Ensembled Seizure Detection Based on Small Training Samples

Pei Feng Tong [ID], Hao Xiang Zhan [ID], and Song Xi Chen [ID]

Abstract—This paper proposes an interpretable ensembled seizure detection procedure using electroencephalography (EEG) data, which integrates data driven features and clinical knowledge while being robust against artifacts interference. The procedure is built on the spatially constrained independent component analysis supplemented by a knowledge enhanced sparse representation of seizure waveforms to extract seizure intensity and waveform features. Additionally, a multiple change point detection algorithm is implemented to overcome EEG signal's non-stationarity and to facilitate temporal feature aggregation. The selected features are then fed into a random forest classifier for ensembled seizure detection. Compared with existing methods, the proposed procedure has the ability to identify seizure onset periods using only a small proportion of training samples. Empirical evaluations on publicly available datasets demonstrated satisfactory and robust performance of the proposed procedure.

Index Terms—Change point detection, few-shot learning, expert system, independent component analysis.

I. INTRODUCTION

EPILEPSY is a chronic neurological disease marked by recurrent seizures which are abnormal brain activities associated with idiosyncratic symptoms [1], [2]. According to the World Health Organization, approximately 50 million people world-wide suffered from epilepsy [3], with 80% of them living in low and middle-income countries. Scalp Electroencephalography (EEG) is a common method used by neurologists for understanding the seizure types and locating the seizure source regions. Since long-term EEG monitoring requires much effort of neurologists, it would be quite helpful to detect seizure automatically via a trained algorithm to relieve doctors from the routine works of inspecting EEG data for diagnostic decision.

There have been studies on automated epileptic seizure detection as reviewed in Baumgartner and Koren [4] and Boonyakitanont et al. [5]. Most of the studies employed feature-based detection algorithms which first extracted features from segmented EEG fragments and then train classifiers to determine which fragments belong to the seizure periods [6], [7]. Bhattacharyya and Pachori [8] proposed a wavelet transform based algorithm for feature extraction with six classifiers to decide the seizure and seizure-free EEG segments. Li et al. [9] developed a multi-scale radial basis function network to deal with the high temporal resolution images and used the support vector machine (SVM) classifier for seizure detection. Wang et al. [10] applied a topological data analysis framework for obtaining epileptic EEG signals. Schröder and Ombao [11] proposed an algorithm which evaluated the seizure evolution over space and time and detected changes in the cross-coherence structure at some specific frequencies before a seizure onset. Sharma and Pachori [12] proposed a time-frequency representation based on eigenvalue decomposition of Hankel matrix and Hilbert transform and least-square SVM. Gupta and Pachori [13] performed Fourier–Bessel series expansion with EEG signals and calculated weighted multiscale Renyi permutation entropy as feature.

The field of automatic seizure detection has a history over 40 years with the goal of developing accurate patient independent seizure detection [14]. However, the performance of the automatic detection still have much room to improve as indicated in recent reviews [15], [16], [17]. A recent review [18] highlights the need to enhance algorithm robustness through the few-shot learning technique, which is a focus of our study.

Nevertheless, there are four major challenges for a more robust and practical seizure detection algorithm. The first challenge comes from the fact that most works employing the independent component analysis (ICA) [19] and related spatially constrained ICA (SCICA) [20] for artifacts removal rather than seizure signal acquisition [21], [22], [23], [24]. ICA and SCICA decompose EEG series to independent components (ICs) and then remove the artifact related ICs via inverse ICA transformation. The ICs identified from the training data may not have good adaptation to the testing data as artifacts may not be stably situated across time and subject domains.

The second challenge is a lack of sparse and informative time-frequency feature representation for seizure signals from EEG records. Many existing time-frequency based feature extraction methods employ complete sets of basis functions, such

as wavelet bases [7], [8], [13], [25], [26]. Using the complete bases tends to produce dense representations of seizure signals, which are usually accompanied by higher variance and reduced signal-to-noise ratio. This is quite contrary to the clinical practice of neurologists, who seek sparse representations by focusing on specific seizure related waveforms [27].

The third challenge comes from that the seizure detection algorithms are often trained on fixed length segments whose start and ending may not match those of the seizure or non-seizure periods, and inevitably generates mixed-statue segments. The mixed statue segments can cause a lower signal-to-noise ratio and lead to biased estimates to the seizure onset times. At the same time, simply remove these segments during training may cause an overestimation of accuracy rates.

The last challenge arises from the need to develop patient specific seizure detection method with small training samples in order to make the detection algorithm worthwhile from clinicians' perspective. Many existing algorithms are trained via the 5-10 folds cross validation, which means using 80% - 90% samples for training and potentially saving only 10%-20% of doctors' time. Training based on small samples (few-shot learning) is more desirable as promoted by the few-shot learning [28], and was emphasized in [18]. One way to realize effective small sample training is to utilize clinical knowledge in the algorithm [29], which is the approach we take.

Our proposals designed to overcome these challenges are made in four aspects. Firstly, the correlations between ICs and labeled data are used to select informative ICs. This offers a systematic IC selection that guides the transfer of learned ICs from training to testing data, and determines seizure related ICs for signal enhancement and artifacts removal. This is different from the existing methods whose use of the ICs were largely concerned with artifact removal. We select the seizure-related ICs by (i) calculating the correlation between the ICs and EEG signals during seizure onset periods, and (ii) formulating an energy ratio statistic between seizure onset and offset periods. The most relevant ICs are those who have high correlations and high energy ratios simultaneously.

The second aspect is in using clinically interpretable waveforms, trying to mimic the practice of neurologists when they examine EEG series. This knowledge based waveform learning generates sparse signal representation due to the use of a clinical knowledge-based waveform dictionary. This is different from the existing orthogonal matching pursuit (OMP) or wavelet based approaches, which largely rely on complete basis functions and hence generate denser and noisier signal representations as mentioned above.

The third aspect is in implementing a statistical change-point detection procedure to segment the denoised time series to stationary pieces after the ICs selection. This offers more flexibility than the existing fixed length segmentation used in training the classifiers. The flexible segment length allows signal enrichment and stronger contrast between seizure and non-seizure periods, leading to better signal discovery, and would be theoretically and practically useful for the signal processing community.

The last challenge to facilitate seizure detection with small training can be overcome by integrating the three proposals mentioned above. Without requiring more labeling efforts in the current clinical workflow, the proposed method is designed to use about one hour EEG records verified by neurologists as training samples, corresponding to a training rate of 5% in the CHB-MIT dataset [30], [31] and 25% in the Helsinki University Hospital (HUH) dataset [32] and the Temple University Hospital EEG Seizure Corpus (TUSZ) [33]. Empirical study demonstrated good performance with accuracies of 99.4%, 95.4% and 97.0% for the CHB-MIT, HUH and TUSZ datasets, respectively, and being able to locate the seizure onset regions with less false positive rates. Compared with existing studies, the proposed procedure uses the least training data while outperforms many of them trained with much higher training rates.

The paper is organized as follows. Section II describes three publicly available datasets and the study plan. Sections III and IV outline the feature extraction and artifact removal via the SCICA and a knowledge enhanced sparse waveform representation, respectively. Section V presents the change-point detection for changes in the covariance among the seizure-related ICs, which are used to generate adaptive segmentation for feature aggregation. The ensemble binary classifier based on the random forest is outlined in Section VI, with the empirical performance of the proposed method reported in Section VII. Section VIII provides a discussion.

II. DATA AND STUDY PLAN

Three publicly available datasets, the CHB-MIT dataset[1], the HUH dataset[2], and the TUSZ dataset[3], are used for empirical evaluation. The detailed analysis of the first two datasets are reported in Section VII, where we leave the last one in the supplementary material (SM). The CHB-MIT dataset consists of long-term (>20 hours) EEG recordings from 24 patients. The HUH dataset comprises EEG time series of 79 newborns with about two hours duration for each newborn, and were independently labeled by three experts. As the three experts did not give identical labelling of seizure events, we took the majority vote in labeling seizure periods, namely label a period as a true seizure if it was regarded so by at least two experts, which resulted in 46 newborns containing at least one seizure onset. We selected these 46 newborns in the analysis. Detailed patient information such as the age and EEG duration are available in Tables S1 and S2 of the SM. EEG signals in the datasets were recorded following the international 10-20 system with a sampling frequency of 256 Hz. The EEG records were preprocessed by a band pass filter from 0.53 to 70 Hz along with a 50 Hz notch filter. As reported in Table S1 of the SM, the CHB-MIT data are very imbalanced with the percentages of seizure onset time ranging from 0.07% to 1.57%, which was the reason that many existing methods were developed with 80% or

---

[1] https://physionet.org/physiobank/database/chbmit/
[2] https://zenodo.org/record/4940267/
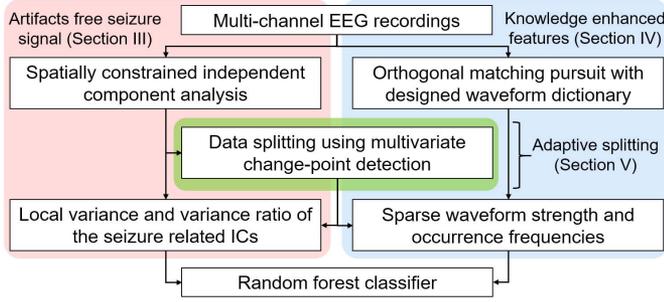[3] https://isip.piconepress.com/projects/tuh_eeg/

Fig. 1. Flow chart of the proposed seizure detection procedure which consists of two parallel flows leading to seizure classification by the random forest. The left flow (in red) uses SCICA to select seizure related ICs and extract features related to EEG oscillations. The right flow (in blue) uses OMP for sparse waveform features. The center block (in green) is a statistical change point detection algorithm for adaptively splitting the data into stationary segments.

higher training samples. We aim at algorithms trained on much smaller samples.

Our goal is to construct an efficient feature extraction procedure that generalized well to the remaining test set. To achieve this, we consider in the next two sections two parallel feature extraction procedures as demonstrated in Fig. 1. One is to identify seizure related ICs by utilizing the SCICA, which are then used to recover artifact-free signals via the inverse ICA transform. The other is a knowledge-based procedure that screens for clinically significant abnormal waveforms. Furthermore, the proposed method is doubly robust as it is a hybrid of the data driven SCICA features and the knowledge guided waveforms, as it has seizure classification power if either of the features is utilized.

Features are extracted from both of these procedures. For the acquired ICs, local variance and variance ratio features are formulated to capture changes in amplitude and variability, which are motivated by clinical observations that the EEG often become more oscillatory cross channels during seizure onsets [27]. By considering the variance features of seizure related ICs, the obtained procedure is robust against interference of artifacts or patient movements, which could otherwise lead to undesirable increases in variance. For the waveform representations, sparse wave strength and frequency are constructed, respectively.

III. SPATIALLY CONSTRAINED INDEPENDENT COMPONENT ANALYSIS

This section outlines the procedure for artifacts free local variance and variance ratio extraction based on a seizure related ICs selection via SCICA. It applies soft constraints when obtaining the seizure related ICs from the EEG series, which makes the ICs transferable from training to testing data.

The unconstrained ICA [19] admits the following linear time series model for the observed multi-channel EEG series:

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t), \qquad (1)$$

where $\boldsymbol{x}(t) = [x_1(t), \ldots, x_p(t)]^T$, $t = 1, \ldots, T$, is the zero-mean p-dimensional EEG observation at time t,

$\boldsymbol{s}(t) = [s_1(t), \ldots, s_{p'}(t)]^T$ is a zero-mean p' dimensional innovation process also called as ICs, whose elements are independent among the channels but are likely temporally dependent, and $\boldsymbol{A}$ is a $p \times p'$ mixing matrix. For the conventional scalp EEG data, $p = p' = 21$. The ICA assumes that at most one IC is Gaussian distributed for the sake of identification.

Consider spatially whitened signals

$$\boldsymbol{z}(t) = \boldsymbol{V}\boldsymbol{x}(t) = \boldsymbol{V}\boldsymbol{A}\boldsymbol{s}(t) := \boldsymbol{H}\boldsymbol{s}(t), \qquad (2)$$

where $\boldsymbol{V} = \boldsymbol{\Sigma}^{-1/2}$ and $\boldsymbol{\Sigma}$ is the covariance of $\boldsymbol{x}(t)$. It is assumed that $\boldsymbol{x}(t)$ are segmented stationary so that their covariance matrices are time-invariant over each time segment. We will provide a change-point detection method to estimate these segments in Section V. Without loss of generality, one may assume $\boldsymbol{H}$ is an orthogonal matrix for identifiability.

To solve for the ICs from $\boldsymbol{z}(t)$, the FastICA algorithm [34] may be used, which maximizes a contrast function based on a measure called negentropy. One may also solve the ICs using the nonparametric maximum likelihood estimation proposed by Samworth and Yuan in [35].

A. Spatially Constrained ICs

A primary purpose of applying ICA to EEG data is to artifact removal. However, such removal usually requires manual selection of the artifact related ICs. We present an automatic approach utilizing SCICA [20] to identify the seizure related ICs with the help of labeled data. The SCICA refines the ICA framework by introducing additional constraints to reflect aspects of clinical knowledge to guide the estimation of the mixing matrix $\boldsymbol{A}$.

The SCICA has the mixing matrix $\boldsymbol{A} = [\boldsymbol{A}_c, \boldsymbol{A}_u]$, where $\boldsymbol{A}_c$ represents spatially constrained time invariant part while $\boldsymbol{A}_u$ for unconstrained time varying part. To reflect the new setting, Model (1) is written as

$$\boldsymbol{x}(t) = \boldsymbol{A}_c\boldsymbol{s}_c(t) + \boldsymbol{A}_u(t)\boldsymbol{s}_u(t) = A\boldsymbol{s}(t), \qquad (3)$$

where $\boldsymbol{s}_c(t)$ and $\boldsymbol{s}_u(t)$ are the spatially constrained and unconstrained ICs, respectively, and the constrained $\boldsymbol{s}_c(t)$ need not be independent. Although $\boldsymbol{A}_u(t)$ is time varying, most of the seizure related information would be in the constrained part $\boldsymbol{A}_c\boldsymbol{s}_c(t)$. As $\boldsymbol{A}_u(t)$ is less important, we solve for the SCICA parameters in time moving segments to deal with the time-changing aspect of $\boldsymbol{A}_u(t)$.

Our treating $\boldsymbol{s}_c(t)$ as seizure-related ICs is different from the existing methods whose spatially constrained ICs are mainly for artifacts [36]. In these studies, the artifacts have well-defined positions, such as ocular artifacts caused by blinks are largely confined at EEG channels Fp1 and Fp2. As we interpret $\boldsymbol{s}_c(t)$ for underlying seizure activities, $\boldsymbol{A}_c$ determines the transformation from $\boldsymbol{s}_c(t)$ to $\boldsymbol{x}(t)$, while the normal neural activities and artifacts are left in $\boldsymbol{s}_u(t)$. By considering seizure related features in $\boldsymbol{s}_c(t)$, we achieve natural artifact removal. One may apply the manual artifact related ICs removal procedure to clean the EEG series before the SCICA step.

For parameter estimation, we first give a $p \times d$ constraint matrix $\boldsymbol{A}_c^0$ with d spatial constraints, and denote

$$\boldsymbol{H} = [\boldsymbol{H}_c, \boldsymbol{H}_u] := \boldsymbol{V}\boldsymbol{A} = [\boldsymbol{V}\boldsymbol{A}_c, \boldsymbol{V}\boldsymbol{A}_u]$$

as defined after (2). We treat the initial $\boldsymbol{H}_c = \boldsymbol{H}_c^0 = \boldsymbol{V}\boldsymbol{A}_c^0$. At each step, we find the unconstrained $\boldsymbol{H}_u$ given previous $\boldsymbol{H}_c$ and update $\boldsymbol{H}_c$ given $\boldsymbol{H}_u$ iteratively, by combining the FastICA method and repeated Gram-Schmidt orthogonalization as detailed in the SM. Let $\boldsymbol{h}_{cj}^0$ and $\hat{\boldsymbol{h}}_{cj}$ be the j-th column of $\boldsymbol{H}_c^0$ and $\hat{\boldsymbol{H}}_c$, respectively, where $\hat{\boldsymbol{H}}_c$ as an estimator of $\boldsymbol{H}_c$. Since $\boldsymbol{H}_c$ is assumed to be close to $\boldsymbol{H}_c^0$, the soft constraint algorithm [20] will update $\hat{\boldsymbol{h}}_{cj}$ if $|\arccos((\boldsymbol{h}_{cj}^0)^T \hat{\boldsymbol{h}}_{cj})| > \alpha$ for an angle α in radians. In that case, $\hat{\boldsymbol{h}}_{cj}$ is projected towards $\boldsymbol{h}_{cj}^0$ such that $|\arccos((\boldsymbol{h}_{cj}^0)^T \hat{\boldsymbol{h}}_{cj})| = \alpha$. More details on the SCICA algorithms are available in the SM.

Since the source of seizures can vary among individuals, we now discuss how to obtain the initial $\boldsymbol{A}_c^0$ for the SCICA. There are unsupervised attempts [37] that used the Renyi entropy and kurtosis to decide the artifact related ICs. Here, we turn to the labeled data, with the state of seizure or otherwise is known from annotations provided by neurologists. We first estimate $\hat{\boldsymbol{A}}^0$ and $\hat{\boldsymbol{s}}^0(t)$ using the unconstrained FastICA algorithm, and then select $s_c(t)$ which are the most correlated with the onset of seizure or have the largest relative energy using the labelled segments. To achieve this, given a training EEG dataset with accompanied clinic labels, let T_1 and T_2 be the labeled time index sets of the seizure and non-seizure periods, respectively. We first calculate Spearman's correlation coefficients $\hat{\rho}_{ij}$ between an IC $\hat{s}_i^0(t)$ and the EEG records $x_j(t)$ over the seizure periods $t \in T_1$. By choosing k ICs that have the k-largest correlations, we obtain candidate seizure related ICs $\hat{s}_i^0(t)$ for $i \in \mathcal{S}_1$ where the selected IC index set

$$\mathcal{S}_1 = \left\{ i : \frac{1}{p} \sum_{j=1}^{p} \hat{\rho}_{ij} > \zeta_{1k} \right\}, \text{ for } i = 1, \dots, p, \quad (4)$$

and ζ_{1k} is a threshold level such that $|\mathcal{S}_1| = k$, and k is a tuning parameter that will be defined later. In addition to the correlation, we also calculate the energy ratio of the ICs between the seizure on-site time set T_1 and off-site time set T_2:

$$\text{ER}_i = \left(|T_1|^{-1} \sum_{t \in T_1} (\hat{s}_i^0(t))^2 \right) \Big/ \left(|T_2|^{-1} \sum_{t \in T_2} (\hat{s}_i^0(t))^2 \right).$$

We then choose the first k ICs that maximize the energy ratio. To be specific, define $\mathcal{S}_2 = \{i : \text{ER}_i > \zeta_{2k}\}$, where ζ_{2k} is another threshold level. Combining the correlation and energy ratio results, the index set of the seizure related ICs is $\mathcal{S} = \mathcal{S}_1 \cap \mathcal{S}_2$. For the empirical study, we chose k so that $|\mathcal{S}| = d$, which then determines the two thresholds ζ_{1k} and ζ_{2k}. Finally, the spatial constraints $\hat{\boldsymbol{A}}_c^0 = \hat{\boldsymbol{A}}_\mathcal{S}^0$, where $\hat{\boldsymbol{A}}_\mathcal{S}^0$ is a $p \times d$ matrix with columns selected from $\mathcal{S}$.

Although the ICA and SCICA algorithms used in the above ICs selection are existing methods, the active selection of seizure related ICs via labeled data outlined above is a new formulation as well as the feature extraction and aggregation procedures in the next subsection.

B. Local Variance Features

Given the estimated $\hat{\boldsymbol{A}}_c^0$, the SCICA can be used for the whole EEG series to estimate $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{s}}(t)$ by applying the SCICA algorithm as detailed in SM. The filtered EEG series $\boldsymbol{y}(t) = [y_1(t), \dots, y_p(t)]^T = \hat{\boldsymbol{A}}_c \hat{\boldsymbol{s}}_c(t)$, which will be used as a signal enhanced version of the raw EEG observation $\boldsymbol{x}(t)$.

We extract two features, the local variance and variance ratio, from the signal enhanced average $\bar{y}(t) = p^{-1} \sum_{i=1}^{p} y_i(t)$. The local variance $\hat{\sigma}_{\bar{y}}^2(t)$ is obtained by applying the Nadaraya-Watson kernel smoothing on the estimated residuals $\hat{\varepsilon}(t) = \bar{y}(t) - \tilde{y}^*(t)$ so that

$$\hat{\sigma}_{\bar{y}}^2(t) = \frac{\sum\limits_z K_b(t - z)\hat{\varepsilon}^2(z)}{\sum\limits_z K_b(t - z)} \text{ and } \tilde{y}^*(t) = \frac{\sum\limits_x K_h(t - x)\bar{y}(x)}{\sum\limits_x K_h(t - x)}, \quad (5)$$

where h and b are two smoothing bandwidths and $K_h(u) = h^{-1}K(u/h)$ for a kernel $K(u)$ which is a probability density function itself. The bandwidths balance the bias and variance trade-off and control the smoothness of the estimated functions, which are determined by the cross-validation method.

Let $\hat{\sigma}_{\bar{x}}^2(t)$ be the local variance of $\bar{x}(t) = p^{-1} \sum_{i=1}^{p} x_i(t)$, that is similarly defined as $\hat{\sigma}_{\bar{y}}^2(t)$. The local variance ratio is the proportion of variance captured by $\bar{y}(t)$ relative to the variance of the raw average $\bar{x}(t)$:

$$\hat{r}_{\bar{y}/\bar{x}}(t) = \hat{\sigma}_{\bar{y}}^2(t)/\hat{\sigma}_{\bar{x}}^2(t). \quad (6)$$

The variance and variance ratio features for a patient from the CHB-MIT dataset are shown in Fig. 2. It is clear that the SCICA preserved most of the seizure information and removed the artifacts such as the high frequency oscillation.

IV. SPARSE WAVEFORM REPRESENTATION

In addition to the amplitude and oscillation features captured by local variance and variance ratio in Section III, it is known by neurologists that certain typical epileptic EEG waveforms are strongly correlated with seizure onset [38], and are commonly used in clinical diagnosis. The existing methods tend to employ a complete set of basis functions to capture the seizure induced waveforms. However, relying on complete bases lead to many basis functions with non-zero coefficients and hence dense representations of seizure waveforms. The latter can cause much feature variability. We take another route aiming for sparse representation by using a limited number of bases to mimic the typical seizure related waveforms, and abandon the completeness of the basis functions. With the basis functions mimicking the clinically recognized waveforms, the interpretability of the detected signals is enhanced.

A. Sparse Waveform Dictionary and OMP

To learn from the clinical knowledge on the epileptic waveforms and attain sparse representation of seizure waveform signals, we consider four basis functions $\{\phi_j(t)\}_{j=1}^4$, whose form are given in Table I. The first three basis functions indicate important epileptiform discharges, which we deliberately make not that many for sparse signal representation. Among them,
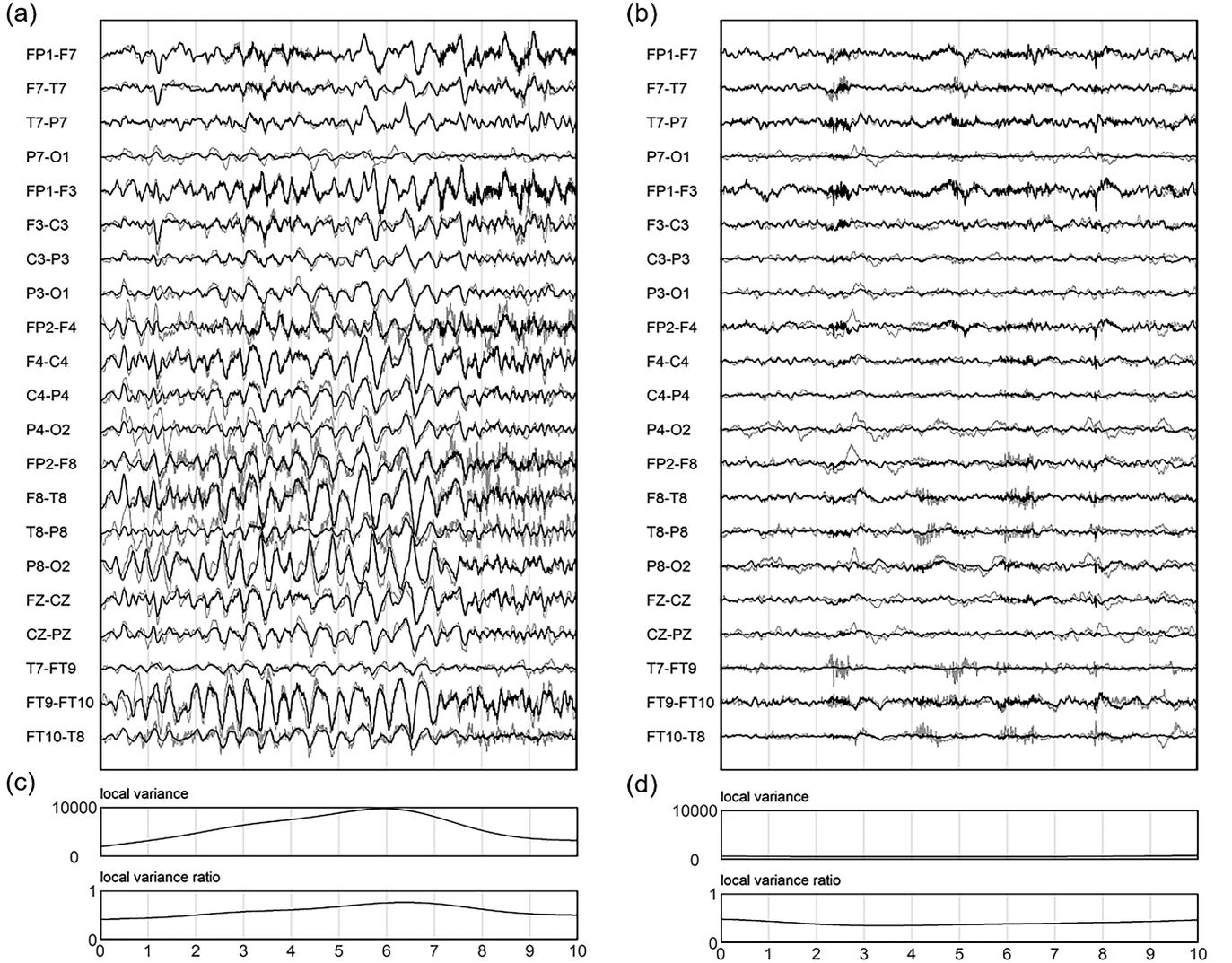
Fig. 2. An SCICA analysis on Patient 1 of the CHB-MIT dataset. The left and right panels display two 10 seconds segments from seizure and non-seizure periods, respectively. The gray lines in Panels (a) and (b) were the raw EEG series $\boldsymbol{x}(t)$ while the black lines were the seizure enhanced signal series $\boldsymbol{y}(t)$ via the SCICA. Panels (c) and (d) display the local variance $\hat{\sigma}_{\bar{y}}^2(t)$ and variance ratio $\hat{r}_{\bar{y}/\bar{x}}(t)$.

the "sinusoidal wave" ($j = 1$) are used to represent the "slow waves", and the "sharp/spike" ($j = 2$) for spike ($40 \sim 70$ms) and sharp ($70 \sim 200$ms) waves. The "sharp wave complex" ($j = 3$) is a combination of the first two members. The "Haar wave" ($j = 4$) acts to alleviate over-fitting instead of being a targeted waveform. It is noted that ϕ_1 can be expressed linearly by ϕ_2 and ϕ_3. Hence, it was not actually used. The basis functions are all compactly supported.

By utilizing these basis functions, a location-scale waveform dictionary is constructed in the form of

$$\Phi_j(t, k, \lambda) = \sqrt{\lambda} \sigma_j (\phi_j(t\lambda - k) - \mu_j), \quad j = 1, \ldots, 4, \quad (7)$$

where k is a location shift parameter and λ is the target frequency attached within to a basis $\phi_j(\cdot)$, μ_j and σ_j are location and scale parameters in the outer layer. The waveform dictionary $\{\Phi_j(t, k, \lambda)\}$ satisfies $\sum_t \Phi_j(t) = 0$ and energy $\sum_t \Phi_j(t)^2 = 1$, echoing the admissibility conditions on the wavelet bases. The candidate frequency range of λ are

provided in Table I, where the frequency of "sinusoidal" part λ_3^1 ranges from 2Hz to 8Hz, covers part of the δ band, the entire θ band and the start of α band. The uncovered δ band (less than 2 Hz) was rarely used in practice, and thus was not considered in the dictionary.
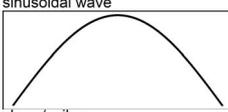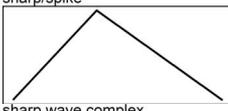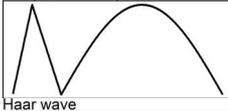
To define the dictionary in matrix form, let $\boldsymbol{\Phi}_{j,k,\lambda} = [\Phi_j(1, k, \lambda), \cdots, \Phi_j(T, k, \lambda)]^T \in \mathbb{R}^T$. Note that there are 20 candidate λ for $j = 2$, 30 λ for $j = 3$ and 7 λ for $j = 4$. Thus, if we stack $\boldsymbol{\Phi}_{j,k,\lambda}$ by columns with all possible values of (j, k, λ), we have a basis matrix $\boldsymbol{\Psi} = [\boldsymbol{\Phi}_{j,k,\lambda}]_{(j,k,\lambda)} \in \mathbb{R}^{T \times 57T}$. Given the observed EEG time series at channel i as $\boldsymbol{x}_i = [x_i(1), \ldots, x_i(T)]^T$, the goal of OMP is to find top m basis functions in $\boldsymbol{\Psi}$ which best match to $\boldsymbol{x}_i$ via regression

$$\boldsymbol{x}_i = \boldsymbol{\Psi}_i \boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, \quad (8)$$

where $\boldsymbol{\Psi}_i \in \mathbb{R}^{T \times m}$ is a submatrix consists of m columns in $\boldsymbol{\Psi}$, $\boldsymbol{\beta}_i$ is the regression coefficient vector indicating the strength of the basis functions and $\boldsymbol{\epsilon}_i$ is the residual. We do not assume (8)

TABLE I

THE DESIGNED SEIZURE WAVEFORM DICTIONARY WITH THE SINUSOIDAL AND THE SHARP/SPIKE WAVE AS BASIC UNITS. THEIR COMBINATION, THE SO CALLED SHARP WAVE COMPLEX PATTERN, IS ALSO CONSIDERED SINCE IT HAS A DIRECT ASSOCIATION TOWARDS THE SEIZURE ONSET. PARAMETERS μ_j AND σ_j ARE CHOSEN SUCH THAT $\int \Phi dt = 0$ AND $\int \Phi^2 dt = 1$, WHILE f REPRESENT THE SAMPLING FREQUENCY

location-scale family	waveform	core formula $\phi_j(t)$	frequency (Hz) λ_j
Not in direct use	sinusoidal wave	$\phi_1(t) = \sin(\pi t/(f)), \quad t \in [0, f).$	/
$\Phi_2(t, k, \lambda_2)$ $= \sqrt{\lambda_2}\sigma_2 \left(\phi_2(t\lambda_2 - k) - \mu_2\right)$	sharp/spike	$\phi_2(t) = \begin{cases} 5t/(2f), & t \in [0, 0.4f); \\ -5t/(3f) + 5/3, & t \in [0.4f, f). \end{cases}$	$\lambda_2 \in \{5, 6, \ldots, 24\}$
$\Phi_3(t, k, \lambda_3^1, \lambda_3^2)$ $= \sigma_3^{1,2} \left(\phi_3(t, k, \lambda_3^1, \lambda_3^2) - \mu_3^{1,2}\right)$	sharp wave complex	$\phi_3(t, k, \lambda_3^1, \lambda_3^2)$ $= \phi_1\left((t - f/\lambda_3^2)\lambda_3^1 - k\right) + \phi_2(t\lambda_3^2 - k)$	$\lambda_3^1 \in \{2, 3, 4, 6, 8\}$ $\lambda_3^2 \in \{5, 7, 10, 15, 20, 25\}$
$\Phi_4(t, k, \lambda_4)$ $= \sqrt{\lambda_4}\sigma_4 \left(\phi_4(t\lambda_4 - k) - \mu_4\right)$	Haar wave	$\phi_4(t) = \begin{cases} 1, & t \in [0, 0.5f); \\ -1, & t \in [0.5f, f). \end{cases}$	$\lambda_4 \in \{2, 4, 8, 16, 32, 64, 128\}$
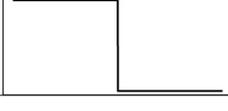
Algorithm 1: The OMP procedure for channel i

Input: $\mathbf{\Psi}$, $\mathbf{x}_i$, m.
Output: $\hat{\mathbf{\Psi}}_i$, $\hat{\boldsymbol{\beta}}_i$

1 Set $l = 1$.
2 Initialize the selected basis set $\hat{\mathbf{\Psi}}_i = 1_T$ to be the unit vector of length T.
3 Set residual $\boldsymbol{\epsilon}_i = \mathbf{x}_i$.
4 **while** $l \leq m$ **do**
5   Select a column $\mathbf{\Phi}_{j,k,\lambda}^l$ in $\mathbf{\Psi}$ that maximizes the absolute value of inner product $|\langle \mathbf{\Phi}_{j,k,\lambda}^l, \boldsymbol{\epsilon}_i \rangle|$;
6   Update the selected basis set $\hat{\mathbf{\Psi}}_i$ to $[\hat{\mathbf{\Psi}}_i, \mathbf{\Phi}_{j,k,\lambda}^l]$;
7   Exclude column $\mathbf{\Phi}_{j,k,\lambda}^l$ from $\mathbf{\Psi}$;
8   Conduct the regression $\mathbf{x}_i = \hat{\mathbf{\Psi}}_i \boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i$, and estimate $\boldsymbol{\beta}_i$ by the ordinary least square estimator $\hat{\boldsymbol{\beta}}_i = (\hat{\mathbf{\Psi}}_i^T \hat{\mathbf{\Psi}}_i)^{-1} \hat{\mathbf{\Psi}}_i^T \mathbf{x}_i$;
9   Updated $\boldsymbol{\epsilon}_i$ to $\mathbf{x}_i - \hat{\mathbf{\Psi}}_i \hat{\boldsymbol{\beta}}_i$.
10 **end**

---

is correctly specified as it is only a vehicle for selecting matched waveforms from the dictionary.

For each channel, the OMP can be implemented as stated in Algorithm 1. The OMP algorithm is essentially a linear regression combined with variable selection, where the selected variables are from the basis vectors which indicate the location and strength of the waveforms in the EEG series. If the maximum number of selected basis m is small, the OMP would have sparse solutions. One may also use the AIC or the BIC as the stopping rule of the OMP procedure.

At the end, the time complexity of OMP is $O(T^2)$ for the inner product and $O(m^2 T + m^3)$ for linear regression. However, since the basis functions have compact support ($\leq f$ in Table I), the time complexity of inner product can be reduced to $O(T)$ using their sparsity. Thus, the total time complexity of OMP is $O(T)$ for a fixed m, making it a suitable algorithm

for high frequency EEG data. In real applications, we chose $m = 8T/256$ to be eight waveforms per seconds given the sampling frequency to be 256Hz. In this case, we prefer applying OMP at some shorter segments in the consideration of computing efficiency.

B. Wave Strength and Frequency Related Features

For the purpose of lowering the feature dimension, we aggregate the OMP coefficients with respect to the wave types and brain regions by extracting the basis functions and coefficients corresponding to "sharp/spike" ($j = 2$) and "sharp wave complex" ($j = 3$).

The coefficient $\hat{\beta}_{i,j,k,\lambda}$ associated with the basis $\mathbf{\Phi}_{j,k,\lambda} \in \hat{\mathbf{\Psi}}_i$ is mapped to a function with exactly one non-zero value at location k. Formally, define two feature time series

$$\text{SHARP}_i(t) = \sum_{k=1}^{T} \sum_{\lambda_2} \sum_{t=1}^{T} |\hat{\beta}_{i,j=2,k,\lambda}| I(t = k) \quad \text{and} \quad (9)$$

$$\text{SWC}_i(t) = \sum_{k=1}^{T} \sum_{\lambda_3^1, \lambda_3^2} \sum_{t=1}^{T} |\hat{\beta}_{i,j=3,k,\lambda}| I(t = k), \quad (10)$$

corresponding to the two waveforms. The feature dimensions were 42 after the aggregation, as there are $p = 21$ EEG channels crossed with two types of waveforms. To reduce the feature dimensions, we aggregate the channel level features (9) and (10) to the eight functional regions of the brain, as shown in Fig. 3. The feature series are aggregated to the eight regions by taking averages over the channels i located in a region R_s in the scalp as

$$\text{R-SHARP}_s(t) = \frac{1}{|R_s|} \sum_{i \in R_s} \text{SHARP}_i(t) \quad \text{and} \quad (11)$$

$$\text{R-SWC}_s(t) = \frac{1}{|R_s|} \sum_{i \in R_s} \text{SWC}_i(t), \quad (12)$$

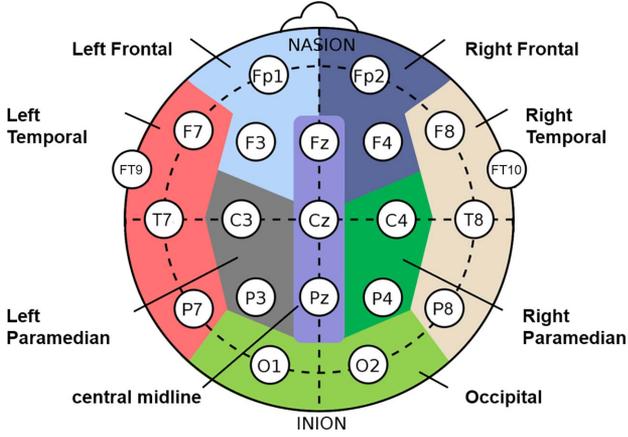where $|A|$ denotes the cardinality of set A.

Fig. 3. Eight functional regions (marked in different colors) defined by our algorithm. The channel level waveform features are further aggregated into these regions for dimension reduction.

V. CHANGE POINT DETECTION

In this section, we introduce a change-point detection algorithm to adaptively partition the time series $s_c(t)$ in (3) to segments for further temporal aggregation of the feature series that have been developed so far. The goal is to establish stationary segments with estimated change points corresponding to the start and end times of an epileptic seizure episode.

The reason for conducting the change-point detection is to counter the non-stationarity exhibited in the spatially constrained seizure related ICs $\hat{s}_c(t)$. For notation simplification, we write $\hat{s}_c(t)$ as $s(t) = \{s_1(t), \ldots, s_d(t)\}$. Since $x(t)$ has been high-pass filtered, both $x(t)$ and $s(t)$ should have zero mean, the non-stationary phenomenon should be largely shown in the second-order covariance. Thus, the focus of the change points detection is on the covariance structure.

The multi-channel seizure related ICs $s(t)$ are assumed to follow a piece-wise locally stationary wavelet (LSW) model [39]. Specifically, denote the locations of N change-points as $0 = \eta_0 < \eta_1 < \eta_2 < \cdots < \eta_N < \eta_{N+1} = T$. Let $u = t/T$ be the rescaled time, and the rescaled change points $\nu_m = \eta_m/T$. For $i = 1, \ldots, d$, the LSW model is expressed as

$$s_i(t) = \sum_{f=-\infty}^{-1} \sum_{k=-\infty}^{\infty} W_{if}(k/T)\psi_f(t-k)\xi_{if}(k), \quad (13)$$

where $\psi_f(k) = 2^{f/2}I(0 \le k \le 2^{-f-1}-1) - 2^{f/2}I(2^{-f-1} \le k \le 2^{-f}-1)$ is the mother wavelet at scale f in the frequency range and location k of the Haar wavelets, $\{\xi_{if}(k)\}$ are stochastic disturbances and $\{W_{if}(k/T)\}$ are the wavelet coefficients. Here scale $f \in \{-\infty, \ldots, -1\}$ is a synonym of frequency band, and the finest frequency band corresponds to $f = -1$. If F denotes the sampling frequency, the frequency band at scale f is roughly $[2^f F, 2^{f+1} F)$. Moreover, given a f, we assume $\boldsymbol{\xi}_f(k) = [\xi_{1f}(k), \ldots, \xi_{df}(k)]^T$ are independently distributed with mean 0 and covariance matrix $\boldsymbol{\Sigma}_f(k/T) = (\Sigma_{(i,j),f}(k/T))$ where $\Sigma_{(i,i),f}(k/T) \equiv 1$, $\text{cov}(\xi_{if}(k), \xi_{jf'}(k')) = \delta_{ff'}\delta_{kk'}\Sigma_{(i,j),f}(k/T)$, and $\delta_{ff'}$ is the Kronecker delta function.

Under the LSW model, the covariance function of $s_i(t)$ is determined by $\boldsymbol{W}_f(k/T) = [W_{1f}(k/T), \ldots, W_{df}(k/T)]^T$ and $\boldsymbol{\Sigma}_f(k/T)$. Following the piece-wise stationary assumption, $W_{if}(k/T) : [0,1] \to R$ and $\Sigma_{(i,j),f}(k/T) : [0,1] \to R$ are piece-wise constant with unknown number of change points, and we denote the change points at scale f as

$$B_{if} = \{z \in (0,1) : \lim_{u \to z-} W_{if}(u) \ne \lim_{u \to z+} W_{if}(u)\} \quad \text{and}$$

$$B_{(i,j),f} = \{z \in (0,1) : \lim_{u \to z-} \Sigma_{(i,j),f}(u) \ne \lim_{u \to z+} \Sigma_{(i,j),f}(u)\}.$$

It is clear that the autocovariance and cross-covariance functions have identical change points inherited from the piecewise constancy of $W_{if}(t)$ and $\Sigma_{(i,j),f}(t)$. Thus,

$$\{\nu_m\}_{m=1}^N = \left\{\cup_{f=-\infty}^{-1} \cup_{i=1}^d B_{if}\right\} \cup \left\{\cup_{f=-\infty}^{-1} \cup_{i,j=1}^d B_{(i,j),f}\right\}.$$

In seizure detection and neuroscience, the EEG signals at a specific frequency band may have clinical meanings. For this purpose, we define

$$\nu^{(f)} = \left\{\cup_{i=1}^d B_{if}\right\} \cup \left\{\cup_{i,j=1}^d B_{(i,j),f}\right\}$$

be the set of change points in the scale f.

The following assumptions are required for establishing consistency of the cumulative sum (CUSUM) change-point detector based on the LSW Model (13).

Assumption 1: (Gap of change points). (i) Any two adjacent change points $\nu_m, \nu_{m+1} \in \nu^{(f)}$ satisfy $|\nu_{m+1} - \nu_m| \ge \delta_T \asymp T^{\Theta-1}$ for a $\Theta \in (\frac{3}{4}, 1]$. (ii) The underlying change points should be balanced so that

$$\max\left(\frac{\nu_m - \nu_{m-1}}{\nu_{m+1} - \nu_{m-1}}, \frac{\nu_{m+1} - \nu_m}{\nu_{m+1} - \nu_{m-1}}\right) \le c_* \in [0.5, 1).$$

Assumption 2: (Signal strength). (i) The wavelet coefficients $W_{if}(u)$ in (13) is bounded such that $|W_{if}(u)| \le 2^{f/2}C$ uniformly over all $f \le -1$ and $i = 1, \ldots, d$. (ii) Let $L_{if} = \sum_{u=-\infty}^{\infty} |(W_{if}(u))^2 - (W_{if}(u-1))^2|$ and $R_{(i,j),f} = \sum_{u=-\infty}^{\infty} |\Sigma_{(i,j),f}(u) - \Sigma_{(i,j),f}(u-1)|$ be the total magnitude of jumps in $(W_{if}(u))^2$ and $\Sigma_{(i,j),f}(u)$, respectively. They satisfy $\sum_{f=-[\log T]}^{-1} 2^{-f} L_{if} = O(\log T)$ uniformly over $i \in \{1, \ldots, d\}$ and $\sum_{f=-[\log T]}^{-1} 2^{-f} R_{(i,j),f} = O(\log T)$ uniformly over $i \ne j \in \{1, \ldots, d\}$.

Assumption 3: (Distribution). (i) The IC $s_i(t)$ has zero mean and is bounded in the sense that there exists a $C > 0$ such that $|s_i(t)| \le C$ for all i and t. (ii) The disturbance $\xi_{if}(k)$ is sub-Gaussian distributed in the sense that there exists some $r > 0$ such that $Ee^{a\xi_{if}(k)} \le e^{r^2 a^2/2}$ for all real number a.

To detect change points in the second-order structure of $s(t)$, we can simply detect change points in the wavelet periodograms and the cross-periodogram sequences, since there is a one-to-one correspondence between the autocovariance ($i = j$) and cross-covariance ($i \ne j$) functions and the wavelet periodograms and cross-periodograms for the multivariate LSW model, respectively [40]. Let $w_{if}(t) = \sum_u s_i(u)\psi_f(t-u)$ be the estimated empirical wavelet coefficients, and $I_{(i,j),f}(t) = w_{if}(t)w_{jf}(t)$, where $I_{(i,i),f}(t)$ is the wavelet periodogram and $I_{(i,j),f}(t)$ for $i \ne j$ is the cross-periodogram.

By Assumption 3 (i), using the Hoeffding's lemma, it can be shown that $s_i(t)$ is sub-Gaussian distributed. Under the LSW Model (13), as the finite summation of sub-Gaussian is also sub-Gaussian, by Assumption 3 (ii), the sample periodogram $I_{(i,j),f}(t)$ can be represented as $\sigma_{(i,j),f} Z(t)^2$ where $Z(t)$ is sub-Gaussian distributed, and $E[I_{(i,j),f}(t)]$ are (asymptotically) piece-wise constant. However, when $i \neq j$, the cross-periodograms $I_{(i,j),f}(t)$ are not scaled sub-exponential distributed, and an adjustment is needed. The adjustment to $I_{(i,j),f}(t)$ is $I_{(i,j),f}(t) = [w_{if}(t) - \text{sgn}\{\widehat{\text{corr}}(w_{if}(t), w_{jf}(t))\} w_{jf}(t)]^2$, for $i \neq j$, where $\widehat{\text{corr}}(\cdot)$ denotes the sample Pearson correlation and $\text{sgn}(\cdot)$ is the sign function extracts the sign of a real number.

To define the thresholding test statistic, let $\boldsymbol{I}_{(i,j),f} = (I_{(i,j),f}(1), \ldots, I_{(i,j),f}(T))^T \in \mathbb{R}^T$ and $\boldsymbol{I}_f = \{\boldsymbol{I}_{(i,j),f}\}_{i,j=1,\ldots,d; i \leq j} \in \mathbb{R}^{T \times d'}$ be a $T \times d'$ matrix, where $d' = d(d+1)/2$ and d is the number of spatially constrained ICs (SCICs). Given a f-th scale or band, on a candidate interval $[t_1, t_2]$ for $1 \leq t_1 < t_2 \leq T$ with sample size $l = t_2 - t_1 + 1$, the CUSUM statistic [41] based on $\boldsymbol{I}_f$ at time $b \in [t_1, t_2]$ is

$$\mathcal{C}_{t_1,t_2}(b, \boldsymbol{I}_f) = \sum_{i,j=1,\ldots,d; i \leq j} \mathcal{C}^*_{t_1,t_2}(b, \boldsymbol{I}_{(i,j),f})$$
$$\times I(C^*_{t_1,t_2}(b, \boldsymbol{I}_{(i,j),f}) > \tau_{(i,j),f}) \text{ where}$$

$$\mathcal{C}^*_{t_1,t_2}(b, \boldsymbol{I}_{(i,j),f}) = \left| \sqrt{\frac{t_2 - b}{l(b - t_1 + 1)}} \sum_{t=t_1}^{b} I_{(i,j),f}(t) \right.$$
$$\left. - \sqrt{\frac{b - t_1 + 1}{l(t_2 - b)}} \sum_{t=b+1}^{t_2} I_{(i,j),f}(t) \right| / \hat{\sigma}_{t_1,t_2}(\boldsymbol{I}_{(i,j),f}),$$

where $\tau_{(i,j),f}$ is a scale f and variable (i,j) specific threshold whose choice will be discussed in Theorem 1, and $\hat{\sigma}_{t_1,t_2}(\boldsymbol{I}_{(i,j),f}) = l^{-1} \sum_{t=t_1}^{t_2} I_{(i,j),f}(t)$ is a scaling factor which standardizes the discrepancy statistic in the numerator.

We first consider the CUSUM statistic $\mathcal{C}_{1,T}(b, \boldsymbol{I}_f)$ over the entire range. If $\mathcal{C}_{1,T}(b, \boldsymbol{I}_f) > 0$ for a $b \in [1, T]$, there might be periodograms or cross-periodograms $\boldsymbol{I}_{(i,j),f}$ which contrast significantly between the adjacent time segments $[1, b]$ and $[b+1, T]$. A detected change point $\hat{\eta}$ over $[1, T]$ at scale f is

$$\hat{\eta} = \underset{1 + c_* l \leq b \leq T - c_* l}{\arg \max} \mathcal{C}_{1,T}(b, \boldsymbol{I}_f) \text{ with } \mathcal{C}_{1,T}(b, \boldsymbol{I}_f) > 0 \quad (14)$$

for any $b \in [\hat{\eta} - \Delta_T, \hat{\eta} + \Delta_T]$, where Δ_T is a tuning parameter whose order is specified in Theorem 1, and the restriction $\mathcal{C}_{1,T}(b, \boldsymbol{I}_f) > 0$ in (14) is to avoid spurious change points. If such a $\hat{\eta}$ cannot be found, we will stop the search over $[1, T]$; otherwise we add $\hat{\eta}/T$ to the detected change point set $\nu^{(f)}$ at scale f and try to locate the next change point using the binary segmentation approach. The binary segmentation splits the interval $[1, T]$ into $[1, \hat{\eta}]$ and $[\hat{\eta} + 1, T]$, and then apply procedure (14) in each of the two sub-intervals, and continues until the entire interval has been searched at scale f. The final detected change points $\{\hat{\eta}_m\}_{m=1}^{\hat{N}}$ is the union of the change points at all interested scales f. We estimated the change points at five frequency bands: 2~4Hz, 4~8Hz, 8~16Hz, 16~32Hz

and 32~64Hz in the empirical study which correspond to $f = -7, \ldots, -3$, respectively, as these frequency bands cover most seizure activities [42], [43].

Cho and Fryzlewicz [40] established the consistency of a change-point detection algorithm based on the binary segmentation with Gaussian noise, namely $\{\xi_{if}(k)\}$ in (13) are Gaussian distributed. As the SCICs of the EEG data may not be Gaussian and in fact it was insisted on being non-Gaussian [19], we establish the consistency of the proposed change points detection for sub-Gaussian innovations in the next theorem whose proof is given in the SM. Recall that δ_T is the lower bound for the gap times between any two adjacent change points in Assumption 1 (i).

Theorem 1: Under Assumptions 1 - 3 and either (i) if the number of change points N is fixed such that $\delta_T \asymp T^0$, choose $\Delta_T \asymp \varepsilon_T = \log^{2+\vartheta}(T)$ for a positive constant ϑ and the threshold levels $\tau_{(i,j),f} = \kappa_{(i,j),f} \log^{1+\omega}(T)$ for a $\omega > \vartheta/2$ or (ii) if $\delta_T \asymp T^{\Theta - 1}$ for a $\Theta \in (\frac{3}{4}, 1)$, choose $\Delta_T \asymp \varepsilon_T = T^{2 - 2\Theta}$ and $\tau_{(i,j),f} = \kappa_{(i,j),f} T^\gamma$ for a $\gamma \in (1 - \Theta, \Theta - 1/2)$ for positive tuning parameters $\kappa_{(i,j),f}$, then as $T \to \infty$, the detected change points $\{\hat{\eta}_m\}_{m=1}^{\hat{N}}$ satisfy that

$$P\{\hat{N} = N; |\hat{\eta}_m - \eta_m| < C\varepsilon_T \text{ for } m = 1, \ldots, N\} \to 1$$

for a positive constant C and with the scale $f \in [-\alpha \log \log T, -1]$ for an $\alpha \in (0, 2 + \vartheta)$.

The theorem shows that the number and location of the change points can be consistently estimated by the proposed change point detection procedure. The threshold level $\tau_{(i,j),f}$ depends on a tuning parameter $\kappa_{(i,j),f}$. The selection of $\kappa_{(i,j),f}$ can be achieved via simulated wavelet periodograms under the null hypothesis of no-change points over an interval, which is detailed in the SM and used in the empirical study. The estimated $\kappa_{(i,j),f}$ for each patient is chosen to be the upper 20% quantile of the simulated results. We chose the other tuning parameter $\Delta_T = (\log T)^2$ in empirical implementation assuming the number of change-point is finite.

VI. FEATURE AGGREGATION AND SEIZURE CLASSIFIER

We are to aggregate the features obtained from the SCICs and the OMP as discussed in Sections III and IV based on the detected change points in Section V. The aggregated features will serve as inputs to the random forest (RF) classifier [44], and the potential seizures can be detected at these segments.

It is inconvenient for the RF classifier to use these feature time series as inputs. We need to aggregate them based on a temporal splitting scheme in the time domain. There are two temporal splitting schemes. One is an adaptive temporal splitting based on the detected change-points $\{\hat{\eta}_m\}_{m=1}^{\hat{N}}$ in Section V; and the other is the fixed splitting scheme with fixed segment length over the entire sampling period, which is commonly used in existing methods. We advocate for the adaptive splitting based on the detected change points.

Recall that the feature time series from the SCICs are the local variance $\hat{\sigma}_{\bar{y}}^2(t)$ and variance ratio $\hat{r}_{\bar{y}/\bar{x}}(t)$ given in (5) and (6) in Section III-B, and the waveform features R-SHARP$_s(t)$ and R-SWC$_s(t)$ in (11) and (12) are defined in Section IV-B.

For the local variance, the adaptive splitting feature aggregation over segment $\{\eta_{k-1}+1, \eta_k\}$ is

$$\text{strength-}\hat{\sigma}_{\bar{y}}^2(k) = \frac{1}{\hat{\eta}_k - \hat{\eta}_{k-1}} \sum_{t=\hat{\eta}_{k-1}+1}^{\hat{\eta}_k} \hat{\sigma}_{\bar{y}}^2(t).$$

This can be viewed as the "strength" aggregation as it takes average of the feature series samples inside the segment. The same aggregation can be applied to the local variance ratio $\hat{r}_{\bar{y}/\bar{x}}(t)$, and the waveform feature series R-SHARP$_s(t)$ and R-SWC$_s(t)$. Besides the "strength" type aggregation, we formulate a "frequency" type aggregation, for instance that towards R-SWC$_s(t)$ is

$$\text{freq-R-SWC}_s(k) = \frac{f}{\hat{\eta}_k - \hat{\eta}_{k-1}} \sum_{t=\hat{\eta}_{k-1}+1}^{\hat{\eta}_k} I(\text{R-SWC}_s(t) \neq 0),$$

where $I(\cdot)$ is the indicator function and f is the sampling frequency. Similar formulation can be made for R-SHARP$_s(t)$. There are a total 34 aggregated features as summarized in Table S3 of SM. For comparison, we also consider feature aggregation base on fixed temporal splitting. For instance,

$$\frac{1}{2f} \sum_{t=2(k-1)f+1}^{2kf} \hat{\sigma}_{\bar{y}}^2(t) \quad \text{and}$$

$$\frac{1}{2} \sum_{t=2(k-1)f+1}^{2kf} I(\text{R-SWC}_s(t) \neq 0)$$

are the local variance and the R-SWC$_s(t)$ with fixed two-second aggregation, respectively.

It is noted that there is a sharp imbalance between the seizure and non-seizure segments in the EEG data. This means that the weighted RF has to be used to reweigh the aggregated features with respect to the segment length and the chance of the seizure. The weight to each aggregated feature is the product of the segment length and the inverse proportion to class frequencies. It was found the overall classification results were robust with respect to the minority weight ranging from 1 to 100. The weighted RF effectively measure the feature importance by the "impurity" score which is constructed via the Gini criterion.

Several issues have to be addressed before we train the weighted RF classifier, which include labeling of the segments for training, and the training and validation samples determination. Given a training sample proportion, say R%, of the total observations T, the training sample for SCICA was selected by a stratified sampling framework as detailed in the SM. The training segments used for the RF, due to the involvement of the change-point detection, are different from the training sample for the SCICA. The training segments are obtained by first conducting the change-point detection outlined Section V on the full data sample, followed by matching the segments determined by the detected change-points with the SCICA training sample. If a detected segment (for either seizure or non-seizure) has more than 50% of its time points contained in the SCICA training sample, it is kept as a training segment for the RF, with the remaining segments used as part of the validation sample. The exact size of the training segments may not be exactly R%, but should be around it.

In the RF implementations, we used the default hyperparameters setting in the Ranger R package [45], which was based on 500 classification trees with unlimited tree depth, and used the Gini index as the default tree splitting criteria with minimum node size 10. The output of the RF classifier is the seizure probability which can be further used to determine the optimal threshold.

As a binary classification problem, it is crucial to determine the probability threshold for identifying seizure segments. The higher the threshold is, the harder a segment is classified as seizure, and the RF classifier is more conservative. This leads to a trade-off between the true positive rate (TPR) and false positive rate (FPR) with respect to the threshold levels. Motivated by Liu et al. [46] in the context of allocation of Gold standard testing, we choose the threshold that strikes a balance between TPR and FPR by maximizing $w_1\text{TPR} - w_2\text{FPR}$, where w_1 and w_2 are weights specified by the user. Since the cost of FPR and benefit of TPR are not the same, we set $w_1 = 30$ and $w_2 = 1$ with the candidate thresholds chosen from $\{0.01, 0.02, \ldots, 0.99\}$ to limit the false negative rate, as it is worse to miss a seizure onset event.

VII. RESULTS

We applied the proposed patient-specific seizure detection algorithm, executed via the weighted RF classifier with the aggregated features constructed in Section VI on both the CHB-MIT and HUH datasets. For the CHB-MIT dataset, we allocated a training rate of 5% (about one hour of data), while for the HUH dataset, we used 25% for training (about half an hour), with the remaining data for validation.

A. Performance Metrics

We first outline four basic attributes used to define the performance metrics on a seizure classifier. As a binary classification problem, all four possible outcomes encompass the following categories: true-positives (TP), false-negatives (FN), true-negatives (TN) and false-positives (FP). These attributes lead to the following performance metrics:

(i) Detection Sensitivity (SENS) TP/(TP+FN);
(ii) Detection Specificity (SPEC) TN/(TN+FP);
(iii) Detection Accuracy (ACC) (TP+TN)/(TP+FN+TN+FP).

These three metrics are defined in the sample level that has a resolution of 256 Hz. We also consider (iv) the area under the receiver operating characteristics curve (AUC) as a performance measure.

Furthermore, we define two metrics related to the detection on seizure events or episodes. A correctly (wrongly) detected seizure event is one detected seizure event that contains at least one (no) seizure label. Let N_1, N_2 and N be the number of correctly detected, wrongly detected and total seizure events, respectively, which lead to the events related metrics:

(v) The percentage of failure to detect events (FDE) $(1 - N_1/N) * 100\%$
(vi) The number of false-positive alarm rate per hour (FAR), which is N_2/T_h,

where $T_h = T/3600/256$ is the number of observation hours.

TABLE II
PERFORMANCE OF SIX EXISTING SEIZURE DETECTION ALGORITHMS AND THE PROPOSED METHOD FOR THE CHB-MIT DATASET. THE PERFORMANCE MEASURES ARE SENSITIVITY (SENS), SPECIFICITY (SPEC), ACCURACY (ACC), AREA UNDER THE RECEIVER OPERATING CHARACTERISTICS (ROC) CURVE (AUC), FAILURE DETECTION EVENT (FDE), AND FALSE-POSITIVE ALARM RATE PER HOUR (FAR). THE LAST COLUMN REPORTS THE NUMBER OF PATIENTS USED IN VARIOUS STUDIES

References	SENS %	SPEC %	ACC %	AUC	FDE %	FAR/h	Training rate %	Studied patients number
Zarei and Asl [47]	96.8	97.3	97.1	-	-	-	90	23
Cimr et al. [48]	97.1	96.9	97.0	-	-	-	90	24
Bhattacharyya and Pachori [8]	97.9	99.6	99.4	1.00	-	-	90	23
Guo et al. [49]	95.6	92.6	92.6	-	0.0	1.38	25	24
Zabihi et al. [50]	91.2	95.2	95.1	0.93	-	-	25	23
Kiranyaz et al. [51]	89.0	94.7	-	-	-	-	25	21
Proposed method	95.7	99.4	99.4	0.99	0.0	1.30	5	24

TABLE III
PATIENT-SPECIFIC PERFORMANCE MEASURES FOR THE PROPOSED SEIZURE DETECTION METHOD IN THE CHB-MIT DATASET. ADAPTIVE SPLITTING AND THE FIXED SEGMENTS DESIGNS HAVE BEEN USED FOR FEATURE AGGREGATION

	Adaptive splitting						Fixed two seconds segments					
	Duration				Events		Duration				Events	
	SENS	SPEC	ACC	AUC	FDE	FAR	SENS	SPEC	ACC	AUC	FDE	FAR
chb01	99.2	100.0	100.0	1.00	0.0	0.00	94.3	100.0	100.0	1.00	0.0	0.00
chb02	99.3	100.0	100.0	1.00	0.0	0.00	100.0	100.0	100.0	1.00	0.0	0.03
chb03	98.8	99.7	99.7	1.00	0.0	0.13	100.0	99.7	99.7	1.00	0.0	0.11
chb04	98.0	99.9	99.9	0.99	0.0	0.24	91.4	99.9	99.9	0.98	0.0	0.45
chb05	93.6	100.0	100.0	1.00	0.0	0.00	96.7	96.7	99.9	1.00	0.0	0.10
chb06	85.5	100.0	100.0	0.95	0.0	1.06	59.9	99.7	99.7	0.94	30.0	14.11
chb07	92.7	100.0	100.0	1.00	0.0	0.05	93.6	100.0	100.0	1.00	0.0	0.00
chb08	97.4	99.9	99.8	0.99	0.0	0.15	95.9	99.5	99.5	0.99	0.0	1.60
chb09	98.0	100.0	100.0	1.00	0.0	0.00	97.1	100.0	100.0	1.00	0.0	0.00
chb10	95.7	99.8	99.8	0.99	0.0	0.08	97.9	99.9	99.9	1.00	0.0	0.08
chb11	97.2	99.7	99.7	1.00	0.0	0.14	100.0	100.0	100.0	1.00	0.0	0.00
chb12	93.1	95.6	95.6	0.94	0.0	9.90	85.9	89.0	88.9	0.93	0.0	16.95
chb13	93.5	98.2	98.1	0.96	0.0	3.73	94.5	96.9	96.9	0.94	0.0	5.09
chb14	93.3	99.7	99.7	0.99	0.0	0.50	90.3	99.5	99.5	0.99	0.0	1.38
chb15	95.5	96.9	96.9	0.96	0.0	2.41	95.2	97.3	97.3	0.97	0.0	3.32
chb16	88.4	98.8	98.8	0.97	0.0	10.65	80.7	99.4	99.3	0.96	0.0	6.65
chb17	99.4	100.0	100.0	1.00	0.0	0.00	98.8	99.9	99.9	0.99	0.0	0.10
chb18	94.9	99.6	99.6	1.00	0.0	1.03	97.3	98.9	98.9	1.00	0.0	1.46
chb19	99.1	99.9	99.9	1.00	0.0	0.00	100.0	100.0	100.0	1.00	0.0	0.00
chb20	92.7	100.0	99.9	1.00	0.0	0.07	95.2	99.6	99.6	1.00	0.0	0.79
chb21	99.7	100.0	100.0	1.00	0.0	0.09	100.0	100.0	100.0	1.00	0.0	0.00
chb22	95.1	100.0	100.0	1.00	0.0	0.00	97.1	100.0	100.0	1.00	0.0	0.00
chb23	99.5	99.9	99.9	1.00	0.0	0.22	100.0	99.9	99.9	1.00	0.0	0.22
chb24	98.4	99.2	99.2	1.00	0.0	0.68	97.3	99.3	99.3	0.99	0.0	1.27
average	95.7	99.4	99.4	0.99	0.0	1.30	94.1	99.0	99.1	0.99	1.3	2.24
(Standard error)	(0.75)	(0.22)	(0.22)	(0.02)	(0.00)	(0.59)	(1.77)	(0.48)	(0.47)	(0.02)	(1.25)	(0.91)

The metrics (i)-(iv) focus on the consistency of the classification over the entire duration, while (v) and (vi) on the classification quality on the event segments. For datasets with extremely imbalanced seizure and non-seizure periods, just reporting the ACC or SPEC can be misleading and overstating model performance. We recommend using the SENS and FAR as the key metrics, since they are of critical importance for clinical acceptance of an algorithm [4].

B. Performance Accuracy

To evaluate the proposed method and the adaptive splitting aggregation strategy, we compared the performance with the recent works on both the CHB-MIT and HUH datasets, as demonstrated in Tables II and IV, while those on the TUSZ dataset are reported in the SM.

For the CHB-MIT dataset, it shows that the proposed procedure achieved an average SENS of 95.7% and an average ACC of 99.4% among all the patients. The proposed method trained on 5% sample had a slightly lower detection sensitivity than those using 90% training data samples but significantly higher than those using 25% training data. This indicates that the proposed method is attractive in the few shot learning setting with satisfactory performance. Compared with human experts, whose average SENS and SPEC were 82% and 99.26%, with the FAR being 0.117 [52], the proposed procedure achieved higher SENS and SPEC. The median FAR of our procedure trained on 5% of the sample is 0.13, which was slightly higher than the human experts.

Table III presents the patient-specific performance using the proposed seizure detection procedure, which also shows that the performance of patients 6, 12 and 16 were slightly lower than the other patients. The three patients are known to be harder

TABLE IV

PERFORMANCE OF FOUR EXISTING SEIZURE DETECTION ALGORITHMS, THREE EXPERTS AND THE PROPOSED METHOD WITH TWO TRAINING RATES (25% AND 50%) FOR THE HUH DATASET. THE PERFORMANCE MEASURES ARE SENSITIVITY (SENS), SPECIFICITY (SPEC), ACCURACY (ACC), AREA UNDER THE RECEIVER OPERATING CHARACTERISTICS (ROC) CURVE (AUC), FAILURE DETECTION EVENT (FDE), AND FALSE-POSITIVE ALARM RATE PER HOUR (FAR). THE LAST COLUMN REPORTS THE NUMBER OF NEWBORNS USED IN VARIOUS STUDIES

References	SENS %	SPEC %	ACC %	AUC	FDE %	FAR/h	Training Rate %	Studied Patients Number
Tanveer et al. [53]	96.7	95.9	96.3	0.99	-	-	90	39
Borovac et al. [54]	79.5	93.7	-	0.92	14.5	1.99	97*	38
Raab et al. [55]	82.6	97.6	90.1	-	-	-	90	12
Mumenin et al. [56]	88.4	93.8	93.4	-	-	-	70	76
Expert A	86.9	98.4	96.8	-	13.4	0.20	0	46
Expert B	90.8	89.7	93.1	-	8.8	1.57	0	46
Expert C	85.8	97.3	96.5	-	11.5	0.86	0	46
Proposed method	93.0	95.5	95.4	0.95	6.6	1.60	25	46
Proposed method	94.5	96.4	96.0	0.96	6.4	1.41	50	46

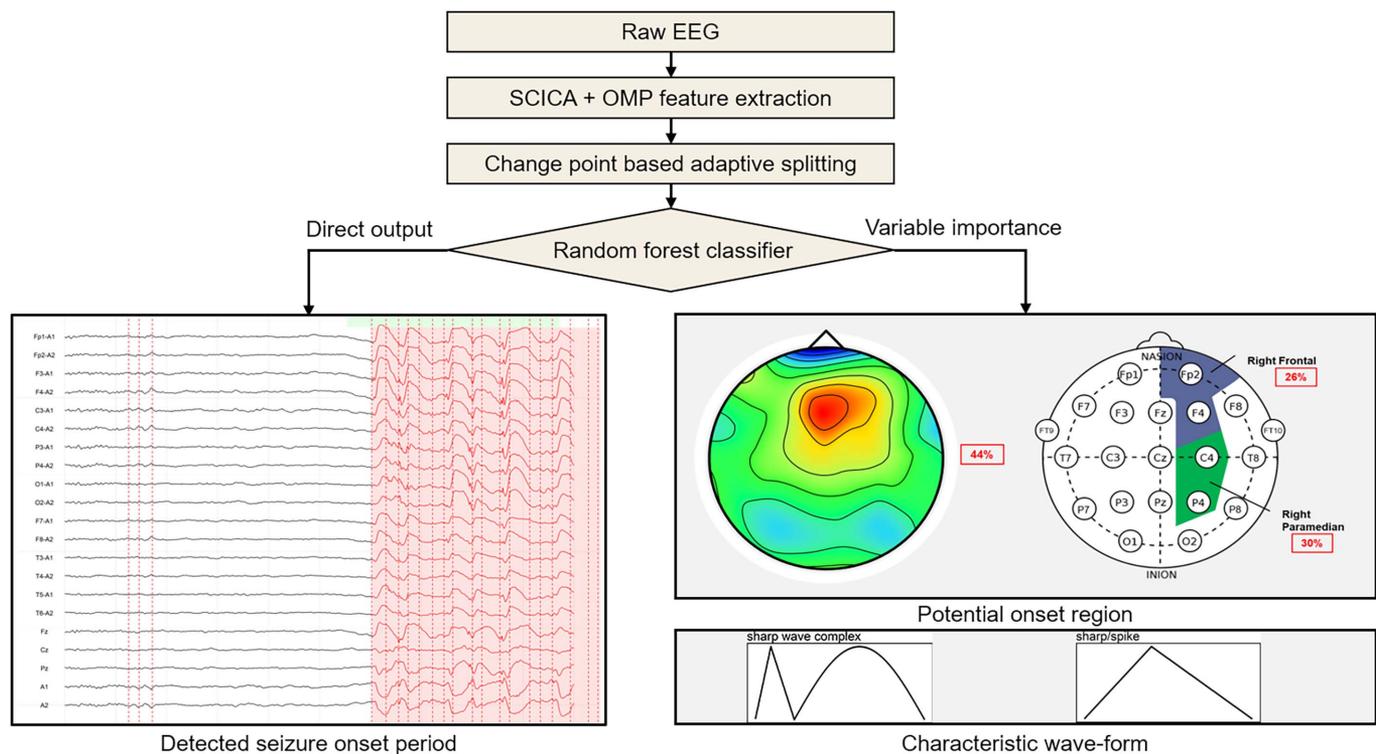*The training rate was attained due to its using the leave one subject out cross-validation.

Fig. 4. A flow chart for the output of the proposed seizure detection procedure. The direct output is the detected seizure onset period (marked as red), while the indirect output is the important features given by the random forest classifier. For example, we can obtain the potential seizure onset region and characteristic waveforms from the top five most important features.

to classify and some published studies [8], [50], [51] skipped the patients. Our analysis showed that, for the three patients, the energy or local variance during seizure periods were not significantly higher than those in interictal period, which may be the reason why they were harder to classify. However, our unique waveform features showed much difference between seizure onset and offset periods and were selected as the top features by the RF classifier.

As indicated above, only a few studies had validated on the entire data series and all patients of the CHB-MIT dataset. In most of the existing studies, only those hourly data files (the CHB-MIT data are organized as hourly files) containing seizure onset have been used for validation, which could significantly reduce the sample size due to the small proportion of onset periods. Such a bias sample selection procedure may elevate the performance metrics if the validation is not conducted over the full data. To reflect this aspect, Table II has the last column marking the number of patients used for validation.

For comparison, Table III also presents the classification results using the same features extracted by the SCICA and waveform dictionary but aggregated based on the fixed two-seconds segments that also had 5% training segments, as an alternative to the adaptive splitting based feature aggregation for the random forest classifier. For majority of the patients, the fixed segment approach had inferior performance than the classification with the adaptive splitting in the feature

aggregation, both in ACC and FAR, especially for the harder to detect Patients 6, 12 and 16. These demonstrated the advantage of the adaptive splitting based feature aggregation.

For the HUH dataset, Table IV reports the performance of the four existing methods, the proposed method and the three experts. Comparing with the four existing methods (excluding the three experts), the proposed method with 25% training rate was the second best in sensitivity (SENS), accuracy (ACC) and the area under the receiver operating characteristics curve (AUC), the third in specificity (SPEC), and the best in both the failure to detect events (FDE) and the false-positive rate per hour (FAR). Although Tanveer et al. (2021) [53] had better performance than the proposed on SENS, SPEC, ACC and AUC, it was trained with 90% of data, much more than the proposed. When the training rate of the proposed method was increased to 50%, all performance metrics improved, and in particular the SPEC exceeded that of Tanveer et al. (2021) [53] while the failure to detect and the false positive rates (FDE and FAR) were even lower.

Comparing with the three experts, the proposed was the best with respect to the sensitivity and FDE, which are important as it has a far more worse consequences if one misses a seizure onset event. The other four metrics were largely comparable with the experts and were better than Expert B in five (six) of the six metrics at 25% (50%) training. The detailed patient level classification performances can be found in Tables S5, S6 and S7 of the SM.

C. Source Regions and Characteristic Waveforms

In additional to the seizure detection with the proposed few shot detection procedure, we can also find features and brain regions influential in the detected seizure onsets. These will enhance the interpretability of the proposed procedure. Interpretability is of great importance for any algorithms being actual used by medical practitioners as they tend to avoid hard to understand methods.

The variable importance measure as a by-product of the RF classifier provides a rank of importance on the derived features. Among the 34 features, 32 features come from 8 brain regions with each region having four features on the two waveforms versus their strength and frequency, and 2 SCICA features on the local variance and the variance ratio information embedded in the five SCICs. The top ranked features can provide information about the seizure onset location and the type of EEG abnormal waveforms, as both the SCICA and OMP features can be linked to specific brain regions.

The procedure to obtain seizure onset regions and characteristic waveforms is demonstrated in Fig. 4. The derivation of the OMP region is straightforward, since it has already been divided into 8 subregions. For the SCICA region, the topographies (brain regions) associated with the i-th SCIC appear in the i-th column of the mixing matrix $\boldsymbol{A}_c^0$, which represent how the i-th SCIC was mapped to the observed EEG series. By mapping the absolute value of the i-th column vector of $\boldsymbol{A}_c^0$ to the corresponding channel location of the brain, we locate the region most associated with the i-th SCIC. To aggregate the information from the five seizure related SCICs, the SCICA
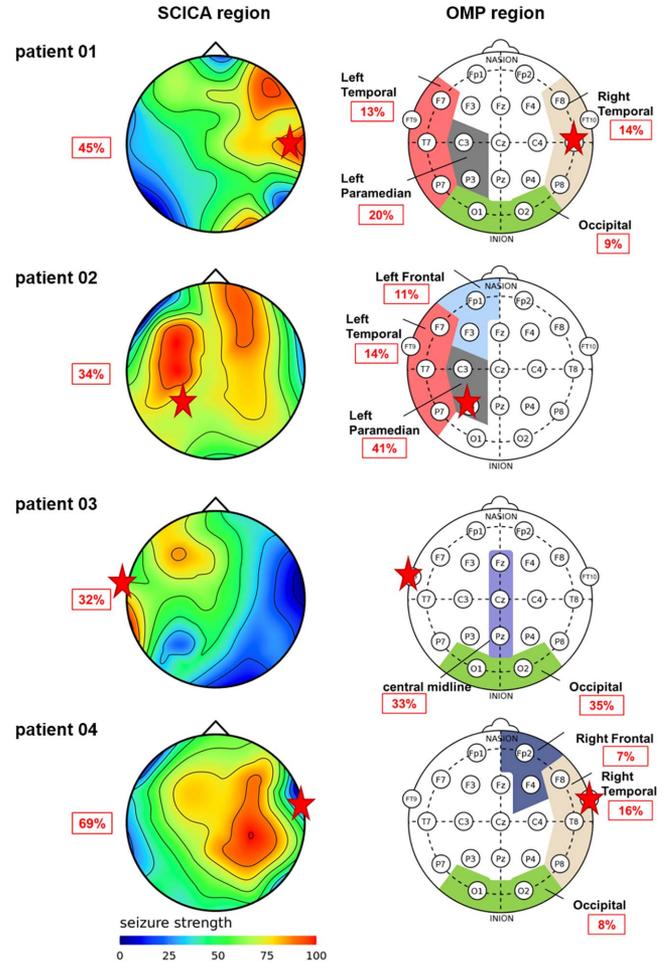

Fig. 5. Seizure source regions identified from the top five features for Patients 1 to 4 in the CHB-MIT dataset. The OMP regions indicate the regions associated with the most important waveform features in the top five features while the SCICA region shows the average of the source regions corresponding to the five SCICs, where the red color indicates the seizure related regions. The confidence towards these regions are shown in percentage, and red star showed the source region suggested by the doctor.

region is then averaged in the sense that the topographies were obtained by first forming averages of absolute $\boldsymbol{A}_c^0$ by rows and then mapping the vector to the channel locations.

The identified source regions for Patients 1 to 4 of the CHB-MIT dataset are displayed in Fig. 5, complemented with the source regions suggested by a collaborative neurologist (marked as red stars). Each region identified by the algorithm is linked with a percentage of its variable importance among the top five features. It is clear the region marked by the algorithm was close to the region marked by the neurologist, especially for patients 1 to 3. The suggested seizure active regions are helpful to neurologists for seizure diagnosis and possible invasive treatments later.

Fig. S1 displays frequencies of the features in the top five important ones via the size of the nodes, and a connection network where an edge between two nodes means the two appeared in the top five features for a patient. It shows that the SCICA features (the local variance and variance ratio) were among the most important five features, except for Patients 6, 12 and 21. Two of the three patients were among the most difficult

cases as conveyed in Table III. For Patient 6, the SCICA regions estimated by the top 5 SCICs were almost non-overlapping, which made spatially constrain the true region difficult. For patient 12, whose seizure events happened more frequently, the non-seizure periods more often fell into the pre-ictal and after-ictal states, which caused the SPEC being the lowest among the 24 patients. In addition, the seizure related waveforms showed less low frequency activities for patients 12 and 16, made their seizure waveforms less distinguishable in terms of energy. Indeed, the dominate feature for these two patients were the spikes occurred at the right paramedian region. Our analysis of the CHB-MIT data suggested that the classification performance was significantly benefited from the waveform features if the seizure period showed less low frequency oscillation.

The connection network in Fig. S1 shows that the 34 features could be divided into five groups, including the two SCICA features and the four OMP features. The local variance obtained by SCICA was the most important, appeared 18 times among the 24 patients. For the OMP features, the strength type and the sharp/spike features were more prominent than the frequency type and the SWC waveform, respectively. For the patients in CHB-MIT dataset, it was found that the left brain region showed more seizure onset. In the meanwhile, the left waveform features were more closely connected, showed a greater connectivity.

VIII. CONCLUSION

Due to the existence of noise and artifacts in the EEG data, finding ways to enhance the signal to noise ratio offers a viable approach in training better algorithms for automatic seizure detection. In this work, we propose informative feature extraction via (i) the SCICA algorithm that also removes the artifacts, (ii) the sparse waveform representation that utilizes the neurologists' knowledge to scan for seizure onset waveforms, and (iii) adaptive feature aggregation via the multivariate change point detection for feature aggregation in the time domain leading to enhanced signal to noise ratio in the extracted features.

Compared with the existing studies, an attractive aspect of our proposal is that it can achieve high detection accuracy with small training samples in the context of the few-shot learning. On the benchmark CHB-MIT dataset, we achieved high sensitivity and accuracy with only 5% of the data for training. Another attraction is that the procedure is built on interpretable features with strong neurological meanings, which would make it more acceptable to medical practitioners to assist their diagnosis by providing seizure onset alerts with the seizure specific waveforms and onset regions.

In this study, only sharp/spike and sharp wave complex waveforms are modeled in the wave-from basis functions. One can add more easily confused non-seizure waveforms, such as the vertex and the k-complex, to reduce the FAR of the algorithm. Another possible future direction is to generalize the patient specific few-shot learning to generalized detection that can transfer the features and waveform features among patients for better diagnosis. Models commonly used in transfer learning or domain generalization can be introduced to construct such label free global detection procedure.

REFERENCES

[1] I. Fried, "Auras and experiential responses arising in the temporal lobe," *J. Neuropsychiatry Clin. Neurosci.*, vol. 9, no. 3, pp. 420–428, 1997.

[2] R. S. Fisher et al., "Epileptic seizures and epilepsy: Definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE)," *Epilepsia*, vol. 46, no. 4, pp. 470–472, 2005.

[3] "Epilepsy: A public health imperative," World Health Organization, Geneva, Switzerland, 2019. [Online]. Available: https://www.who.int/publications/i/item/epilepsy-a-public-health-imperative

[4] C. Baumgartner and J. P. Koren, "Seizure detection using scalp-EEG," *Epilepsia*, vol. 59, no. S1, pp. 14–22, 2018.

[5] P. Boonyakitanont, A. Lek-Uthai, K. Chomtho, and J. Songsiri, "A review of feature extraction and performance evaluation in epileptic seizure detection using EEG," *Biomed. Signal Process. Control*, vol. 57, 2020, Art. no. 101702.

[6] V. Joshi, R. B. Pachori, and A. Vijesh, "Classification of ictal and seizure-free EEG signals using fractional linear prediction," *Biomed. Signal Process. Control*, vol. 9, pp. 1–5, Jan. 2014.

[7] D. Bhati, M. Sharma, R. B. Pachori, and V. M. Gadre, "Time–frequency localized three-band biorthogonal wavelet filter bank using semidefinite relaxation and nonlinear least squares with epileptic seizure EEG signal classification," *Digit. Signal Process.*, vol. 62, pp. 259–273, Mar. 2017.

[8] 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.

[9] Y. Li, W.-G. Cui, H. Huang, Y.-Z. Guo, K. Li, and T. Tan, "Epileptic seizure detection in EEG signals using sparse multiscale radial basis function networks and the Fisher vector approach," *Knowl.-Based Syst.*, vol. 164, pp. 96–106, 2019.

[10] Y. Wang, H. Ombao, and M. K. Chung, "Topological data analysis of single-trial electroencephalographic signals," *Ann. Appl. Statist.*, vol. 12, no. 3, pp. 1506–1534, 2018.

[11] A. L. Schröder and H. Ombao, "FreSpeD: Frequency-specific change-point detection in epileptic seizure multi-channel EEG data," *J. Am. Statist. Assoc.*, vol. 114, no. 525, pp. 115–128, 2019.

[12] R. R. Sharma and R. B. Pachori, "Time–frequency representation using IEVDHM–HT with application to classification of epileptic EEG signals," *IET Sci., Meas. Technol.*, vol. 12, no. 1, pp. 72–82, 2018.

[13] V. Gupta and R. B. Pachori, "Epileptic seizure identification using entropy of FBSE based EEG rhythms," *Biomed. Signal Process. Control*, vol. 53, 2019, Art. no. 101569.

[14] R. Cherian and E. G. Kanaga, "Theoretical and methodological analysis of EEG based seizure detection and prediction: An exhaustive review," *J. Neurosci. Methods*, vol. 369, 2022, Art. no. 109483.

[15] E. Reus, G. Visser, J. van Dijk, and F. Cox, "Automated seizure detection in an EMU setting: Are software packages ready for implementation?" *Seizure*, vol. 96, pp. 13–17, Mar. 2022.

[16] T. M. Ganguly et al., "Seizure detection in continuous inpatient EEG," *Neurology*, vol. 98, no. 22, pp. e2224–e2232, 2022.

[17] A. A. Ein Shoka, M. M. Dessouky, A. El-Sayed, and E. E.-D. Hemdan, "EEG seizure detection: Concepts, techniques, challenges, and future trends," *Multimedia Tools Appl.*, vol. 82, pp. 1–31, Nov. 2023.

[18] A. Shoeibi et al., "An overview of deep learning techniques for epileptic seizures detection and prediction based on neuroimaging modalities: Methods, challenges, and future works," *Comput. Biol. Med.*, vol. 149, 2022, Art. no. 106053.

[19] P. Comon, "Independent component analysis, A new concept?" *Signal Process.*, vol. 36, no. 3, pp. 287–314, 1994.

[20] C. Hesse and C. James, "The FastICA algorithm with spatial constraints," *IEEE Signal Process. Lett.*, vol. 12, no. 11, pp. 792–795, Nov. 2005.

[21] O. Dimigen, "Optimizing the ICA-based removal of ocular EEG artifacts from free viewing experiments," *NeuroImage*, vol. 207, 2020, Art. no. 116117.

[22] S. Raghu, N. Sriraam, G. P. Kumar, and A. S. Hegde, "A novel approach for real-time recognition of epileptic seizures using minimum variance modified fuzzy entropy," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 11, pp. 2612–2621, Nov. 2018.

[23] A. Subasi and M. I. Gursoy, "EEG signal classification using PCA, ICA, IDA and support vector machines," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8659–8666, 2010.

[24] A. Delorme, T. Sejnowski, and S. Makeig, "Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis," *Neuroimage*, vol. 34, no. 4, pp. 1443–1449, 2007.

[25] Y. Li, X.-D. Wang, M.-L. Luo, K. Li, X.-F. Yang, and Q. Guo, "Epileptic seizure classification of EEGs using time–frequency analysis based multiscale radial basis functions," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 2, pp. 386–397, Mar. 2017.

[26] L. S. Vidyaratne and K. M. Iftekharuddin, "Real-time epileptic seizure detection using EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 2146–2156, Nov. 2017.

[27] A. Varsavsky, I. Mareels, and M. Cook, *Epileptic Seizures and the EEG: Measurement, Models, Detection and Prediction*. Boca Raton, FL, USA: CRC Press, 2016.

[28] M. Fink, "Object classification from a single example utilizing class relevance metrics," in *Proc. Adv. Neural Inf. Process. Syst.*, 2005, vol. 17, pp. 449–456.

[29] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a few examples: A survey on few-shot learning," *ACM Comput. Surv. (CSUR)*, vol. 53, no. 3, pp. 1–34, 2020.

[30] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[31] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, MIT, Cambridge, MA, USA, 2009.

[32] N. J. Stevenson, K. Tapani, L. Lauronen, and S. Vanhatalo, "A dataset of neonatal EEG recordings with seizure annotations," *Scientific Data*, vol. 6, no. 1, 2019, Art. no. 190039.

[33] V. Shah et al., "The Temple University Hospital seizure detection corpus," *Frontiers Neuroinformat.*, vol. 12:83, 2018, doi: 10.3389/fninf.2018.00083

[34] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, May 1999.

[35] R. J. Samworth and M. Yuan, "Independent component analysis via nonparametric maximum likelihood estimation," *Ann. Statist.*, vol. 40, no. 6, pp. 2973–3002, 2012.

[36] C. Hesse and C. James, "On semi-blind source separation using spatial constraints with applications in EEG analysis," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2525–2534, Dec. 2006.

[37] N. Mammone, F. La Foresta, and F. C. Morabito, "Automatic artifact rejection from multichannel scalp EEG by wavelet ICA," *IEEE Sensors J.*, vol. 12, no. 3, pp. 533–542, Mar. 2012.

[38] S. J. M. Smith, "EEG in the diagnosis, classification, and management of patients with epilepsy," *J. Neurol. Neurosurg. Psychiatry*, vol. 76, no. suppl 2, pp. ii2–ii7, 2005.

[39] G. P. Nason, R. Von Sachs, and G. Kroisandt, "Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum," *J. Roy. Statist. Soc., Ser. B (Statist. Methodol.)*, vol. 62, no. 2, pp. 271–292, 2000.

[40] H. Cho and P. Fryzlewicz, "Multiple-change-point detection for high dimensional time series via sparsified binary segmentation," *J. Roy. Statist. Soc. Ser. B-Statist. Methodol.*, vol. 77, no. 2, pp. 475–507, 2015.

[41] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.

[42] M. Mursalin, Y. Zhang, Y. Chen, and N. V. Chawla, "Automated epileptic seizure detection using improved correlation-based feature selection with Random Forest classifier," *Neurocomputing*, vol. 241, pp. 204–214, Jun. 2017.

[43] M. G. Tsipouras, "Spectral information of EEG signals with respect to epilepsy classification," *EURASIP J. Adv. Signal Process.*, vol. 2019, no. 1, 2019, Art. no. 10.

[44] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[45] M. N. Wright and A. Ziegler, "Ranger: A fast implementation of random forests for high dimensional data in C++ and R," *J. Statist. Softw.*, vol. 77, no. 1, pp. 1–17, 2017.

[46] T. Liu, J. W. Hogan, L. Wang, S. Zhang, and R. Kantor, "Optimal allocation of gold standard testing under constrained availability: Application to assessment of HIV treatment failure," *J. Am. Statist. Assoc.*, vol. 108, no. 504, pp. 1173–1188, 2013.

[47] 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.

[48] D. Cimr, H. Fujita, H. Tomaskova, R. Cimler, and A. Selamat, "Automatic seizure detection by convolutional neural networks with computational complexity analysis," *Comput. Methods Programs Biomed.*, vol. 229, Feb. 2023, Art. no. 107277.

[49] Y. Guo et al., "Epileptic seizure detection by cascading isolation forest-based anomaly screening and EasyEnsemble," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 915–924, Mar. 2022.

[50] M. Zabihi, S. Kiranyaz, V. Jantti, T. Lipping, and M. Gabbouj, "Patient-specific seizure detection using nonlinear dynamics and nullclines," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 2, pp. 543–555, Feb. 2020.

[51] S. Kiranyaz, T. Ince, M. Zabihi, and D. Ince, "Automated patient-specific classification of long-term electroencephalography," *J. Biomed. Inform.*, vol. 49, pp. 16–31, Jun. 2014.

[52] S. B. Wilson, M. L. Scheuer, C. Plummer, B. Young, and S. Pacia, "Seizure detection: Correlation of human experts," *Clin. Neurophysiol.*, vol. 114, no. 11, pp. 2156–2164, 2003.

[53] M. A. Tanveer, M. J. Khan, H. Sajid, and N. Naseer, "Convolutional neural networks ensemble model for neonatal seizure detection," *J. Neurosci. Methods*, vol. 358, Jul. 2021, Art. no. 109197.

[54] A. Borovac et al., "Ensemble learning using individual neonatal data for seizure detection," *IEEE J. Transl. Eng. Health Med.*, vol. 10, pp. 1–11, Aug. 2022.

[55] D. Raab, A. Theissler, and M. Spiliopoulou, "XAI4EEG: Spectral and spatio-temporal explanation of deep learning-based seizure detection in EEG time series," *Neural Comput. Appl.*, vol. 35, no. 14, pp. 10051–10068, 2023.

[56] K. M. Mumenin, P. Biswas, M. A.-M. Khan, A. S. Alammary, and A.-A. Nahid, "A modified aquila-based optimized XGBoost framework for detecting probable seizure status in neonates," *Sensors*, vol. 23, no. 16, 2023, Art. no. 7037.

Pei Feng Tong received the B.S. degree from Peking University, Beijing, China, in 2020. He is currently working toward the Ph.D. degree in statistics with Peking University. His research interests include interpretable models, change-point analysis, and signal processing theory and algorithms.

Hao Xiang Zhan received the B.S. degree from Peking University, Beijing, China, in 2023. He is currently working toward the Ph.D. degree in statistics. His research interests focus on high dimensional data analysis.

Song Xi Chen is a Chair Professor with Peking University. He was elected as a fellow of the Institute of Mathematical Statistics (IMS) and American Statistical Association in 2009, a fellow of AAAS in 2018, and a fellow of Chinese Academy of Science in 2021. He served as a council member of IMS from 2016 to 2019, and the Scientific Secretary of Bernoulli Society from 2019 to 2023. He has published more than 120 refereed papers. His research interests include ultrahigh dimensional statistical inference, nonparametric/empirical likelihood theory, econometrics, large scale air quality monitoring, and assessment.