Epileptic Seizure Detection by Cascading Isolation Forest-Based Anomaly Screening and EasyEnsemble

Yao Guo, Xinyu Jiang, Linkai Tao, Long Meng, *Graduate Student Member, IEEE*, Chenyun Dai, Xi Long, *Senior Member, IEEE*, Feng Wan, *Senior Member, IEEE*, Yuan Zhang, *Senior Member, IEEE*, Johannes van Dijk, *Senior Member, IEEE*, Ronald M. Aarts, *Fellow, IEEE*, Wei Chen, *Senior Member, IEEE*, and Chen Chen

***Abstract*— The electroencephalogram (EEG), for measuring the electrophysiological activity of the brain, has been widely applied in automatic detection of epilepsy seizures. Various EEG-based seizure detection algorithms have already yielded high sensitivity, but training those algorithms requires a large amount of labelled data. Data labelling is often done with a lot of human efforts, which is very time-consuming. In this study, we propose a hybrid system integrating an unsupervised learning (UL) module and a supervised learning (SL) module, where the UL module can significantly reduce the workload of data labelling. For preliminary seizure screening, UL synthesizes amplitude-integrated EEG (aEEG) extraction, isolation forest-based anomaly detection, adaptive segmentation, and silhouette coefficient-based anomaly detection evaluation. The UL module serves to quickly locate the determinate subjects (seizure segments and seizure-free segments) and the indeterminate subjects (potential seizure candidates). Afterwards, more robust seizure detection for the indeterminate subjects is performed by the SL using an EasyEnsemble algorithm. EasyEnsemble, as a class-imbalance learning method, can potentially decrease the generalization error of the seizure-free segments. The proposed method can significantly reduce the workload of data labelling while guaranteeing satisfactory performance. The proposed seizure detection system is evaluated using the Children's Hospital Boston-Massachusetts Institute of Technology (CHB-MIT) scalp EEG dataset, and it achieves a mean accuracy of 92.62%, a mean sensitivity of 95.55%, and a mean specificity of 92.57%. To the best of our knowledge, this is the first epilepsy seizure detection study employing the integration of both the UL and the SL modules, achieving a competitive performance superior or similar to that of the state-of-the-art methods.**

***Index Terms*— Seizure detection system, unsupervised learning, supervised learning, EEG, aEEG, anomaly detection.**

Manuscript received September 30, 2021; revised January 29, 2022; accepted February 27, 2022. Date of publication March 30, 2022; date of current version April 13, 2022. This work was supported in part by the Shanghai Municipal Science and Technology Major Project under Grant 2017SHZDZX01, in part by the National Natural Science Foundation of China under Grant 62001118 and Grant 62172340, in part by the Shanghai Committee of Science and Technology under Grant 20S31903900, and in part by the Shanghai Municipal Science and Technology International Research and Development Collaboration Project under Grant 20510710500. *(Corresponding authors: Wei Chen; Chen Chen.)*

Yao Guo, Xinyu Jiang, Long Meng, and Chenyun Dai are with the Center for Intelligent Medical Electronics, School of Information Science and Technology, Fudan University, Shanghai 200438, China (e-mail: yguo19@fudan.edu.cn; jiangxy18@fudan.edu.cn; lmeng18@fudan.edu.cn; chenyundai@fudan.edu.cn).

Linkai Tao is with the Department of Industrial Design, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands (e-mail: 148243215@qq.com).

Xi Long and Ronald M. Aarts are with the Department of Electrical Engineering, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands (e-mail: x.long@tue.nl; r.m.aarts@tue.nl).

Feng Wan is with the Centre for Cognitive and Brain Sciences and the Department of Electrical and Computer Engineering, Faculty of Science and Technology, Institute of Collaborative Innovation, University of Macau, Macau 999078, China (e-mail: fwan@um.edu.mo).

Yuan Zhang is with the College of Electronic and Information Engineering, Southwest University, Chongqing 400715, China (e-mail: yuanzhang@swu.edu.cn).

Johannes van Dijk is with the Epilepsy Center Kempenhaeghe, 5590 AB Heeze, The Netherlands, also with the Department of Electrical Engineering, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands, and also with the Department of Orthodontics, University of Ulm, 89081 Ulm, Germany (e-mail: j.p.v.dijk@tue.nl).

Wei Chen is with the Center for Intelligent Medical Electronics, School of Information Science and Technology, and the Human Phenome Institute, Fudan University, Shanghai 200438, China (e-mail: w_chen@fudan.edu.cn).

Chen Chen is with the Human Phenome Institute, Fudan University, Shanghai 200438, China (e-mail: chenchen_fd@fudan.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2022.3163503

I. Introduction

EPILEPSY, as one of the common neurological diseases, affects more than 50 million people of all ages worldwide and leads to a major international public health concern [1]. It is characterized by irregular seizures, repeated seizures or severe convulsions, and may bring out deterioration of the physical condition, intellectual development and even become life-threatening. Electroencephalography (EEG), an electrophysiological monitoring method by recording the electrical activity of the brain using multi-channel sensors, is one of the primary methods for diagnosing and detecting epilepsy [2], [3]. However, labelling the beginning and end time of seizures is time-consuming. Moreover, with the fatigue caused by long-duration work, manual inspection is prone to increase errors. Therefore, developing an effective automatic epileptic seizure detection method can considerably reduce

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

the workload of medical staff, and greatly support the life of epileptic patients [4].

Nowadays, supervised learning (SL) is the mainstream machine learning approach for EEG-based automatic seizure detection [5]. In recent years, many SL methods based on different classifiers have been developed for seizure detection, such as support vector machines (SVM) [6], convolutional neural networks (CNN) [7], and Takagi-Sugeno-Kang (TSK) fuzzy system [8]–[10]. Bhattacharyya *et al*. [11] extracted discriminative features from different oscillatory levels using the multi-channel extension of empirical wavelet transform (EWT). Recently, deep learning-based seizure detection methods have predominantly attracted researchers' interest, mainly due to their superiority in automatic feature learning directly from the raw data, and due to the high performance and robustness in detecting seizures. Tian *et al*. [7] designed a multi-view deep feature extraction for seizure detection method based on fast Fourier transform (FFT), wavelet packet decomposition (WPD), CNN and TSK fuzzy system. These works are evaluated on CHB-MIT dataset and their results all exhibit the potential of machine learning and deep learning methods for seizure detection.

However, the above studies are all SL-based methods, where large amounts of labelled data are required, especially for deep learning methods requiring a large number of parameters to be optimized. With a small labelled dataset, the deep learning-based methods may have constrained generalizability due to the over-fitting problem. An effective approach to address these issues is to utilize an anomaly detection method based on unsupervised learning (UL), which does not require data labelling. Tsiouris *et al*. [12] designed a simple rule-based seizure detection method for detecting an anomaly seizure epoch. You *et al*. [13] utilized complex generative adversarial networks as an anomaly detection method to detect seizures automatically. However, the seizure detection rate (SDR) of the UL method adopted in these two studies was much worse than SL methods. In consideration of the trade-off between the requirement of the data labelling and the performance of seizure detection, a hybrid seizure detection method that combines the UL method and the SL method is proposed. It is dedicated to utilizing a small amount of labelled data for training while still aiming for favorable performance in seizure detection. In the UL module, a preliminary screening based on the amplitude-integrated EEG (aEEG), isolation forest and silhouette coefficient (SC) is performed to quickly distinguish between the determinate subjects (seizure segments and seizure-free segments) and indeterminate subjects. The subjects whose SC value higher than the threshold belong to determinate subjects (Group U) and these results of determinate subjects from UL modules contribute as the partial final results of our proposed system. For the subjects whose SC is lower than the threshold, they belong to indeterminate subjects (Group S). Afterwards, the indeterminate subjects are passed to the SL module for further detection of seizures, which is expected to perform better than the UL module.

This paper proposes a novel pipeline framework for automatic seizure detection combining UL and SL from a small labelled dataset. The main contributions are summarized as follows: First, this is the first work to integrate UL and SL methods in automatic seizure detection and this is also the first work that attempts to utilize the aEEG and isolation forest for the identification of seizure segments. Second, with the UL module, preliminary screening that can quickly assign subjects into determine subjects and indeterminate subjects. For these determined subjects, the seizure-free segments and seizure segments can be obtained using UL method. Third, as the time-compressed EEG on a semi-logarithmic scale, aEEG has been explored in the preliminary screening. The feasibility of the aEEG is investigated and verified in the identification of seizure segments and seizure-free segments. Fourth, the isolation forest method combined with the SC are proposed in the UL part. The impacts of SC on the performance and the optimal range of SC are investigated.

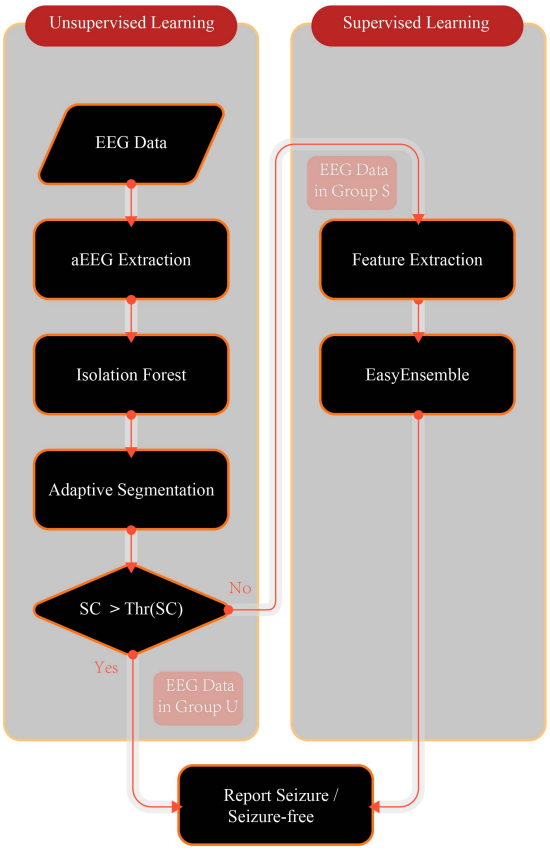

Fig. 1. The flowchart of the proposed method.

The remainder of the paper is organized as follows. Section II provides the details of our proposed method. Results are demonstrated in Section III and discussed in Section IV. Finally, we conclude the paper in Section V.

II. Materials and Methodology

An overview of our seizure detection system is stated in this section. The UL module includes aEEG extraction, anomaly detection, adaptive segmentation and anomaly detection evaluation. Once a recording is classified as a dubious recording, all data from the same subject will be fed to the SL module for further analysis. The SL module includes feature extraction

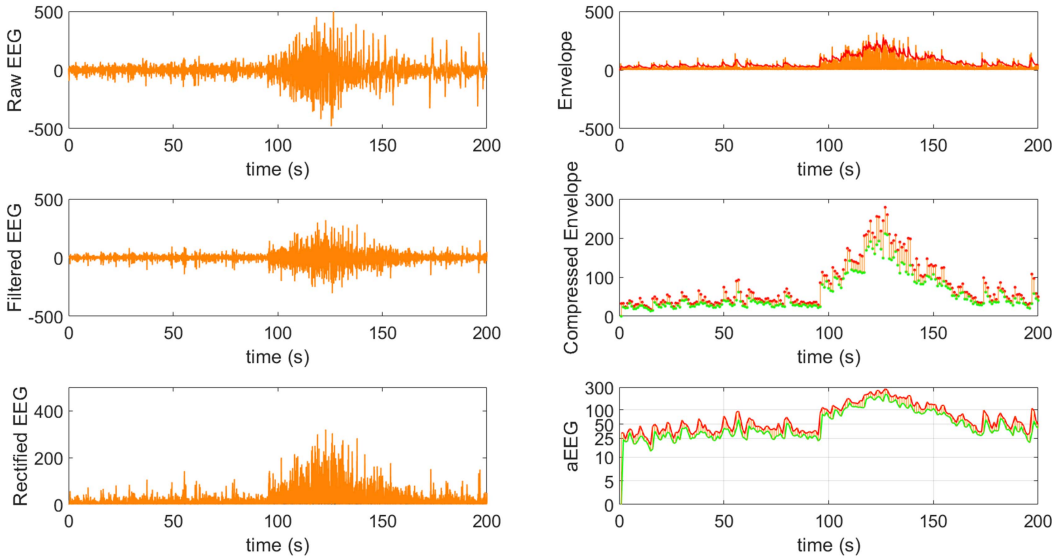

Fig. 2. Schematic diagram for extracting aEEG:(a) raw EEG, (b) filtered EEG, (c) rectified EEG, (d) envelope signal (Red line), (e) the upper margins (red point) and the lower margins (green point) of compressed envelope, and (f) aEEG signal.

and EasyEnsemble. The flowchart of our proposed method is shown in Fig. 1.

A. Datasets

The CHB-MIT dataset, including 24 patients, was collected from the Children's Hospital Boston [14], [15]. It is a widely used dataset for evaluating the performance of automatic seizure detection algorithms. The recordings with less than two seizures were excluded from our work. 18 channels that all patients have, including FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ and CZ-PZ, were used in our work.

B. UL Module

The UL module performs preliminary screening on the input data, and distinguishes indeterminate subjects from the determinate subjects.

1) aEEG Extraction: aEEG, as a compressed transformation of EEG signals, provides a global overview of brain cerebral activity and has been widely applied in seizure detection [16], [17]. Many commercial EEG devices, such as CerebraLogik, Brainz, and UniqueC, can export EEG and aEEG [18]. As a type of time-domain feature of EEG, aEEG is far less time-consuming for visual inspection than traditional EEG features and is easy to interpret even for non-specialists. We employed aEEG in our work.

The procedure of aEEG extraction includes four steps: asymmetrical filtering, rectification, envelope detection, and time compression [18]. Details of each step are introduced as follows:

a) Asymmetrical filtering: An asymmetric bandpass filter based on the Park-McClellan algorithm is designed. Within the passband (2-15 Hz), the amplification of signals increases with a slope of 12 dB per decade, while frequencies below 2 Hz and above 15 Hz are heavily attenuated. The signal after asymmetrical filtering is shown in Fig. 2(b).

b) Rectification: The filtered signal is rectified by converting negative voltages to positive values. The rectified signal is shown in Fig. 2(c).

c) Envelope detection: The envelope signal is acquired by squaring and utilizing the low-pass filtering-based envelope detection method with a first-order to produce an approximate smooth curve, representing the trend of the raw EEG data amplitude variations. A first-order Butterworth filter with a cut-off frequency of 0.375 Hz is applied. The envelope signal is shown as the red line in Fig. 2(d).

d) Time compression: To highlight the macro trend, time compression was achieved by compressing the tracing in the time axis. The sampling rate in the CHB-MIT dataset is 256 Hz, the compression rate is 256:1 on the time axis. Then, each epoch's maximum and minimum amplitudes (every 256 points) were extracted as the upper and the lower terminal points of the associated aEEG vertical line. To reduce the dynamic range of large fluctuations, amplitude integration is applied by performing logarithm operation on the values between 10 UV and 100 UV. After the whole transformation procedure, aEEG features can effectively enlarge the spike shape wave to identify epileptoid discharge. The signal after time compression is shown in Fig. 2(e). The upper margin of aEEG tracings (as shown in Fig. 2(f)) is used for further analysis.

2) Anomaly Detection: After extracting aEEG tracing, the next step is to detect the seizure epoch using anomaly detection. Anomaly detection, an effective method to solve two-class imbalanced classification tasks without any training process, has not been systematically investigated in the context of seizure detection yet [19]. In our work, anomaly detection mainly includes isolation forest and adaptive segmentation.

The isolation forest, with linear computational complexity, is an efficient anomaly detector based on the concept that

anomaly sample points are closer to the tree's root than normal ones [20]. The isolation forest algorithm involves two stages:

a) Modeling stage: An iTree is a complete binary tree, and its child nodes can only be 0 or 2. An iTree isolates each sample to a leaf node after building the iTree. Let X be the transformed aEEG after feature extraction.

The main modeling stage involves two steps: 1) To construct an iTree, each attribute q from X and a split value p generated between the maximum and minimum values of the attribute q in the data are randomly selected. If the data value in attribute p is less than q, it will be placed on the left child tree of the current tree node. Otherwise, it will be assigned to the right subtree. 2) Repeating the previous step to construct the left and right subtree, until either: (a) the tree reaches an artificially designed height limit, (b) $|X| = 1$, or (c) all data in X have the same values. By repeating the above steps to produce n iTrees, we can construct the iForest anomaly detection model.

b) Evaluating stage: The path length $h(x)$ is a metric of the degree of susceptibility to isolation. $h(x)$ is measured by the number of edges M_j traverses an iTree from the root node until the terminating node. The length of the path determines whether the data is an anomaly or not. A deep path length represents measurement sample is more likely to be classified as normal instances, whereas the shorter path means low susceptibility of isolation.

For any anomaly detection, an anomaly score is required for evaluating the degree of anomaly. The difficulty of obtaining an iTree score is that the complexities of the increase in the maximum feasible height and the average possible height are different. The former complexity is in the order of n, while the latter one is in the order of $log(n)$. Therefore, the normalization of $h(x)$ by any of the above terms cannot be directly compared. Due to the equivalent structure to Binary Search Tree (BST), the average $h(x)$ estimation is the same as BST. Therefore, the estimation of the average path length method for BST is used in iTree. The average path length of unsuccessful search in BST is given in Equation (1):

$$c(n) = 2H(n-1) - (2(n-1)/n) \tag{1}$$

where the harmonic number $H(i)$ can be estimated by the sum of $ln(i)$ and Euler's constant ε ($\varepsilon = 0.5772156649$). The $c(n)$, the average of $h(x)$ given n, is used for normalizing $h(x)$. The anomaly score s of sample x is calculated by:

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

where $E(h(x))$ is the average of path length from a collection of iTrees.

3) Adaptive Segmentation: Because an epileptic seizure is continuous for dozens of seconds, and the number of seizure are not fixed for each recording, setting the anomaly proportion manually to determine abnormal point is not suited for seizure detection problems. Therefore, we design an adaptive segmentation used for abnormal segment detection problems with unfixed anomaly proportions.

After obtaining the anomaly score s, when multiple high s values (potential abnormal sample) appear continuously, the potential seizure can be detected. In addition, we divide the seizure sample into starting sample of seizure, the intermediate state of seizure and the terminal sample of seizure. For each recording, each anomaly score vector can be obtained. The main steps of adaptive segmentation are as follows. (1) The anomaly score of high-frequency noise (caused by, for example, swallowing, frown) is high but only lasts a few seconds. Therefore, a moving average filter is applied to prevent the classification of the segment with a high score but short length as seizure samples. (2) The six local maximum in anomaly score are searched and ranked by the amplitude. The onset for one recording is at most six times. The global maximum (the maximum in local maximum) is the first intermediate state sample. Because the seizures in the same patient often belong to the same type, their anomaly scores are likely similar. Accordingly, the rest global maximum higher than 90% of the global maximum will be recognized as intermediate state samples. (3) After recognizing the intermediate state samples, the starting and terminal samples of a seizure can be determined by identifying a vector's front and back valleys using a differential coefficient. Finally, the seizure samples and normal samples can be obtained as the UL module result.

4) Anomaly Detection Evaluation: Anomaly detection evaluation: The cluster results in seizure/normal samples can be obtained after adaptive segmentation. However, the cluster performance cannot be guaranteed because of missing the training procedure. Therefore, the anomaly detection evaluation indicator is adopted to select those indeterminate subjects for further analysis using the SL method. These indeterminate subjects have probably low performance.

The clustering evaluation index, SC, is widely applied to calculate the performance of a clustering technique [21]. It is also used for clustering quality evaluation in anomaly detection. SC can be calculated by the average of intra-cluster distance a and the average of nearest-cluster distance b for each sample, as shown in Equation (3) [22]. Usually, the SC value ranges from -1 to 1. An SC value closing to 1 represents an excellent clustering performance [23]. An SC value closing to -1 represents an unreliable clustering performance [24].

$$SC = (b-a)/max(a, b) \tag{3}$$

After calculating the SC of each recording for all patients, the minimum of SC for each patient, denoted by Min(SC), is used for anomaly detection evaluation. Additionally, we set the threshold value of Min(SC), namely Thr(SC) (see Eq. (4)), as the mean of all Min(SC) values from all patients. When Min(SC) for a certain subject is higher than the threshold, the data were assigned into Group U (determinate subjects). The results using these data are more likely to show a high SDR. When the Min(SC) of a certain subject is lower than Thr(SC), data from this subject were assigned into Group S (indeterminate subjects), and a continuous data process using SL was implemented in this case to improve the sensitivity further. In addition, we implemented additional experiments by setting Thr(SC) from 0 to 0.5 to evaluate the rationality of setting Thr(SC).

$$Thr_{Min(SC)} = mean(Min(SC)) \tag{4}$$

C. SL Module

The SL module mainly includes data pre-processing, feature extraction (extracting nine types of features based on four sub-bands) and EasyEnsemble, as shown in Fig. 1.

1) Data Processing: The EEG data is filtered from 0.1-32 Hz using a 5-order Butterworth filter. Then, each EEG recording is divided into 5s EEG signals as segments for binary classification.

2) Feature Extraction: The delta (0-4 Hz), theta (4-8 Hz), alpha (8-16 Hz), and beta (16-32 Hz) bands are attractive in clinical application [25]. We use discrete wavelet transforms (DWT) with the Daubechies order-1 (db1) wavelet to decompose the original signal into four sub-band signals. Then, nine various types of features, including absolute mean, skewness, kurtosis, line length, variance, energy, Teager-Kaiser, peak-to-peak amplitude and average power, were extracted in each sub-band.

3) Classification: EasyEnsemble is an effective method for the class imbalance problem, which focuses on minority class by generating T relative balanced subproblems [26]. The procedures of EasyEnsemble are as follows: Firstly, it divides the whole seizure-free data into T equal parts. Afterwards, it combines the whole seizure data and each $1/T$ seizure-free data. The final decision is obtained by combining results obtained from AdaBoost. Compared with common methods used in the anomaly detection module (such as over-sampling and threshold moving), the EasyEnsemble demonstrates its benefit in decreasing information loss from random under-sampling [26]. The classification result is used for evaluating seizure epoch recognition ability.

4) Post-Processing: Because all epileptic seizures last at least six seconds, a seizure event typically consists of at least two seizure epochs. As a result, in order to minimize FAR, the SL module's post-processing is utilized, which considers more than two successive epochs as seizure epochs as a single seizure event.

D. Performance Evaluation

In this study, we employed five performance assessment indices: accuracy, sensitivity, specificity, SDR, and false alarm rate (FAR). The accuracy, sensitivity, and specificity are common segment-based level evaluation indicators. SDR and FAR are event-based performance indicators. SDR denotes the number of detected seizure events to the total number of seizure events. The signal from the onset of a seizure to the finish of a seizure is generally considered an event. Once two consecutive seizure epochs in one seizure event are correctly identified, we defined that this event is correctly identified both for UL and SL modules. FAR denotes the number of false alarms reported in an hour.

In our work, two-fold recording-independent cross-validation is applied for performance estimation. Specifically, all recordings for the ith patient is divided into two sections. One is used for training, and the other is used for testing. The same procedure repeats with the training and testing roles of the two sections switching. The training rate for the SL module is 50%. The training rate (TR) is calculated according to the following formula.

$$TR = \frac{NS}{TNP} * 50\% \tag{5}$$

where NS presents the number of patients in Group S, and TNP is the total number of patients.

To evaluate the UL module and SL module, we perform the additional two experiments on these two single modules in all 24 patients and the Wilcoxon test is used to compare their performance.

TABLE I

THE PERFORMANCE OF THE PROPOSED ALGORITHM EVALUATED ON CHB-MIT (ACC: ACCURACY, SEN: SENSITIVITY, SPEC: SPECIFICITY. THE RESULT OF PATIENTS IN BOLD PRINT (GROUP U) ARE OBTAINED BY UL METHOD. THE PATIENTS MARKED WITH UNDERLINE (GROUP S) ARE OBTAINED BY SL METHOD)

Patient	Acc (%)	Sen (%)	Spec (%)	SDR (%)	FAR (/h)
1	86.42	100	86.26	100	0.57
2	73.49	100	72.43	100	1.91
3	90.24	100	90.07	100	0.71
4	88.14	98.33	88.09	100	0.37
5	97.48	97.59	97.48	100	0
6	92.39	91.82	92.39	100	2.84
7	97.90	93.29	97.95	100	0.44
8	96.41	96.07	96.50	100	0.2
9	97.08	100	97.06	100	0.1
10	99.01	97.62	99.03	100	0.34
11	91.66	100	91.43	100	1.07
12	95.99	82.25	96.35	100	0.67
13	87.65	85.93	87.68	100	4.39
14	92.18	94.12	92.17	100	2.94
15	98.30	96.40	98.39	100	0.29
16	89.20	100	89.17	100	5.02
17	94.48	97.94	94.38	100	2.12
18	87.74	86.61	87.75	100	3.04
19	97.20	91.76	97.33	100	0
20	96.45	97.41	96.43	100	1.42
21	96.05	98.53	96.01	100	1.06
22	97.41	99.11	97.38	100	0
23	88.26	100	88.12	100	0.63
24	91.77	88.51	91.82	100	3.01
Mean	92.62	95.55	92.57	100	1.38

III. RESULTS

The accuracy, sensitivity, specificity, SDR and FAR of each patient and the mean of the 24 patients for the whole system combining the UL module and the SL module are summarized in Table I. Here Thr(SC) is set to be 0.19 based on the calculation by Eq. (4). The patients in bold print in Table I are those Min(SC) values that are higher than 0.19 to be assigned to Group U (Group U: patient 1, 2, 3, 4, 5, 7, 8, 9, 11, 19, 22, 23). The results of Group U in Table I are obtained using the UL method. The remaining patients are those Min(SC) are lower than 0.19 to be assigned to Group S (Group S: patients 6, 10, 12, 13, 14, 15, 16, 17, 18, 20, 21, 24). The result of Group S is obtained using the SL method. Due to 12 patients

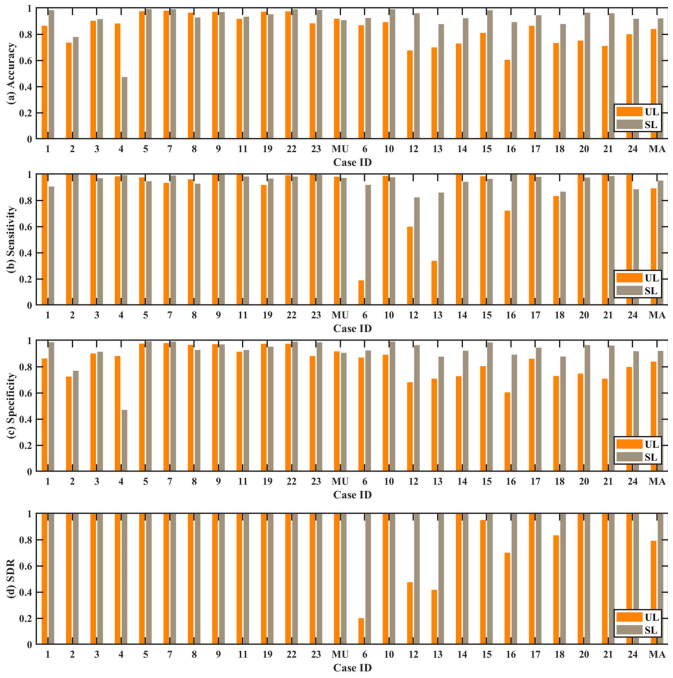

Fig. 3. The accuracy, sensitivity, specificity and SDR of the 24 patients using UL and SL methods (MU: mean result of Group U, MA: mean result of the 24 patients).

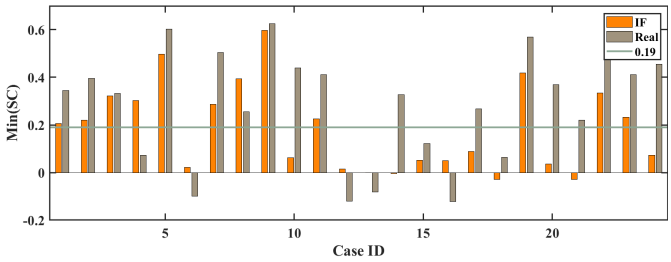

Fig. 4. The Min(SC) for the 24 patients obtained by detection results and ground truth label (The orange bar and the gray bar represent the Min(SC) value calculated by ground truth label and detection results, respectively; the green line around 0.2 are the Thr(SC)).

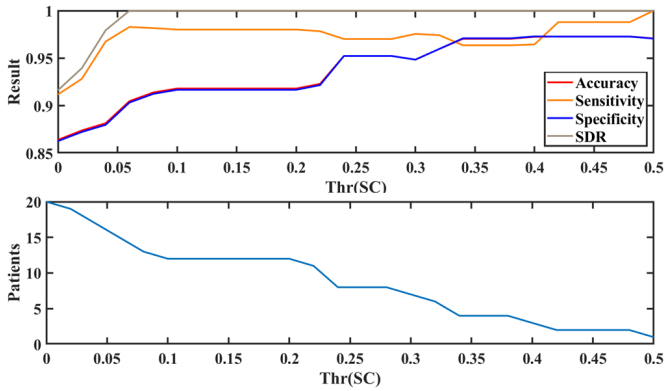

Fig. 5. The changing trend of accuracy, sensitivity, specificity, and SDR, the number of selected patients (the number of patients in Group U) when the Thr(SC) changes from 0 to 0.5.

in Group U, the training rate of the overall system is 25%, according to Eq.(5). We achieved an accuracy of 92.62% ± 1.17%, a sensitivity of 95.55% ± 1.06%, a specificity of 92.57% ± 1.21%, an SDR of 100% and an FAR of 1.38 (/h) with 25% of training rate.

To compare the performance of the UL method and the SL method, the UL and SL modules are applied on the 24 patients, and the accuracy, sensitivity, specificity, and SDR are shown in Fig. 3. The results of Group U indicate that the performance of the UL method is slightly better than the SL method, but the performance of these two methods have no statistical differences in term of accuracy ($p>0.05$), sensitivity ($p>0.05$) and specificity ($p>0.05$). The results of Group S in Figs. I (a), (c) and (d) show that using the SL method in Group S contributes to better performance than using the UL method, especially for accuracy ($p<0.05$), specificity ($p<0.05$) and SDR ($p<0.05$).

The Min(SC) score for each patient calculated by the detection results and the ground truth label are shown in Fig. 4. As shown in Fig. 4, the Min(SC) of patients 6, 12, 13, and 16 for the ground truth label are less than 0. Negative values in SC value generally indicate that a sample does not assign to the right cluster. Meanwhile, the SDRs of these four patients are the lowest among the 24 patients and all lower than 50%. In addition, the patients (e.g., patients 14, 17, 20, 21, and 24) have an SDR of 100%, but their Min(SC) obtained by isolation forest is lower than the Thr(SC). These five patients' specificity values are lower than 88%.

Fig. 5 shows the accuracy, sensitivity, specificity, SDR and the number of selected patients (the number of patients in Group U) when the threshold value of Min(SC) increases from 0 to 0.5. Because the specificity and accuracy have similar values in seizure detection tasks (imbalance problem), they are very close. Sensitivity and SDR are similar evaluation indices where specificity is for seizure segments detection and SDR is for seizure events detection. Therefore, they are very close to each other. There are four observations from Fig. 5: (1) When the threshold of Min(SC) increases from 0 to 0.5, the number of selected patients decreases from 20 to 1. (2) The overall trend is firmly upwards for accuracy, sensitivity, specificity and SDR when the threshold increases from 0 to 0.06. (3) When the threshold increases from 0.06 to 0.5, the accuracy, sensitivity, specificity, and SDR have no evident changes. (4) When the threshold of Min(SC) increases from 0.1 to 0.2, the accuracy, sensitivity, specificity, SDR and the number of the selected patients do not have any change. These findings indicate that when Thr(SC) ranges from 0.06 to 0.2, the present result and the number of patients in Group U can be maintained.

IV. Discussion

In recent years, automatic seizure detection methods have been extensively studied. These methods can be generally divided into two categories: the SL-based and the UL-based methods. For the SL-based methods, most previous studies demonstrated superiority in terms of the sensitivity and specificity in patient-dependent evaluation with a high training rate. Although high performance can be achieved, these studies

TABLE II
COMPARISON OF EXISTING METHODS AND THE PROPOSED METHOD EVALUATED ON CHB-MIT DATASET (TR: TRAINING RATE; INTER-RECORDING (IR): THE TRAINING SET AND TESTING SET COME FROM DIFFERENT RECORDINGS; -: NO REPORTED VALUES OR NOT APPLICABLE)

Author (Year)	Method	Patient	TR	IR	Acc (%)	Sen (%)	Spec (%)	SDR (%)	FAR (/h)
Peng *et al.* (2021) [27]	dictionary learning and sparse representation	24	90%	No	95.06	95.38	94.33	-	-
Chakrabarti *et al.* (2020) [28]	DWT and ANN	10	90%	No	95.3	97.2	93.5	-	-
Zarei *et al.* (2021) [6]	DWT, orthogonal matching pursuit and SVM	23 (24)	90%	No	97.09	96.81	92.36	-	-
Jiang *et al.* (2020) [29]	symplectic geometry eigenvalues and SVM	22 (12, 24)	90%	No	99.72	97.17	99.62	-	-
Tian *et al.* (2019) [7]	CNN and multi-view TSK fuzzy system	24	90%	No	98.43	96.51	98.95	-	-
Selvakumari *et al.* (2019) [5]	high dimensional phase space, hybrid classifier	24	50%	No	96.28	94.5	97.5	-	-
Zabihi *et al.* (2019) [30]	nullcline analysis, LDA	23 (15)	25%	Yes	95.11	91.15	95.16	-	-
Bhattacharyya *et al.* (2017) [11]	empirical wavelet transform	23 (12)	50%	No	99.41	97.91	99.57	-	-
Tian *et al.* 2019 [7]	multi-view CNN model	24	50%	No	97.16	94.68	98.40	-	-
Li *et al.* (2020) [31]	Nested LSTM	23 (24)	-	Yes	95.29	95.43	95.29	94.07	0.66
Khanmohammadi *et al.* (2017) [25]	PCA, CSP and adaptive distance-based change point detector	23 (24)	-	Yes	-	-	-	96	0.12
Tang *et al.* (2020) [32]	PCA, CSP, multivariate multiscale sample entropy and SVM	20 (4, 6, 12, 16)	-	Yes	97.8	-	-	97.2	0.64
Our work	time-frequency feature and EasyEnsemble	24	90%	No	97.74	98.25	97.73	-	-
Our work	aEEG, isolation forest, time-frequency feature and EasyEnsemble	24	25%	Yes	92.62	95.55	92.57	100	1.38

also bring out a significant issue that the practical use of these methods may be constrained with the requirement on a large training dataset. A few studies also adopted a relatively small training rate to verify their performance. However, the performance also sharply decreases, especially for the sensitivity [33]. The UL-based method is an emerging method, and its superiority is without requiring labelling data. However, it has an intractable defect, where a high SDR (>90%) is difficult to obtain. Therefore, combining UL and SL methods is a novel and alternative idea in automatic seizure detection. However, there is a problem to be solved: how to combine UL and SL modules effectively. Fortunately, the internal evaluation index SC can effectively select indeterminate subjects, thereby combining the UL and SL modules.

Only a few studies have employed UL methods for seizure detection. Tsiouris *et al.* [12] designed a rule-based seizure detection logic method based on four conditions for identifying segments where the EEG signal energy is accumulated intensively among the delta (< 4 Hz), theta (4–7 Hz) and alpha (8–13 Hz) frequency bands, achieving an SDR of 88% with 5% inspection value [12]. Compared with this work, our proposed UL method based on aEEG and isolation forest does not require predefined rules and parameters while achieving a higher performance (accuracy: 83.96%, sensitivity: 89.20%, specificity: 89.89%) using all recording data (13 recordings of patient 12 are excluded in [12]). However, the current performance still is lower than SL methods. As shown in Fig 3, the accuracy, specificity and SDR of the SL method are

significantly better than the UL method for Group S, which shows that the UL module is difficult or impossible to use alone.

In addition, the rationality of the internal evaluation index SC needs to be further explored. We can observe from Fig. 3 that patients 6, 12, 13, and 16 have a poor SDR in the UL module. Many works proved that seizure segments of these four patients are very difficult to identify [12], [33], [34]. The reason may concluded as follows: (1) The proportion of seizure-free segments with noise and artifact is very high compared with the other patients; (2) If seizure duration is too short, the training model will classify it as seizure-free epoch with noise. The average seizure length for patient 16 is just 8.4 seconds, with the lowest epoch lasting only 6 seconds. Furthermore, as shown in Fig. 4, the Min(SC) value derived by the ground truth label for these four patients (patients 6, 12, 13, and 16) is less than 0. Meanwhile, the performance of these four patients obtained by the UL module is also dissatisfactory. These findings suggest SC value for these four indeterminate subjects is significantly correlated with the classification result. In addition, the patients, including patients 14, 17, 20, 21, and 24 have an SDR of 100%, but their Min(SC) obtained by isolation forest is lower than the Thr(SC). These five patients' specificity values are lower than 88%. That is because a lot of seizure-free epochs were misdetected as seizures, decreasing the specificity and Min(SC) for this type of patient. Besides, calculating the difference between Min(SC) by predicted labels (Fig. 4) and Min(SC) by the ground truth label (Fig. 4) can judge whether the result is accurate. For example, the difference for patients 10, 14, 20, 21, 24 is more than 0.25, and their accuracy is lower than 90%. When the difference between Min(SC) obtained by detection results and the ground truth label is lower, the detection results are relatively accurate. These contents demonstrate that the SC in our system is an effective evaluation index in primary screening indistinguishable patients. Besides, the changing trend in Fig. 5 shows that when the threshold value changes between 0.06 and 2, the performance and selected patients only slightly change, which proves that the performance is not limited by Thr(SC).

To evaluate the performance of the SL method, 10-fold cross-validation was performed. The comparison of the results of our SL method and five comparative studies evaluated on the same CHB-MIT dataset using the 10-fold cross-validation published from 2019 to 2020 are presented in Table II [6], [27]–[29], [7]. These works with a 90% training rate all are used 10-fold cross-validation. The sparse representation [27], artificial neural network (ANN) [28], SVM [6], [29], multi-view TSK fuzzy system [7] were used or designed to improve the discriminant ability of the seizure detection system. However, the EasyEnsemble classifier differs from classifiers mentioned above, which can make full use of potential useful seizure-free epochs by randomly and repeatedly sampling majority classes to solve imbalance tasks. As a result, the proposed SL module can achieve the highest sensitivity (98.25%) compared with the other five works ($< 97.5\%$).

To evaluate the performance of our overall seizure detection method, the results of state-of-the-art methods using leave-one-record-out cross-validation are shown in Table II for comparison [25], [31], [32]. These three related works are based on leave-one-record-out cross-validation. These works applied different post-processing for reducing FAR. Li *et al.* used multi-channel integration, moving average filter, and collar technique to remove sporadic false detection, decreasing FAR. Tang *et al.* set the necessary time for continuous monitoring as three epochs. Although our FAR is lower than these three works, we can achieve an SDR of 100%. Actually, when we adjust the necessary time for continuous monitoring in post-processing as three epochs, the corresponding SDR and FAR will change as 97.5% and 0.91(/h). This finding indicates that post-processing can adjust the trade-off between SDR and FAR.

TABLE III
COMPARISON OF DIFFERENT FEATURE EXTRACTION METHODS USING THE UL MODULE FOR 24 PATIENTS AND GROUP U PATIENTS

Patients	Feature	Acc (%)	Sen (%)	Spec (%)
24 patients	aEEG	83.96	89.20	83.36
	Time-frequency features	92.25	76.74	91.68
Group U	aEEG	91.81	98.01	91.67
	Time-frequency features	94.88	91.20	95.33

To evaluate the performance of our overall seizure detection method, the results of four state-of-the-art methods with a training rate of less than or equal to 50% are shown in Table II for comparison [5], [7], [11], [30]. These studies designed different structures/models to reduce the performance degradation caused by the decrease of the training rate. Our system obtained much higher sensitivity than works using the same training rate [30]. Although the work by Zabihi *et al*. [30] achieved a higher accuracy (95.11%) and specificity (95.16%), it excludes patient 15 and do not use cross-validation. In [5], [11] and [7], with a training rate of 50%, high accuracy and specificity were obtained, while lower sensitivity was presented as compared to our model evaluated with a training data of 25%.

Compared with these previous works, we use a different view in which the partial patients are fed into the SL module, and the remaining patients using the UL method have no training procedures, reducing the training rate of the entire database. The side effect of the previous view is that when training data are less for a patient-specific model, the sensitivity would have a high probability of drastic reduction. For example, patient 22 only has three times seizures. When only one seizure is fed into the classifier as training data, the final sensitivity is much lower than the case with two seizures as the training set [30]. However, our results prove that adding an unsupervised learning method in an automatic seizure detection system is an effective and novel idea to obtain superior performance than the state-of-the-art methods and decrease the overall system's training data.

A lot of ensemble methods in epilepsy seizure detection tasks often use the same feature extraction strategy in various

modules [35]. Our system uses two different feature extraction methods for three reasons. Firstly, the function of the UL module in our system is designed for the preliminary screen, which can quickly determinate those determinate subjects whose seizure epoch can be clustered using isolation forest. We hope the feature extraction part of the UL module is relatively simple, highly efficient, and easy-to-understand. It means the features that can be easily exported or quickly extracted are the best choice. For aEEG, it meets all our requirements. aEEG can be quickly exported from a lot of EEG devices like CerebraLogik, Brainz, and UniqueC. aEEG also can be easily and quickly extracted, as demonstrated in our previous work [18]. Meanwhile, physician are more knowledgeable about the relationships between the shifting trend of the aEEG and the seizure state. Thus, aEEG is served as an appropriate feature in the UL module. Secondly, we have also added an experiment to compare the performance of the UL module using aEEG and time-domain features. The results are listed in TABLE III. The time-frequency features indicated in TABLE III, including the absolute mean, skewness, kurtosis, line length, variance, energy, Teager-Kaiser, peak-to-peak amplitude, and average power, are also employed in the SL module. The results show that aEEG can obtain higher sensitivity in the UL module. Finally, although aEEG is already obtained by the UL module, aEEG is not suitable for the SL module. The reasons are shown as follows: asymmetrical filtering (an important step of obtaining aEEG) requires that the frequency of input data must cover 2-15 Hz. There are no sub-band (delta, theta, alpha, and beta) for the SL module that its frequency can cover 2-15Hz. Therefore, aEEG cannot be extracted in the SL module. The above reasons result in two different feature extraction methods used in our system.

However, the current system still has some limitations, which will be addressed in our future work. Firstly, due to the missing specific seizure type information in the CHB-MIT dataset, we cannot obtain the link of result of seizure type. In the future, more clinical data will be collected to further verify our proposed method. Based on these clinical data, the result of different seizure type data using UL and SL methods will be analyzed. Secondly, several subjects using UL method can obtain 100% of SDR, but they still were classified into Group S using SC. In addition, the highest SC calculated by the ground truth label is far below 1. All internal cluster validity indices, including SC, are mainly proposed for dealing with data-balanced tasks. Therefore, the internal cluster validity indices used for seizure detection problems will be designed in the future.

V. Conclusion

This paper proposes a hybrid method that utilizes a small amount of labelled data for model training while achieving satisfactory seizure detection performance. It combines both the UL and the SL methods, where the UL method is adopted as the preliminary screening to distinguish determinate subjects and indeterminate subjects, and the SL method is performed as the enhanced classification of the indeterminate subjects. To our best knowledge, it is the first work synthesizing UL and SL methods in seizure detection. Meanwhile, it is also the first work to utilize the aEEG and isolation forest for prescreening. The proposed method based on the UL and the SL is evaluated on 24 patients from CHB-MIT, achieving an accuracy of 92.62%, a sensitivity of 95.55% and a specificity of 92.57% with a training rate of 25%. Extensive experimental results demonstrate that the proposed method outperforms other state-of-the-art methods. The present study on seizure detection provides a practical option based on sEEG signals in clinical diagnosis.

References

[1] M. J. Brodie and J. A. French, "Management of epilepsy in adolescents and adults," *Lancet*, vol. 356, no. 9226, pp. 323–329, Jul. 2000.

[2] X. Jiang, K. Xu, R. Zhang, H. Ren, and W. Chen, "Redundancy removed dual-tree discrete wavelet transform to construct compact representations for automated seizure detection," *Appl. Sci.*, vol. 9, no. 23, p. 5215, Nov. 2019.

[3] L. Wang, X. Long, R. M. Aarts, J. P. van Dijk, and J. B. A. M. Arends, "A broadband method of quantifying phase synchronization for discriminating seizure EEG signals," *Biomed. Signal Process. Control*, vol. 52, pp. 371–383, Jul. 2019.

[4] L. Wang, X. Long, J. B. A. M. Arends, and R. M. Aarts, "EEG analysis of seizure patterns using visibility graphs for detection of generalized seizures," *J. Neurosci. Methods*, vol. 290, pp. 85–94, Oct. 2017.

[5] R. S. Selvakumari, M. Mahalakshmi, and P. Prashalee, "Patient-specific seizure detection method using hybrid classifier with optimized electrodes," *J. Med. Syst.*, vol. 43, no. 5, p. 121, May 2019.

[6] A. Zarei and B. M. Asl, "Automatic seizure detection using orthogonal matching pursuit, discrete wavelet transform, and entropy based features of EEG signals," *Comput. Biol. Med.*, vol. 131, Apr. 2021, Art. no. 104250.

[7] X. Tian *et al.*, "Deep multi-view feature learning for EEG-based epileptic seizure detection," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 10, pp. 1962–1972, Oct. 2019.

[8] Y. Jiang *et al.*, "Recognition of epileptic EEG signals using a novel multiview TSK fuzzy system," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 1, pp. 3–20, Feb. 2017.

[9] Y. Jiang *et al.*, "Smart diagnosis: A multiple-source transfer TSK fuzzy system for EEG seizure identification," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 16, no. 2, pp. 1–21, Jul. 2020.

[10] Y. Jiang *et al.*, "Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 12, pp. 2270–2284, Dec. 2017.

[11] A. Bhattacharyya and R. B. Pachori, "A multivariate approach for patient-specific EEG seizure detection using empirical wavelet transform," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2003–2015, Sep. 2017.

[12] K. M. Tsiouris, S. Markoula, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, "A robust unsupervised epileptic seizure detection methodology to accelerate large EEG database evaluation," *Biomed. Signal Process. Control*, vol. 40, pp. 275–285, Feb. 2018.

[13] S. You *et al.*, "Unsupervised automatic seizure detection for focal-onset seizures recorded with behind-the-ear EEG using an anomaly-detecting generative adversarial network," *Comput. Methods Programs Biomed.*, vol. 193, Sep. 2020, Art. no. 105472.

[14] H. Daoud and M. A. Bayoumi, "Efficient epileptic seizure prediction based on deep learning," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 5, pp. 804–813, Oct. 2019.

[15] Y. Zhang, Y. Guo, P. Yang, W. Chen, and B. Lo, "Epilepsy seizure prediction on EEG using common spatial pattern and convolutional neural network," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 465–474, Feb. 2020.

[16] C. Chen, C. Sun, S. Leonhardt, P. Andriessen, H. Niemarkt, and W. Chen, "Amplitude-integrated electroencephalography applications and algorithms in neonates: A systematic review," *IEEE Access*, vol. 7, pp. 141766–141781, 2019.

[17] J. Wang, R. Ju, Y. Chen, G. Liu, and Z. Yi, "Automated diagnosis of neonatal encephalopathy on aEEG using deep neural networks," *Neurocomputing*, vol. 398, pp. 95–107, Jul. 2020.

[18] C. Chen *et al.*, "A digitized approach for amplitude-integrated electroencephalogram transformation towards a standardized procedure," *Biomed. Signal Process. Control*, vol. 66, Apr. 2021, Art. no. 102433.

[19] Y. Zhao, Z. Nasrullah, and Z. Li, "PyOD: A Python toolbox for scalable outlier detection," 2019, *arXiv:1901.01588*.
[20] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proc. 8th IEEE Int. Conf. Data Mining*, Dec. 2008, pp. 413–422.
[21] Z. Lin, F. Wen, and Y. Ding, "Data-driven coherency identification for generators based on spectral clustering," *IEEE Trans. Ind. Informat.*, vol. 14, no. 3, pp. 1275–1285, Mar. 2018.
[22] A. Granados, K. Koroutchev, and F. D. B. Rodriguez, "Discovering data set nature through algorithmic clustering based on string compression," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 3, pp. 699–711, Mar. 2015.
[23] J. Li, R. Huang, G. He, Y. Liao, Z. Wang, and W. Li, "A two-stage transfer adversarial network for intelligent fault diagnosis of rotating machinery with multiple new faults," *IEEE/ASME Trans. Mechatronics*, vol. 26, no. 3, pp. 1591–1601, Jun. 2021.
[24] B. Trevizan, J. Chamby-Diaz, A. L. C. Bazzan, and M. Recamonde-Mendoza, "A comparative evaluation of aggregation methods for machine learning over vertically partitioned data," *Expert Syst. Appl.*, vol. 152, Aug. 2020, Art. no. 113406.
[25] S. Khanmohammadi and C.-A. Chou, "Adaptive seizure onset detection framework using a hybrid PCA–CSP approach," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 1, pp. 154–160, Jan. 2018.
[26] T.-Y. Liu, "EasyEnsemble and feature selection for imbalance data sets," in *Proc. Int. Joint Conf. Bioinf., Syst. Biol. Intell. Comput.*, 2009, pp. 517–520.
[27] H. Peng *et al.*, "A novel automatic classification detection for epileptic seizure based on dictionary learning and sparse representation," *Neurocomputing*, vol. 424, pp. 179–192, Feb. 2021.
[28] S. Chakrabarti, A. Swetapadma, A. Ranjan, and P. K. Pattnaik, "Time domain implementation of pediatric epileptic seizure detection system for enhancing the performance of detection and easy monitoring of pediatric patients," *Biomed. Signal Process. Control*, vol. 59, May 2020, Art. no. 101930.
[29] Y. Jiang, W. Chen, and M. Li, "Symplectic geometry decomposition-based features for automatic epileptic seizure detection," *Comput. Biol. Med.*, vol. 116, Jan. 2020, Art. no. 103549.
[30] M. Zabihi, S. Kiranyaz, V. Jäntti, T. Lipping, and M. Gabbouj, "Patient-specific seizure detection using nonlinear dynamics and nullclines," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 543–555, Feb. 2020.
[31] Y. Li *et al.*, "Automatic seizure detection using fully convolutional nested LSTM," *Int. J. Neural Syst.*, vol. 30, no. 4, Apr. 2020, Art. no. 2050019.
[32] F.-G. Tang, Y. Liu, Y. Li, and Z.-W. Peng, "A unified multi-level spectral–temporal feature learning framework for patient-specific seizure onset detection in EEG signals," *Knowl.-Based Syst.*, vol. 205, Oct. 2020, Art. no. 106152.
[33] K. Samiee, P. Kovács, and M. Gabbouj, "Epileptic seizure detection in long-term EEG records using sparse rational decomposition and local Gabor binary patterns feature extraction," *Knowl.-Based Syst.*, vol. 118, pp. 228–240, Feb. 2017.
[34] X. Yao, X. Li, Q. Ye, Y. Huang, Q. Cheng, and G.-Q. Zhang, "A robust deep learning approach for automatic classification of seizures against non-seizures," *Biomed. Signal Process. Control*, vol. 64, Feb. 2021, Art. no. 102215.
[35] K. Abualsaud, M. Mahmuddin, M. Saleh, and A. Mohamed, "Ensemble classifier for epileptic seizure detection for imperfect EEG data," *Sci. World J.*, vol. 2015, pp. 1–15, Dec. 2015.