Term EEG Monitoring, Journal of Clinical Neurophysiology. 18 (2001) 442–455. https://doi.org/10.1097/00004691-200109000-00009.

[2] A. Alkhachroum, B. Appavu, S. Egawa, B. Foreman, N. Gaspard, E.J. Gilmore, L.J. Hirsch, P. Kurtz, V. Lambrecq, J. Kromm, P. Vespa, S.F. Zafar, B. Rohaut, J. Claassen, Electroencephalogram in the Intensive Care Unit: A Focused Look at Acute Brain Injury, Intensive Care Medicine. 48 (2022) 1443–1462. https://doi.org/10.1007/s00134-022-06854-3.

[3] J.-M. Guérit, Neuromonitoring in the Operating Room: Why, When, and How to Monitor?, Electroencephalography and Clinical Neurophysiology. 106 (1998) 1–21. https://doi.org/10.1016/s0013-4694(97)00077-1.

[4] J.H. Rodríguez Quintana, S.J. Bueno, J.L. Zuleta-Motta, M.F. Ramos, A. Vélez-van-Meerbeke, Utility of Routine EEG in Emergency Department and Inpatient Service, Neurology: Clinical Practice. 11 (2020) e677–e681. https://doi.org/10.1212/cpj.0000000000000961.

[5] A.D. Patel, B. Haridas, Z.M. Grinspan, J. Stevens, Utility of Long-Term Video-EEG Monitoring for Children with Staring, Epilepsy & Behavior. 68 (2017) 186–191. https://doi.org/10.1016/j.yebeh.2017.01.002.

[6] S. Wang, W. Wang, G. Yu, L. Wan, Y. Fan, H. Wang, T. Liu, T. Ji, Q. Liu, L. Cai, X. Liu, Safety and Efficacy of Rapid Withdrawal of Anti-seizure Medications during Long-Term Video-Electroencephalogram Monitoring in Children with Drug Resistant Epilepsy: A Retrospective Study, Epilepsia Open. (2023). https://doi.org/10.1002/epi4.12680.

[7] C. Nouboue, S. Selfi, E. Diab, S. Chen, B. Périn, W. Szurhaj, Assessment of an Under-Mattress Sensor as a Seizure Detection Tool in an Adult Epilepsy Monitoring Unit, Seizure. 105 (2023) 17–21. https://doi.org/10.1016/j.seizure.2023.01.005.

[8] M.H. Adenan, M. Khalil, K.S. Loh, L. Kelly, A. Shukralla, S. Klaus, R. Kilbride, G.

Mullins, P. Widdess-Walsh, M. Kinney, N. Delanty, H. El-Naggar, A Retrospective Study of the Correlation between Duration of Monitoring in the Epilepsy Monitoring Unit and Diagnostic Yield, Epilepsy & Behavior. 136 (2022) 108919. https://doi.org/10.1016/j.yebeh.2022.108919.

[9] S.W. Terman, S.S. O'Kula, M.M. Asmar, K.A. Davis, D.M. Gazzola, R. Lesanu, L. George, L.M. Selwa, S.M. Glynn, C.E. Hill, Inpatient Long-Term Video-Electroencephalographic Monitoring Event Capture Audiovisual Diagnostic Quality, Epilepsy & Behavior. 137 (2022) 108947. https://doi.org/10.1016/j.yebeh.2022.108947.

[10] C.M. Fleseriu, I. Sultan, J.A. Brown, A. Mina, J. Frenchman, D.J. Crammond, J. Balzer, K.M. Anetakis, K. Subramaniam, V. Shandal, F. Navid, P.D. Thirumala, Role of Intraoperative Neurophysiological Monitoring in Preventing Stroke after Cardiac Surgery, in: The Annals of Thoracic Surgery, 2023. https://doi.org/10.1016/j.athoracsur.2023.01.004.

[11] A.L. Brian, T.H. M'hamed, F.T. Stephen, B.T. Brian, J.M. Austin, M.K. Tara, B.L. Varina, M. Iris, A.A. Todd, A.D. Phillip, Quantitative Electroencephalography after Pediatric Anterior Circulation Stroke, Journal of Clinical Neurophysiology. 39 (2020) 610–615. https://doi.org/10.1097/wnp.0000000000000813.

[12] M. Diachenko, S.J. Houtman, E.L. Juarez-Martinez, J.R. Ramautar, R. Weiler, H.D. Mansvelder, H. Bruining, P. Bloem, K. Linkenkaer-Hansen, Improved Manual Annotation of EEG Signals through Convolutional Neural Network Guidance, eNeuro. 9 (2022) ENEURO.0160-22.2022. https://doi.org/10.1523/eneuro.0160-22.2022.

[13] J.A. Urigüen, B. Garcia-Zapirain, EEG Artifact Removal—State-of-the-Art and Guidelines, Journal of Neural Engineering. 12 (2015) 031001. https://doi.org/10.1088/1741-2560/12/3/031001.

[14] S. Sadiya, T. Alhanai, M.M. Ghassemi, Artifact Detection and Correction in EEG Data: A Review, in: 2021 10th International IEEE/EMBS Conference on Neural Engineering (NER), 2021. https://doi.org/10.1109/ner49283.2021.9441341.

[15] M.M.N. Mannan, M.A. Kamran, M.Y. Jeong, Identification and Removal of Physiological Artifacts from Electroencephalogram Signals: A Review, IEEE Access. 6 (2018) 30630–30652. https://doi.org/10.1109/ACCESS.2018.2842082.

[16] M.K. Islam, A. Rastegarnia, Z. Yang, Methods for Artifact Detection and Removal from Scalp EEG: A Review, Neurophysiologie Clinique/Clinical Neurophysiology. 46 (2016) 287–305. https://doi.org/10.1016/j.neucli.2016.07.002.

[17] W. Mumtaz, S. Rasheed, A. Irfan, Review of Challenges Associated with the EEG Artifact Removal Methods, Biomedical Signal Processing and Control. 68 (2021) 102741. https://doi.org/10.1016/j.bspc.2021.102741.

[18] X. Jiang, G. Bian, Z. Tian, Removal of Artifacts from EEG Signals: A Review, Sensors (Basel, Switzerland). 19 (2019) 987. https://doi.org/10.3390/s19050987.

[19] X. Chen, X. Xu, A. Liu, S. Lee, X. Chen, X. Zhang, M.J. McKeown, Z.J. Wang, Removal of Muscle Artifacts from the EEG: A Review and Recommendations, IEEE Sensors Journal. 19 (2019) 5353–5368. https://doi.org/10.1109/jsen.2019.2906572.

[20] M.E. O'Sullivan, G. Lightbody, S.R. Mathieson, W.P. Marnane, G.B. Boylan, J.M. O'Toole, Development of an EEG Artefact Detection Algorithm and Its Application in Grading Neonatal Hypoxic-Ischemic Encephalopathy, Expert Systems with Applications. 213 (2023) 118917. https://doi.org/10.1016/j.eswa.2022.118917.

[21] T.M. Ingolfsson, A. Cossettini, S. Benatti, L. Benini, Energy-Efficient Tree-Based EEG Artifact Detection, in: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2022. https://doi.org/10.1109/embc48229.2022.9871413.

[22] F. Paissan, V.P. Kumaravel, E. Farella, Interpretable CNN for Single-Channel Artifacts Detection in Raw EEG Signals, in: 2022 IEEE Sensors Applications Symposium (SAS), 2022. https://doi.org/10.1109/sas54819.2022.9881381.

[23] J. Wang, J. Cao, D. Hu, T. Jiang, F. Gao, Eye Blink Artifact Detection with Novel Optimized Multi-Dimensional Electroencephalogram Features, IEEE

Transactions on Neural Systems and Rehabilitation Engineering. 29 (2021) 1494–1503. https://doi.org/10.1109/tnsre.2021.3099232.

[24] S. Stalin, V. Roy, P.K. Shukla, A. Zaguia, M.M. Khan, P.K. Shukla, A. Jain, A Machine Learning-Based Big EEG Data Artifact Detection and Wavelet-Based Removal: An Empirical Approach, Mathematical Problems in Engineering. 2021 (2021) 1–11. https://doi.org/10.1155/2021/2942808.

[25] O. Komisaruk, E. Nikulchev, Neural Network Model for Artifacts Marking in EEG Signals, International Journal of Advanced Computer Science and Applications. 12 (2021) 28–35. https://doi.org/10.14569/ijacsa.2021.0121204.

[26] J. Cao, L. Chen, D. Hu, F. Dong, T. Jiang, W. Gao, F. Gao, Unsupervised Eye Blink Artifact Detection from EEG with Gaussian Mixture Model, IEEE Journal of Biomedical and Health Informatics. 25 (2021) 2895–2905. https://doi.org/10.1109/JBHI.2021.3057891.

[27] H. Tiwary, A. Bhavsar, Time-Frequency Representations for EEG Artifact Classification with CNNs, in: 2021 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), 2021. https://doi.org/10.1109/aipr52630.2021.9762201.

[28] M. Tosun, Ö. Kasım, Novel Eye-Blink Artefact Detection Algorithm from Raw EEG Signals Using FCN-Based Semantic Segmentation Method, IET Signal Processing. 14 (2020) 489–494. https://doi.org/10.1049/iet-spr.2019.0602.

[29] K. Yasoda, R.S. Ponmagal, K.S. Bhuvaneshwari, K. Venkatachalam, Automatic Detection and Classification of EEG Artifacts Using Fuzzy Kernel SVM and Wavelet ICA (WICA), Soft Computing. 24 (2020) 16011–16019. https://doi.org/10.1007/s00500-020-04920-w.

[30] W. Peh, Y. Yao, J. Dauwels, Transformer Convolutional Neural Networks for Automated Artifact Detection in Scalp EEG, in: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2022. https://doi.org/10.1109/embc48229.2022.9871916.

[31] F. Barban, M. Chiappalone, G. Bonassi, D. Mantini, M. Semprini, Yet Another Artefact Rejection Study: An Exploration of Cleaning Methods for Biological and

Neuromodulatory Noise, Journal of Neural Engineering. 18 (2021) 0460c2. https://doi.org/10.1088/1741-2552/ac01fe.

[32] S. Blum, N.S.J. Jacobsen, M.G. Bleichner, S. Debener, A Riemannian Modification of Artifact Subspace Reconstruction for EEG Artifact Handling, Frontiers in Human Neuroscience. 13 (2019) 1–10. https://doi.org/10.3389/fnhum.2019.00141.

[33] L.A. Bradshaw, A. Myers, W.O. Richards, W. Drake, J.P. Wikswo, Vector Projection of Biomagnetic Fields, Medical & Biological Engineering & Computing. 43 (2005) 85–93. https://doi.org/10.1007/bf02345127.

[34] I. Winkler, S. Brandl, F. Horn, E. Waldburger, C. Allefeld, M. Tangermann, Robust Artifactual Independent Component Classification for BCI Practitioners, Journal of Neural Engineering. 11 (2014) 035013. https://doi.org/10.1088/1741-2560/11/3/035013.

[35] H. Zhang, M. Zhao, C. Wei, D. Mantini, Z. Li, Q. Liu, EEGdenoiseNet: A Benchmark Dataset for Deep Learning Solutions of EEG Denoising, Journal of Neural Engineering. 18 (2021) 056057. https://doi.org/10.1088/1741-2552/ac2bf8.

[36] J. Yin, A. Liu, C. Li, R. Qian, X. Chen, Frequency Information Enhanced Deep EEG Denoising Network for Ocular Artifact Removal, IEEE Sensors Journal. 22 (2022) 21855–21865. https://doi.org/10.1109/jsen.2022.3209805.

[37] G. Tamburro, P. Fiedler, D. Stone, J. Haueisen, S. Comani, A New ICA-Based Fingerprint Method for the Automatic Removal of Physiological Artifacts from EEG Recordings, PeerJ. 6 (2018) e4380. https://doi.org/10.7717/peerj.4380.

[38] M. Chavez, F. Grosselin, A. Bussalb, F. De Vico Fallani, X. Navarro-Sune, Surrogate-Based Artifact Removal from Single-Channel EEG, IEEE Transactions on Neural Systems and Rehabilitation Engineering. 26 (2018) 540–550. https://doi.org/10.1109/tnsre.2018.2794184.

[39] B. Somers, T. Francart, A. Bertrand, A Generic EEG Artifact Removal Algorithm Based on the Multi-Channel Wiener Filter, Journal of Neural Engineering. 15 (2018) 036007. https://doi.org/10.1088/1741-2552/aaac92.

[40] C. Liu, C. Zhang, Remove Artifacts from a Single-Channel EEG Based on VMD

and SOBI, Sensors. 22 (2022) 6698. https://doi.org/10.3390/s22176698.

[41] L. Xu, M. Xu, Y. Ke, X. An, S. Liu, D. Ming, Cross-Dataset Variability Problem in EEG Decoding with Deep Learning, Frontiers in Human Neuroscience. 14 (2020). https://doi.org/10.3389/fnhum.2020.00103.

[42] A. Delorme, S. Makeig, EEGLAB: An Open Source Toolbox for Analysis of Single-Trial EEG Dynamics Including Independent Component Analysis, Journal of Neuroscience Methods. 134 (2004) 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009.

[43] A. Gramfort, M. Luessi, E. Larson, D.A. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen, M.S. Hämäläinen, MNE Software for Processing MEG and EEG Data, NeuroImage. 86 (2014) 446–460. https://doi.org/10.1016/j.neuroimage.2013.10.027.

[44] M. Jas, D.A. Engemann, Y. Bekhti, F. Raimondo, A. Gramfort, Autoreject: Automated Artifact Rejection for MEG and EEG Data, NeuroImage. 159 (2017) 417–429. https://doi.org/10.1016/j.neuroimage.2017.06.030.

[45] F.T. Liu, K.M. Ting, Z.-H. Zhou, Isolation-Based Anomaly Detection, ACM Transactions on Knowledge Discovery from Data. 6 (2012) 1–39. https://doi.org/10.1145/2133360.2133363.

[46] F.T. Liu, K.M. Ting, Z.-H. Zhou, Isolation Forest, in: 2008 Eighth IEEE International Conference on Data Mining, 2008. https://doi.org/10.1109/icdm.2008.17.

[47] L. Zhao, Y. Zhang, X. Yu, H. Wu, L. Wang, F. Li, M. Duan, Y. Lai, T. Liu, L. Dong, D. Yao, Quantitative Signal Quality Assessment for Large-Scale Continuous Scalp EEG from a Big Data Perspective, Physiological Measurement. (2022). https://doi.org/10.1088/1361-6579/ac890d.

[48] L. Dong, J. Li, Q. Zou, Y. Zhang, L. Zhao, X. Wen, J. Gong, F. Li, T. Liu, A.C. Evans, P.A. Valdes-Sosa, D. Yao, WeBrain: A Web-Based Brainformatics Platform of Computational Ecosystem for EEG Big Data Analysis, NeuroImage. 245 (2021) 118713. https://doi.org/10.1016/j.neuroimage.2021.118713.

[49] S. Saba-Sadiya, E. Chantland, T. Alhanai, T. Liu, M. M. Ghassemi, Unsupervised EEG Artifact Detection and Correction, Frontiers in digital health. 2 (2021). 608920. https://doi.org/10.3389/fdgth.2020.608920.

Biographies

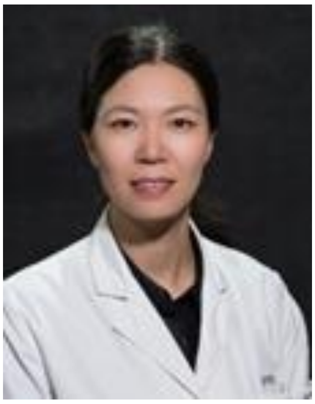
Runkai Zhang received the B.S. degree from Henan University. He is currently working toward his Ph.D. degree at Southeast University. His research interests include signal processing and epilepsy.

Rong Rong received the M.S. degree from Nanjing Medical University. She is now with the Department of Neurology, Nanjing Drum Tower Hospital. Her research interests focus on epilepsy.

John Q. Gan received the Ph.D. degree in biomedical electronics from Southeast University in 1991. His current research interests include biomedical engineering, machine learning theory and algorithms.

Yun Xu received the Ph.D. degree from the Department of Biochemistry at Nanjing University in 2006. She is currently the chairman in the Department of Neurology at Nanjing Drum Tower Hospital.

Haixian Wang received the Ph.D. degree in computer science from Anhui University in 2005. His research interests focus on brain-computer interfaces and machine learning.



Xiaoyun Wang received the M.S. degree in clinical and scientific research of neurology from the Xiangya School of Medicine. Her research interests focus on electroencephalography and epilepsy.