Automatic Epileptic Seizure Detection Using Graph-Regularized Non-Negative Matrix Factorization and Kernel-Based Robust Probabilistic Collaborative Representation

Shasha Yuan, Jianwei Mu, Weidong Zhou, Ling-Yun Dai, Jin-Xing Liu, *Member, IEEE*, Juan Wang, and Xiang Liu

***Abstract*— Automatic seizure detection system can serve as a meaningful clinical tool for the treatment and analysis of epilepsy using electroencephalogram (EEG) and has obtained rapid development. An automatic detection of epileptic seizure method based on kernel-based robust probabilistic collaborative representation (ProCRC) combined with graph-regularized non-negative matrix factorization (GNMF) is proposed in this work. The raw EEG signals are pre-processed through the wavelet transform to obtain time-frequency distribution of EEG signals as preliminary feature information and GNMF is further employed for dimension reduction, retaining and enhancing the productive feature information of EEG signals. Then, the test sample is represented using robust ProCRC that can decide whether the testing sample belongs to each class (seizure or non-seizure) by jointly maximizing the likelihood. In addition, the kernel trick is applied to improve the separability of non-linear high dimensional EEG signals in robust ProCRC. Finally, post-processing techniques are introduced to generate more accurate and reliable results. The average epoch-based sensitivity of 96.48%, event-based sensitivity of 93.65% and specificity of 98.55% are acquired in this method, which is evaluated on the public Freiburg EEG database.**

***Index Terms*— Electroencephalogram, seizure detection, graph-regularized non-negative matrix factorization, kernel method, robust probabilistic collaborative representation.**

I. Introduction

EPILEPSY is a chronic neurological disorder and manifests by repeated spontaneous seizures that result of excessive neuronal activity and 1% of the population worldwide live with epilepsy [1], [2]. Electroencephalogram plays an important role in seizure detection, which reflects the brain electrical activities [3]. However, the technique could be time consuming for trained electroencephalographers to perform visual inspection on the long term EEG recordings over an extended period of time. Sometimes this process may lead to inaccuracies [4]. Besides, inconsistent identification results are possible for the same EEG recording because of the subjective nature of the analysis. Therefore, automatic seizure detection systems are needed for epilepsy monitoring and treatment as an effective and reliable tool to aide medical staff [5].

In the early 1980s, the method that decomposed EEG signals into half waves and extracted sharpness, slope and rhythmicity for classification was one of the earliest seizure detection algorithms to be introduced in [6]. Subsequently, automatic seizure detection systems have generated considerable interest and various methods have been introduced for epileptic detection, with varying degrees of success. Srinivasan *et al.* [7] introduced an epileptic detection method that applied the neural network and combined it with approximate entropy. Wu *et al.* proposed an automatic seizure detection method based on complementary ensemble empirical mode decomposition (CEEMD) and extreme gradient boosting (XGBoost) [8]. Khan *et al.* combined local binary patterns (LBP) and discrete wavelet transform (DWT) to perform time-frequency decomposition of signals, using univariate and bivariate as features for epilepsy detection [9].

Recently, deep learning has also been widely applied to seizure detection or prediction, which achieved good performance. Karácsony *et al.* proposed an end-to-end deep learning approach based on a combination of Mask R-CNN, Inflated 3D Convnet feature extraction and LSTM-FC for the classification of Frontal and Temporal Lobe epileptic seizures [10]. Daoud and Bayoumi applied the DCNN and Bi-LSTM network learning spatial and temporal EEG features, as well as DCAE based Semi-supervised learning approach for patient-specific seizure prediction [11]. A deep learning-based model named pyramidal one-dimensional convolutional neural network was designed for the detection of epilepsy in [12]. Meanwhile, Zhao *et al.* [13] introduced a novel graph attention network

Manuscript received 21 January 2022; revised 24 July 2022; accepted 28 August 2022. Date of publication 5 September 2022; date of current version 22 September 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61701279, Grant 61902215, Grant 62172254, and Grant 62172253. *(Corresponding author: Shasha Yuan.)*

Shasha Yuan, Jianwei Mu, Ling-Yun Dai, Jin-Xing Liu, Juan Wang, and Xiang Liu are with the School of Computer Science, Qufu Normal University, Rizhao 276826, China (e-mail: jiayouyss@126.com; jyxiaomu@126.com; dailingyun_1@163.com; sdcavell@126.com; wangjuansdu@163.com; liuxiang12456@163.com).

Weidong Zhou is with the School of Microelectronics, Shandong University, Jinan 250100, China (e-mail: wdzhou@sdu.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2022.3204533

for automatic epileptic seizure detection. Duan *et al.* presented one-dimensional convolutional embedding modules and a deep metric learning model with a stage-wise training strategy to detect seizures [14].

In many image or signal processing systems, discrete wavelet transform (DWT) is usually used to decompose the original record, and then the decomposed coefficients are reconstructed to the sub-signals of different frequency bands. Indeed, due to the non-stationary characteristics of EEG signals, the characteristics of time series EEG data are difficult to directly analyze, therefore, DWT is introduced to process the raw EEG recordings to provide the frequency and time domain information for the classification algorithm, which has been widely applied because of its suitable performance [3], [15], [16].

Non-negative matrix factorization (NMF) has been applied successfully to various fields, including image processing [17], [18] and data clustering [19], [20] as a powerful matrix factorization technique. However, NMF does not preserve the geometrical information, which is essential for data processing. Compared with NMF, the graph-regularized non-negative matrix factorization (GNMF) proposed in [21] retains the intrinsic geometric structure, and therefore, more information retained by graph regularization, making it an ideal method for matrix factorization and data reduction. As redundancy exists in long-term EEG signals, GNMF can be used to perform data dimension reduction and retain the main feature information at the same time.

Collaborative representation based classification (CRC) and sparse representation based classification (SRC) have achieved good performance in many fields, including pattern classification [22], [23] and facial recognition [24], [25], [26]. CRC and SRC both represent testing samples as an efficient linear combination through training samples. Wright *et al.* [27], performed face classification by evaluating the residuals that were obtained through SRC. However, the complicated computation procedure of the 1-norm to determine sparsity associated with SRC is time-consuming. Later, Zhang *et al.* [28] presented CRC as a development of SRC and achieved good performance in accuracy, with low complexity. To clarify the intrinsic reason for CRC's satisfactory performance, Cai *et al.* [29] gave a probabilistic collaborative representation method from a probability viewpoint for pattern classification. In this paper, we utilize the probabilistic collaborative representation-based classification (ProCRC) to represent the difference in EEG signals as an efficient classification method for seizure detection. Meanwhile, for inseparable EEG recordings in the original space, the kernel trick provides a linear separable representation method that maps the samples into the high-dimensional feature space, and which has performed well in machine learning and data classification [23], [24], [25]. The non-linear separability of long-term EEG signals makes it difficult to obtain effective classification results through linear separable methods. Therefore, the kernel trick is combined with ProCRC in this seizure detection system to help obtaining more accurate classification results.

The main contributions of this work can be summarized as follows:

- The DWT and GNMF are applied for retaining and enhancing the productive feature information, which can discover the intrinsic discriminating structure of the EEG data.
- The kernel method is combined with robust probabilistic collaborative representation for classification, with the advantage that it could improve the linear separability and efficiently calculate the maximum probability that a test EEG sample belongs to each class.
- The pre-processing with the normalization based on differential operator, as well as the post-processing including multi-label fusion rule, optimize the performance of the algorithm.
- The competitive experiment results are achieved on the long-term EEG database with low computational cost.

The structure of this paper is as follows: section 2 briefly introduces the database background and the details of the datasets. The detailed algorithm for seizure detection is then exhibited in section 3. Next, the results are showed in section 4 and section 5 is devoted to a detailed discussion of the algorithm's performance. Finally, section 6 presents the conclusions for this work.

II. EEG Database

A. The Freiburg Database

The database from the Epilepsy Center of the University Hospital of Freiburg was used to evaluate the automatic seizure detection system proposed in this paper and this database provided the EEG signals for 21 patients. A Neurofile NT digital video EEG system was used to sample the raw EEG signals with a 256 Hz sampling rate. A 16-bit A/D converter was used to record the epilepsy activities. Besides, six channels were selected previously by well-trained epileptologists, including three extra-focal and three focal channels. The EEG data recorded in the three focal channels were used in this paper.

Based on the clinical manifestation, the onset and offset of epilepsy activities were determined by well-trained epileptologists. For each patient, the recording time containing seizures ranged from 2 to 5 h because each patient had different types of epilepsy. For each recorded seizure activity, the durations were also different; for example, some were less than 12 s while others were more than 15 min. A summary description is shown in Table I.

B. Training and Testing Datasets

In this automatic seizure detection system, the database was divided into two parts: the training set and testing set. To address the imbalance quantity in the seizure and non-seizure signals, the number of two types of signals in training set was selected according to the following rule: for most patients, one or two seizure recordings were chosen and double non-seizure recordings were selected to form the training set. It should be noted that for patient 18, three seizures were used because the durations of seizure activities were too short. Then data augmentation is performed on the seizure data in the way that these data are resampled twice applied the synthetic minority oversampling technique (SMOTE), which makes the amounts of seizure and non-seizure data equal.

In total, for the training set, there are 0.33 h of seizure EEG recordings and 0.66 h of non-seizure recordings, which

TABLE I
DETAILED INFORMATION OF THE FREIBURG EEG DATABASE

Patient	Seizure type	Sex	Average seizure durations (s)	Number of seizures	Testing EEG length (h)
1	CP,SP	F	13.1	4	26
2	CP,SP,GTC	M	118.2	3	26
3	CP, SP	M	72.7	5	27
4	CP, SP, GTC	F	87.4	5	28
5	CP, SP, GTC	F	44.9	5	27
6	CP,GTC	F	66.9	3	26
7	CP,SP,GTC	F	153.5	3	25
8	CP,SP	F	163.7	2	25
9	CP,GTC	M	114.7	5	27
10	CP,SP,GTC	M	411.0	4	25
11	CP,SP,GTC	F	157.3	4	26
12	CP,SP,GTC	F	55.1	4	28
13	CP,SP,GTC	F	158.3	2	25
14	CP,GTC	F	216.4	4	25
15	CP,SP,GTC	M	145.4	4	33
16	CP,SP,GTC	F	121.0	4	27
17	CP,SP,GTC	M	86.2	5	27
18	CP,SP	F	13.7	5	26
19	CP,SP,GTC	F	12.5	4	24
20	CP,SP,GTC	M	85.7	5	27
21	CP,SP	M	83.1	5	27
Total	-	-	114.3	85	557

M: Male, F: Female, CP refers to complex partial seizures type, SP refers to simple partial seizures type, GTC refers to generalized tonic-clonic seizures type.

were divided into 885 segments with 295 seizure segments and 590 non-seizure segments. In addition, the testing set encompassed 557 h EEG recordings with 56 seizures from 21 patients are employed for the evaluation of the proposed method, Specifically, it contains 1.40 h of seizure EEG recordings with 1256 segments and 555.60 h of non-seizure recordings with 50044 segments, which is long enough to evaluate the algorithm well.

III. METHODOLOGIES

The architecture of the whole algorithm for seizure detection is depicted in Fig. 1, where (a) includes pre-processing, GNMF, kernel-based robust ProCRC and post-processing; (b) shows the procedure of kernel-based robust ProCRC.

A. Pre-Processing

The long-term raw EEG recordings were divided into 4-s segments using the sliding window in a non-overlapping way. Then, the DWT was applied to the segments with five decomposition levels. After decomposition, five detailed coefficients for each segment were obtained, including d1 (64-128 Hz), d2 (32-64 Hz), d3 (16-32 Hz), d4 (8-16 Hz) and d5 (4-8 Hz), as well as the approximation coefficient A5 (0-4 Hz). Here, we used the Daubechies-4 wavelet as the wavelet function, because it has shown outstanding performance in capturing EEG signals characteristics [30], [31]. Then, EEG sub-signals were reconstructed using d3, d4 and d5 for the detection system because the frequency for most seizure activities was often below 30 Hz [32] and these sub-signals produce less artifacts and background noise than the other sub-signals. At the same time, if other detailed coefficients were used, more data redundancy would occur and more significant effects would still not be guaranteed.

Considering that the changes are not significant for some seizure events comparing to non-seizure recordings, the normalization based on differential operator is applied to enhance the contrast between the background (non-seizure data) and seizure data. After normalization, EEG signals can also satisfy the non-negative constraint for GNMF. Here, the operator is defined as:

$$P = \exp(\frac{1}{m}|diff(p_i)|) \tag{1}$$

Here, $diff$ refers to the first-order derivative, $diff(p_i) = p_i(t) - p_i(t-1)$, refers to the EEG signals after reconstruction and m is the positive parameter, chosen as 10000 in this

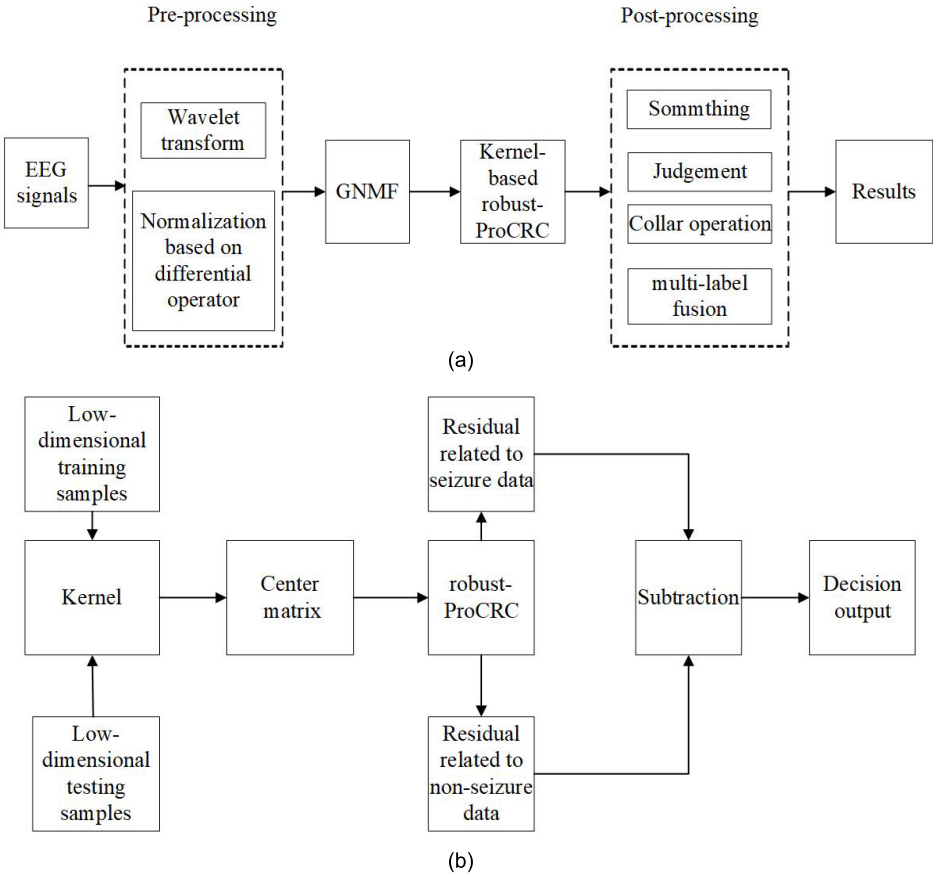

Fig. 1. (a) Represents the whole process of the detection system, which can be divided into four parts: pre-processing; GNMF for data dimension reduction; kernel-based robust ProCRC for determining the classification labels; and post-processing. (b) Outline of the kernel-based robust ProCRC method.

paper. After normalization, the seizure activity becomes more clearly compared with the background, which suggests that the differential operator-based normalization could indeed enhance the true seizure activity.

B. GNMF

NMF has been applied successfully to various fields including computer vision and data clustering as a powerful matrix factorization technique. However, it fails to preserve the geometrical information, which is essential in data processing. Hence, GNMF was proposed to overcome this limitation.

For the data matrix $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N] \in \mathbf{R}^{M \times N}$, NMF aims at decomposing it into the product of two non-negative factors as:

$$\mathbf{X} \approx \mathbf{U} \times \mathbf{V}^T \tag{2}$$

$\mathbf{V} = [v_{jk}] \in \mathbf{R}^{N \times K}$ and $\mathbf{U} = [u_{ik}] \in \mathbf{R}^{M \times K}$ refer to the coefficient matrix and the base matrix, respectively.

The Euclidean distance between $\mathbf{V}$ and $\mathbf{U}$ is used as the objective function:

$$O_{NMF} = \left\| \mathbf{X} - \mathbf{U}\mathbf{V}^T \right\|^2 \tag{3}$$

On the basis of the manifold learning theory, the GNMF algorithm was proposed in [21] in which the local geometric structure information was retained after matrix factorization. The objective function is introduced for GNMF as follows:

$$O_{GNMF} = \left\| \mathbf{X} - \mathbf{U}\mathbf{V}^T \right\|^2 + \lambda Tr\left(\mathbf{V}^T\mathbf{L}\mathbf{V}\right) \tag{4}$$

The local optimal updating rules used to obtain the two matrices are as follows:

$$u_{ik} \leftarrow u_{ik} \frac{(\mathbf{X}\mathbf{V})_{ik}}{\left(\mathbf{U}\mathbf{V}^{\mathbf{T}}\mathbf{V}\right)_{ik}} \tag{5}$$

$$v_{jk} \leftarrow v_{jk} \frac{\left(\mathbf{X}^{\mathbf{T}}\mathbf{U} + \lambda\mathbf{W}\mathbf{V}\right)_{jk}}{\left(\mathbf{V}\mathbf{U}^{\mathbf{T}}\mathbf{U} + \lambda\mathbf{D}\mathbf{V}\right)_{jk}} \tag{6}$$

After GNMF, the base matrix $\mathbf{U}$ is obtained and we can convert the raw EEG data into low-dimensional space from the original space. The representation is defined as follows:

$$y_i = \mathbf{U}^{\dagger} x_i \tag{7}$$

where $\mathbf{U}^{\dagger}$ is the pseudo-inverse matrix of base matrix $\mathbf{U}$. $\mathbf{U}^{\dagger} = (\mathbf{U}^{\mathbf{T}}\mathbf{U})^{-1}\mathbf{U}^{\mathbf{T}}$.

In this paper, GNMF retains the main feature information after dimensionality reduction by a new representation based on $\mathbf{U}^{\dagger}$. Considering EEG data will have non-negative values after DWT decomposition, so the differential operator-based normalization provides non-negative values that satisfy the non-negative constraint for GNMF. We used the training data after pre-processing to perform GNMF and obtain the base matrix $\mathbf{U}$ in this paper. Furthermore, the pseudo-inverse

matrix $\mathbf{U}^{\dagger}$ was computed and performed to transform the data as a low-dimensional representation and the main feature information of x_i was retained in the new representation for y_i.

C. The Robust ProCRC

Compared with CRC, robust ProCRC investigates the CRC algorithm from a probabilistic view and its performance, accuracy and time cost are better than either SRC or CRC. Suppose that data sets $\mathbf{Y} = [\mathbf{Y_1}, \mathbf{Y_2}, \ldots, \mathbf{Y}_k]$ is a collection of training samples. The training samples belongs to K classes and each column in $\mathbf{Y}_k$ represents a sample vector. And $\mathbf{Y}_k$ refers to the data of k th class. According to [29], for a data points y inside S, a probability is defined through a Gaussian function:

$$P(l(y) \in l_{\mathbf{Y}}) \propto \exp(-c \|\alpha\|^2) \tag{8}$$

And y has the higher probability approaching to the S center if $\|\alpha\|_2$ is smaller. $l(y)$ refers to the label of data point y, $l_{\mathbf{Y}}$ refers to the label set for the collaborative subspace S spanned by training samples and c refers to a constant.

Practically, the testing sample z may be outside the collaborative subspace and, if this is the case, the probability that the measure z belongs to S is defined as:

$$\begin{aligned} P(l(z) \in l_{\mathbf{Y}}) &= P(l(z) = l(y) | l(y) \in l_{\mathbf{Y}}) P(l(y) \in l_{\mathbf{Y}}) \\ &\propto \exp(-k \|z - \mathbf{Y}\alpha\|_2^2 + c \|\alpha\|_2^2) \end{aligned} \tag{9}$$

Here, $\lambda = c/k$ and y is in the subspace S.

For seizure detection, the EEG signals have two classes (seizure and non-seizure), so the algorithm should decide that the testing sample belongs to each class signal. For a sample y inside S, the collaborative representation can be defined as $y = \mathbf{Y}\alpha = \sum_{K=1}^{K} \mathbf{Y}_k \alpha_k$, where α_k is the coding vector for $\mathbf{Y}_k$ and $\alpha = [\alpha_1; \alpha_2; \ldots; \alpha_K]$. To represent a seizure data y having the same class label as y_k, the probability can be calculated as:

$$P(l(y) = k | l(y) \in l_{\mathbf{Y}}) \propto \exp(-\varsigma \|y - \mathbf{Y}_k \alpha_k\|_2^2) \tag{10}$$

ς is a constant.

In addition, for a testing EEG sample z outside S, the probability can be computed as:

$$\begin{aligned} P(l(z) = k | l(z) = l(y)) &= P(l(y) = k) \cdot P(l(z) \\ &= l(y) | l(y) = k) \propto \exp(-(\|z - \mathbf{Y}\alpha\|_2^2 \\ &\quad + \lambda \|\alpha\|_2^2 + \gamma \|\mathbf{Y}\alpha - \mathbf{Y}_k \alpha_k\|_2^2)) \end{aligned} \tag{11}$$

Suppose that a common data point y can be determined to maximize the joint probability and $l(z) = k$ is also independent, we can then obtain the class label of z as:

$$\begin{aligned} P(l(z) = k) &= \max P(l(z) = 1, \ldots, l(z) = K) = \max \\ &\quad \prod\nolimits_k P(l(z) = k) \propto \max \exp(-(\|z - \mathbf{Y}\alpha\|_2^2 \\ &\quad + \lambda \|\alpha\|_2^2 \frac{\gamma}{K} \sum\nolimits_{i=1}^{K} (\|\mathbf{Y}\alpha - \mathbf{Y}_k \alpha_k\|_2^2))) \end{aligned} \tag{12}$$

By applying the logarithm operator in the previous formula, the result is:

$$\begin{aligned} (\hat{\alpha}) &= \arg\min_{\alpha} \Big\{ \|z - \mathbf{Y}\alpha\|_2^2 \\ &\quad + \lambda \|\alpha\|_2^2 + \frac{\gamma}{K} \sum\nolimits_{k=1}^{K} (\|\mathbf{Y}\alpha - \mathbf{Y}_k \alpha_k\|_2^2) \Big\} \end{aligned} \tag{13}$$

where γ and λ are parameters.

So, we obtain the probability that

$$\begin{aligned} P(l(z) = k) \propto \exp(-(+\lambda \left\|\hat{\alpha}\right\|_2^2 + \left\|z - \mathbf{Y}\hat{\alpha}\right\|_2^2 \\ + \gamma \left\|\mathbf{Y}\hat{\alpha} - \mathbf{Y}_k \hat{\alpha}_k\right\|_2^2)) \end{aligned} \tag{14}$$

In many classification problems, l_1 norm has a better performance in enhancing the robustness, so the Laplacian kernel is used and the probability is re-defined as:

$$P(l(z) = l(y) | l(y) \in l_{\mathbf{Y}}) \propto \exp(-k \|z - y\|_1) \tag{15}$$

The iterative re-weighted least square algorithm is applied to solve the sparse coefficient vector $(\hat{\alpha})$ for robust ProCRC and so we have (16), as shown at the bottom of the next page.

Then, the sparse coefficient is applied to represent testing EEG signals and the residual related to seizure and non-seizure samples can be further obtained for classification.

D. Kernel-Based Robust ProCRC

As an effective data processing method, the kernel method is applied to improve the linear separability for EEG signals and the new linear representation of the testing samples are obtained after mapping. The non-linear mapping is defined as: $R^m \to R^F$. R^m refers to the original data space and R^F refers to the high-dimensional space. $\langle \Phi(y_i), \Phi(y_j) \rangle_{R^F}$ refers to the inner products for the transformed samples and is defined by the kernel function as:

$$\langle \Phi(y_i), \Phi(y_j) \rangle_{R^F} = \Phi(y_i)^T \Phi(y_j)^T = K(y_i, y_j) \tag{17}$$

Here, $K(,)$ refers to the kernel function in the original data space. After mapping, the training samples y and testing samples z become $K(\mathbf{M}, y)$ and $K(\mathbf{M}, z)$, respectively. Here, $\mathbf{M}$ is called the center matrix. The sparse representation then becomes:

$$K(\mathbf{M}, z) = K(\mathbf{M}, y)\beta \tag{18}$$

In this paper, the training data are applied to obtain matrix M. For each class $\mathbf{Y}_i = [y_i^1, y_i^2, \ldots, y_i^{n_i}]$ of the training set, the mean sample can be computed as $u_i = (\sum_{j=1}^{n_i} y_i^j)/n_i$. Then, the nearest half samples of u_i are selected to compose the matrix $\mathbf{M}_i = [u_i, y_i'^1, y_i'^2, \ldots, y_i'^{n_i}/2]$. Next, the center matrix $\mathbf{M}$ is generated as:

$$\mathbf{M} = [\mathbf{M}_1, \mathbf{M}_2, \ldots, \mathbf{M}_K] \tag{19}$$

In addition, the representation coefficient β is re-defined for the high space R^F after implementation of the kernel method as:

$$\begin{aligned} (\hat{\beta}) = (\frac{\gamma}{K} \sum\nolimits_{k=1}^{K} \overline{(K(\mathbf{M}, \mathbf{Y})_k')}^T \cdot \overline{K(\mathbf{M}, \mathbf{Y})_k}' \\ + \lambda I + K(\mathbf{M}, \mathbf{Y})^T W_{\mathbf{Y}} K(\mathbf{M}, \mathbf{Y}))^{-1} \\ K(\mathbf{M}, \mathbf{Y})^T W_{\mathbf{Y}} K(\mathbf{M}, z) \end{aligned} \tag{20}$$

After calculating the sparse representation coefficient, we can compute the residual of seizure class and non-seizure class data through the following definition:

$$\gamma_i(z) = \left\| K(\mathbf{M}, z) - K(\mathbf{M}, \mathbf{Y}_i)\hat{\beta}_i \right\|_2^2 \tag{21}$$

After mapping, the robust ProCRC is performed in the new feature space. The residuals relating to the two class signals are further compared to obtain the classification results.

E. Post-Processing

To further achieve more accurate classification results, post-processing techniques were performed on the kernel based robust ProCRC classifier outputs to obtain the classification label. These techniques included smoothing, threshold judgment, adaptive collar operation and multi-label fusion.

The moving average filter is first used to the classifier outputs to remove the accidental burrs and noise in the smoothing step. The filter is defined as

$$g(n) = \frac{1}{2C+1} \sum_{c=-c}^{C} y(n+c) \tag{22}$$

Here, y is the output of kernel based robust ProCRC as the filter input signal, g is the output signal and $2C+1$ is the smoothing length.

For this system, we define the label "0" represents a non-seizure segment and "1" represents a seizure segment. To obtain the binary labels, the threshold judgement is applied to the smoothing result g by compared to a fixed threshold r. If $g > r$, the label is set to "1" (seizure segment), otherwise, it is set to "0" (non-seizure segment). The threshold value r is determined based on the classification verification results of the training data in the training step, which is different for each patient.

For the three-channel signals used in this paper, there are three decision labels for one testing segment corresponding to each channel. Besides, in the pre-processing part, all testing segments are decomposed into three sub-signals that also generate three decision labels. Therefore, multi-label fusion is required, and the specific steps are as follows. Firstly, for the results of three channels of each sub-signal, if two or three channels decided as '1' simultaneously, the segment will be marked as '1'. In this way, three decision labels connected with the three sub-signals are obtained. Next, for the three decision labels, if three is at least one marked as '1', the segment will be considered as seizure data finally. Otherwise, the segment will be marked as non-seizure data. Fig 2 shows the details of multi-label fusion process.

For this long-term EEG database, the changes of seizure activity are not obvious at the beginning and ending stages. At the same time, smoothing causes these changes to be less obvious. These factors make some seizure segments missed. Therefore, the adaptive collar operation is used in the last step to compensate for the missing segments. In this step, both sides are extended m segments for each detected seizure event. Considering that seizure amplitude is in proportion to the seizure duration [3], so the value of m is patient specific and selected based on the initial duration of seizure activities for this technique.

IV. Experiment Results

The Freiburg EEG database was adopted to evaluate the proposed detection algorithm and the experiments were executed in the MATLAB2016 with an Inter Core processor with a 2.4Hz environment on a personal computer. To evaluate the proposed seizure detection algorithm, two types of evaluation criteria were used in this paper. The raw EEG signals were divided into 4 s segments in the pre-processing stage, so the first approach was segment-based evaluation, which contained three contents: sensitivity, recognition accuracy and specificity. The segments detected by this method were compared with the segments labeled by experts, and three measures were defined as follows:

Sensitivity: the number of seizure segments that were detected correctly by our system divided by all seizure segments labeled by experts. This measure represents the accurate ability to detect epilepsy seizure signals. Specificity: the non-seizure segments that were labeled by our system divided by all non-seizure segments labeled by experts. The specificity represents the accurate ability to detect non-seizure data. Recognition accuracy: EEG segment numbers that were distinguished correctly divided by the total EEG segments.

Although high sensitivity and specificity are the goals of most epilepsy detection systems, these two goals are contradictory. Specifically, improving sensitivity reduces specificity as non-seizure signals may be incorrectly marked as seizure signals. Meanwhile, if higher specificity is obtained, the sensitivity is reduced and some seizure activities may be missed. In addition, the quantity of non-seizure data is much longer than seizure data leading to a particular imbalance between two types of data. Hence, the appropriate methods for each stage of this algorithm are needed to solve the data imbalance problem and obtain both better sensitivity and specificity in this work.

The results evaluated by the first approach are depicted in Table II. The average results that were achieved included a sensitivity of 96.48%, recognition accuracy of 98.56% and specificity of 98.55%. For most patients (18 out of 21), the sensitivity was 100%, which means that our method could detect all seizure segments for these seizure activities. The lowest sensitivity of 66.67% was obtained in patient 1 and 19 because the short duration of seizure events resulted in unclear epilepsy activities. Besides patient 10, whose results had the lowest specificity of 91.31%, all other patients result achieved a specificity higher than 92%, with 15 patient results achieving a specificity greater than 99%.

The second approach was an event-based evaluation that focused on the number of seizure events, not the number of EEG segments. Therefore, the number of seizure activities detected by this method was compared with the number of activities labeled by experts. Here, the sensitivity was defined as the number of seizure events correctly detected by our method divided by all seizure events labeled by experts. Another criterion was the false detection rate, which was used to focus on the non-seizure data that were marked as seizure data by our system and calculated as the average number of false seizures detected per hour. The results are shown as Table III.

In total, 56 seizure events from testing set were adopted to evaluate this method and 53 events were detected correctly,

$$(\hat{\alpha}) = \arg\min_{\alpha} \left\{ \frac{\gamma}{K} \sum\nolimits_{K=1}^{K} \|\mathbf{Y}\alpha - \mathbf{Y}_k\alpha_k\|_2^2 + \lambda \|\alpha\|_2^2 + (\mathbf{Y}\alpha - z)^T W_{\mathbf{Y}} (\mathbf{Y}\alpha - z) \right\} \tag{16}$$

TABLE II
THE RESULTS EVALUATED UNDER SEGMENT STANDARDS

Patient	Sensitivity (%)	Accuracy (%)	Specificity (%)
1	66.67	92.54	92.55
2	100	99.96	99.96
3	100	99.46	99.46
4	100	99.58	99.58
5	100	98.63	98.63
6	100	99.69	99.69
7	100	99.93	99.93
8	100	99.59	99.59
9	100	97.99	97.98
10	100	91.31	91.30
11	100	97.09	97.09
12	100	99.94	99.94
13	100	99.58	99.58
14	100	99.56	99.56
15	92.73	99.40	99.43
16	100	99.49	99.49
17	100	99.95	99.95
18	100	99.15	99.15
19	66.67	97.74	97.75
20	100	99.79	99.79
21	100	99.16	99.16
Average	96.48	98.55	98.56

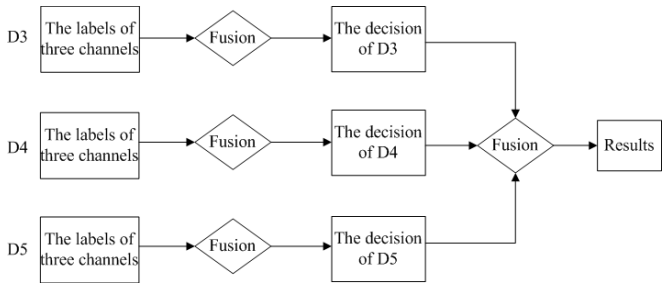

Fig. 2. The multi-label fusion steps used in the post-processing stage.

with obtaining an average sensitivity of 93.65% for all patients. The average false detection rate of 0.44/h was achieved for 21 patients, and except for patient 11 and 19 with false detection rates higher than 1/h, the false detection rates of most patients were relatively ideal. The high false detection rates of these two patients were likely caused by high-amplitude activities. In addition, 18 patients were able to correctly detect all seizure events, except patients 1, 15 and 19. For these three patients, each had one seizure event missed because the durations for the three seizure activities were too short.

TABLE III
THE RESULTS EVALUATED UNDER EVENT-BASED METHOD

Patient	Number of seizures experts marked	Number of true detections	Sensitivity (%)	False detection rate/h
1	2	1	50	0.54
2	2	2	100	0.08
3	3	3	100	0.19
4	4	4	100	0.18
5	3	3	100	0.48
6	2	2	100	0.35
7	2	2	100	0.08
8	1	1	100	0.12
9	4	4	100	0.52
10	2	2	100	0.40
11	3	3	100	1.12
12	3	3	100	0.11
13	1	1	100	0.08
14	3	3	100	0.16
15	3	2	66.67	0.06
16	3	3	100	0.15
17	4	4	100	0.11
18	2	2	100	0.96
19	2	1	50	3.21
20	3	3	100	0.11
21	4	4	100	0.30
Total	56	53	93.65	0.44

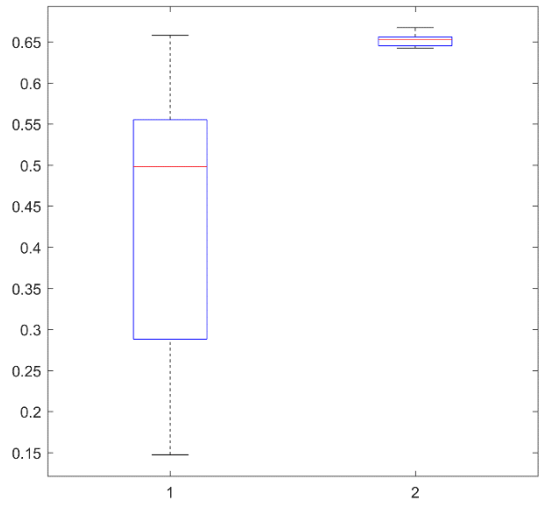

Fig. 3. The residual difference of one-minute recordings from patient 2; "1" refers to the residual difference of seizure data and "2" refers to the residual difference of non-seizure data.

V. DISCUSSIONS

Many automatic seizure detection algorithms require finding and computing suitable features and classifiers to generate classification results. However, selecting valid features is time-consuming, and the choice of classifiers is not always appropriate to best distinguish seizure from non-seizure data. In our

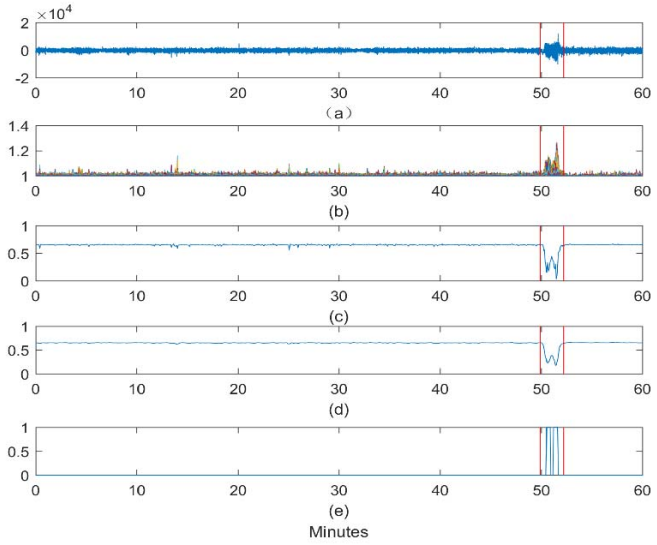

Fig. 4. The diagram of the experimental procedure for 1 h EEG data from patient 2. (a) the raw EEG signals from one channel; (b) the signals after pre-processing; (c) the results after kernel based roust probabilistic collaborative representation; (d) the results after smoothing; (e) the binary value after threshold judgment for one channel.

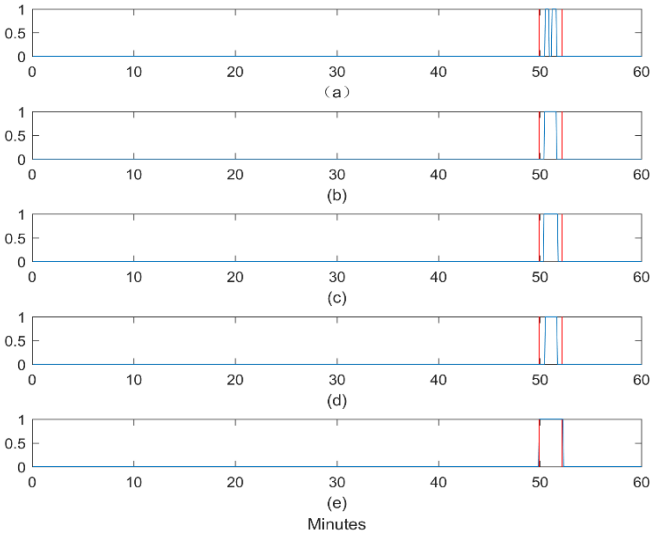

Fig. 5. The steps of multi-label fusion and the adaptive collar operation of 1 h EEG data from patient 2. (a) (b) and (c) the binary value EEG signals from the three different channels; (d) the results after multi-label fusion with the three channels; (e) the final classification results obtained by the adaptive collar operation for the three channels.

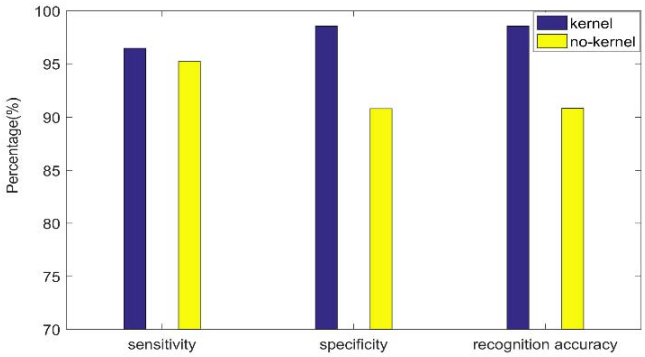

Fig. 6. The comparison of results between the robust ProCRC with kernel method and without kernel method.

proposed algorithm, only wavelet transform and GNMF are used for simple preprocessing, feature extraction and dimensionality reduction, and its computation is not complicated. In addition, in the classification step, we use the training data to represent the testing samples through kernel-based robust ProCRC algorithm and the residuals obtained related to the two class signals are compared to determine the classification results of the test samples, which is also easy to implement and has fewer parameters. Fig. 3 shows the difference between the residuals of the seizure stage and the non-seizure stage for patient 2, which can be seen that there is a significant difference and can be effectively classified.

Besides, to obtain more accurate classification results for the long-term nonstationary EEG signals, our method also applies different methods in the pre-processing and post-processing parts. The DWT is used to decompose the raw signals into five sub-signals and to provide the frequency and time information. Then, normalization based on the differential operator is further adopted to heighten the difference between the two class signals, which can capture abrupt changes and improve classification accuracy.

In the post-processing stage, as isolated incorrect detection points exist in the whole detection process because the raw EEG signals have noise in the recording process, the smoothing step is used to remove these errors. In an epileptic seizure activity, due to the slow change, it is difficult to completely detect the segments of the beginning and the end stages, and some parts may be missed. In this method, the adaptive collar operation is performed to compensate for the missed segments. These procedures are shown in Fig. 4, which include smoothing and threshold judgment, while Fig.5 shows the multi-label fusion and adaptive collar techniques.

The kernel method can transform the EEG samples into a more linear separable form through kernel mapping. In order to verify this effect, we conducted a comparison experiment using the robust ProCRC with kernel method and the robust ProCRC without kernel method, and 95.24% average sensitivity, 90.83% recognition accuracy and 90.82% average specificity were obtained through the robust ProCRC with kernel method which was shown in Fig.6. The experiment using the kernel method obtained better classification results, which suggests that the linear representation of the testing sample becomes more accurate for classification after kernel mapping.

To further quantify the performance of our method, we calculated the running time for each step: The training step had a duration of 4.902 s, while for 1 h of EEG data, it took 41.54 s to obtain classification representation in the testing stage. In the study of Yuan *et al.*, the training step had a duration of 15 s, while it took 1 min to obtain classification representation in the testing stage for 1 h of data using kernel collaborative representation [33]. Overall, the efficiency in our proposed algorithm indicates that it is feasible as a real-time seizure detection system.

Table IV provides a comparison between our method and other methods that used the Freiburg database as well. The seizure detection method introduced by [34] used

TABLE IV
COMPARISON OF METHODS EVALUATED FROM THE SAME DATABASE

Method	patient number	EEG data(h)	Sensitivity (%)	Specificity (%)	Recognition accuracy (%)	False detection rate(/h)
Yuan et al.[33]	21	595	94.41	96.97	96.87	0.26
Mahmoodian et al.[34]	20	560	95.83	96.70	96.84	0.24
Tzimourta et al.[35]	21	28.6	99.74	97.30	97.74	0.21
Geng et al.[36]	20	680	98.08	98.69	98.69	0.24
Yu et al.[37]	21	564.38	97.48	96.81	96.83	0.53
Ma et al.[38]	21	566.57	95.12	97.60	97.60	-
Zhou et al.[39]	21	-	93.70	97.20	95.40	-
Hussain et al.[40]	21	-	99.46	98.45	99.27	-
Li et al.[41]	21	564.03	97.47	96.17	96.17	0.487
Ours	21	590	96.48	98.56	98.55	0.44

cross-bispectrum for feature extraction and obtained a lower false detection rate; but the sensitivity and specificity were both lower than ours and their method only detected the EEG data of 20 patients. In the work of Tzimourta. *et al.*, they proposed a multicenter methodology applying DWT with five decomposition levels [35]. The random forest classifier was performed on 28.6 h of EEG data from 21 patients and the results were obtained with a sensitivity of 99.74% and false detection rate of 0.21/h. However, while the classification results were quite satisfactory, the experimental data contained only 28.6 h of EEG data, which was much shorter than the EEG data we used, while our results were obtained from 590 h of EEG data. Geng *et al.* introduced an efficient method using bidirectional long short-term memory neural networks and stockwell transform to perform seizure detection [36]. Even though 680 h of EEG data were used to evaluate the performance, a lower false detection rate of 0.24/h and a higher sensitivity of 98.09% were obtained, the patient 10 in this database is not considered in their method, whereas we analyzed data from all 21 patients. Furthermore, the time complexity of the training stage is about 48 s, which is relatively longer than our method with 4.902s, and their method uses far more hyperparameters than our method.

As an efficient classification method, the CRC was introduced in [33] for epileptic detection. This method was used to analyze 595 h EEG data from 21 patients and achieved a sensitivity of 94.41% and false detection rate of 0.26/h, but the specificity and sensitivity were both lower than ours. Later, in the work of Yu *et al.*, the authors proposed a method that applied kernel methodology combined with robust probabilistic collaborative representation [37]. A sensitivity of 97.48%, specificity of 96.81% and false detection rate of 0.53/h were obtained using their method which used only 52 seizures in the testing stage. Compared to their algorithm, our system evaluated 56 seizures. Even a slightly lower sensitivity was obtained, while better specificity and false detection rate were achieved. It shows that the recognition performance of our algorithm for seizure EEG data is not bad, and has a better recognition effect on non-seizure data. Tensor, as an efficient method, has been used for seizure classification. Ma *et al.* [38] calculated tensor distance as EEG feature and applied the BLDA classifier for the detection algorithm. This method obtained a sensitivity of 95.12% and specificity of 97.60% for the EEG data from 21 patients, which were lower than those obtained with our method.

Moreover, many deep learning algorithms have also been validated and evaluated on this database. Zhou *et al.* proposed a convolutional neural network (CNN) to distinguish ictal, preictal, and interictal stages for seizure detection and achieved the average accuracies of 95.4% for interictal and ictal classification using frequency domain signals, which is lower than our methods [39]. Hussain *et al.* used 1 D-convolutional long short-term memory neural networks to automatically generate customized features for better classification of ictal, interictal, and preictal segments, which achieved a highest classification accuracy of 99.27%. Even though the results were better than our method, they used 80% of the EEG segments for training the model and the remaining 20% of signals were used to test, while Our algorithm requires very little training data [40]. Li *et al.* combined the fully convolutional network and the Nested Long Short-Term Memory (NLSTM) model for end-to-end automatic seizure detection and the average sensitivity of 97.47%, specificity of 96.17%, and false detection rate of 0.487/h are yielded [41]. Compared with their method, a higher accuracy and a lower false detection rate are observed in our proposed method. Therefore, our algorithm has comparable performance with lower algorithm complexity and does not require too many labeled samples for supervised learning.

VI. CONCLUSION

An efficient system using GNMF and kernel-based robust ProCRC is proposed for epileptic seizure detection. For long-term raw EEG signals, GNMF is mainly applied to dimensionality reduction and preserve the useful EEG feature information, after which the kernel method is used to enhance the linear separable representation. After performing kernel-based robust ProCRC, the training set is applied to represent the testing samples sparsely and the residuals are compared to distinguish the two classes of EEG signals and obtain the class

label. The classification results and low time complexity also indicate the potential application value for real-time clinical application of the proposed algorithm.

REFERENCES

[1] F. Fahimi, Z. M. Yaseen, and A. El-Shafie, "Application of soft computing based hybrid models in hydrological variables modeling: A comprehensive review," *Theor. Appl. Climatol.*, vol. 128, nos. 3–4, pp. 875–903, May 2017, doi: 10.1007/s00704-016-1735-8.

[2] V. Nagaraj, A. Lamperski, and T. I. Netoff, "Seizure control in a computational model using a reinforcement learning stimulation paradigm," *Int. J. Neural Syst.*, vol. 27, no. 7, Nov. 2017, Art. no. 1750012, doi: 10.1142/S0129065717500125.

[3] S. Yuan, W. Zhou, J. Li, and Q. Wu, "Sparse representation-based EMDand BLDA for automatic seizure detection," *Med. Biol. Eng. Comput.*, vol. 55, pp. 1227–1238, Jun. 2017, doi: 10.1007/s11517-016-1587-5.

[4] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Epileptic seizure detection in EEGs using time-frequency analysis," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 5, pp. 703–710, Sep. 2009, doi: 10.1109/TITB.2009.2017939.

[5] A. R. Hassan, S. Siuly, and Y. Zhang, "Epileptic seizure detection in EEG signals using tunable-Q factor wavelet transform and bootstrap aggregating," *Comput. Methods Programs Biomed.*, vol. 137, pp. 247–259, Dec. 2016, doi: 10.1016/j.cmpb.2016.09.008.

[6] J. Gotman, "Automatic recognition of epileptic seizures in the EEG," *Electroencephalogr. Clin. Neurophysiol.*, vol. 54, no. 5, pp. 530–540, Nov. 1982, doi: 10.1016/0013-4694(82)90038-4.

[7] V. Srinivasan, C. Eswaran, and N. Sriraam, "Approximate entropy-based epileptic EEG detection using artificial neural networks," *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, no. 3, pp. 288–295, May 2007, doi: 10.1109/TITB.2006.884369.

[8] J. Wu, T. Zhou, and T. Li, "Detecting epileptic seizures in EEG signals with complementary ensemble empirical mode decomposition and extreme gradient boosting," *Entropy*, vol. 22, no. 2, p. 140, Jan. 2020, doi: 10.3390/e22020140.

[9] K. A. Khan, S. P. Pan, Y. U. Khan, and O. Farooq, "A hybrid local binary pattern and wavelets based approach for EEG classification for diagnosing epilepsy," *Expert Syst. Appl.*, vol. 140, Feb. 2020, Art. no. 112895, doi: 10.1016/j.eswa.2019.112895.

[10] T. Karacsony, A. M. Loesch-Biffar, C. Vollmar, S. Noachtar, and J. P. S. Cunha, "A deep learning architecture for epileptic seizure classification based on object and action recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 4117–4121.

[11] 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, doi: 10.1109/TBCAS.2019.2929053.

[12] I. Ullah, M. Hussain, E.-U.-H. Qazi, and H. Aboalsamh, "An automated system for epilepsy detection using EEG brain signals based on deep learning approach," *Expert Syst. Appl.*, vol. 107, pp. 61–71, Oct. 2018, doi: 10.1016/j.eswa.2018.04.021.

[13] Y. Zhao, G. Zhang, C. Dong, Q. Yuan, F. Xu, and Y. Zheng, "Graph attention network with focal loss for seizure detection on electroencephalography signals," *Int. J. Neural Syst.*, vol. 31, no. 7, Jul. 2021, Art. no. 2150027, doi: 10.1142/S0129065721500271.

[14] L. Duan, Z. Wang, Y. Qiao, Y. Wang, Z. Huang, and B. Zhang, "An automatic method for epileptic seizure detection based on deep metric learning," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 5, pp. 2147–2157, May 2022, doi: 10.1109/JBHI.2021.3138852.

[15] A. Subasi, "EEG signal classification using wavelet feature extraction and a mixture of expert model," *Expert Syst. Appl.*, vol. 32, no. 4, pp. 1084–1093, May 2007, doi: 10.1016/j.eswa.2006.02.005.

[16] M. M. Rashid, F. Johnson, and A. Sharma, "Identifying sustained droug-ht anomalies in hydrological records: A wavelet approach," *J. Geophys. Res., Atmos.*, vol. 123, pp. 7416–7432, Jul. 2018, doi: 10.1029/2018JD028455.

[17] W. He, H. Zhang, and L. Zhang, "Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3909–3921, Jul. 2017, doi: 10.1109/TGRS.2017.2683719.

[18] C. Leng, G. Cai, D. Yu, and Z. Wang, "Adaptive total-variation for non-negative matrix factorization on manifold," *Pattern Recognit. Lett.*, vol. 98, pp. 68–74, Oct. 2017, doi: 10.1016/j.patrec.2017.08.027.

[19] Z. Shu *et al.*, "Parameter-less auto-weighted multiple graph regularized nonnegative matrix factorization for data representation," *Knowl.-Based Syst.*, vol. 131, pp. 105–112, Sep. 2017, doi: 10.1016/j.knosys.2017.05.029.

[20] Z. Shu, C. Zhao, and P. Huang, "Local regularization concept factorizati-on and its semi-supervised extension for image representation," *Neurocomputing*, vol. 158, pp. 1–12, May 2015, doi: 10.1016/j.neucom.2015.02.014.

[21] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Dec. 2011, doi: 10.1109/TPAMI.2010.231.

[22] Y. Chi and F. Porikli, "Classification and boosting with multiple collaborative representations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1519–1531, Aug. 2014, doi: 10.1109/TPAMI.2013.236.

[23] Y. Chi and F. Porikli, "Connecting the dots in multi-class classification: From nearest subspace to collaborative representation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3602–3609.

[24] X. Jiang and J. Lai, "Sparse and dense hybrid representation via dictionary decomposition for face recognition," *IEEE Trans. Pattern Analysis Mach. Intell.*, vol. 37, no. 5, pp. 1067–1079, May 2015, doi: 10.1109/TPAMI.2014.2359453.

[25] W. Deng, J. Hu, and J. Guo, "Extended SRC: Undersampled face recognition via intraclass variant dictionary," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1864–1870, Jan. 2012, doi: 10.1109/TPAMI.2012.30.

[26] S. Dongcheng, X. Yidan, and D. Guangyi, "Research on technology of compressed sensing for face recognition," in *Proc. 7th Int. Conf. Image Graph.*, Jul. 2013, pp. 505–508.

[27] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009, doi: 10.1109/TPAMI.2008.79.

[28] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 471–478.

[29] S. Cai, L. Zhang, W. Zuo, and X. Feng, "A probabilistic collaborative representation based approach for pattern classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2950–2959.

[30] D. Gajic, Z. Djurovic, J. Gligorijevic, S. Di Gennaro, and I. Savic-Gajic, "Detection of epileptiform activity in EEG signals based on time-frequency and non-linear analysis," *Frontiers Comput. Neurosci.*, vol. 9, p. 38, Mar. 2015, doi: 10.3389/fncom.2015.00038.

[31] O. Faust, U. R. Acharya, H. Adeli, and A. Adeli, "Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis," *Seizure*, vol. 26, pp. 56–64, Mar. 2015, doi: 10.1016/j.seizure.2015.01.012.

[32] S. Grewal and J. Gotman, "An automatic warning system for epileptic seizures recorded on intracerebral EEGs," *Clin. Neurophysiol.*, vol. 116, no. 10, pp. 2460–2472, Oct. 2005, doi: 10.1016/j.clinph.2005.05.020.

[33] S. Yuan *et al.*, "Kernel collaborative representation-based automatic seizure detection in intracranial EEG," *Int. J. Neural Syst.*, vol. 25, no. 2, Mar. 2015, Art. no. 1550003, doi: 10.1142/S0129065715500033.

[34] N. Mahmoodian, A. Boese, M. Friebe, and J. Haddadnia, "Epileptic seizure detection using cross-bispectrum of electroencephalogram signal," *Seizure*, vol. 66, pp. 4–11, Jul. 2019, doi: 10.1016/j.seizure.2019.02.001.

[35] K. D. Tzimourta *et al.*, "A robust methodology for classification of epileptic seizures in EEG signals," *Health Technol.*, vol. 9, no. 2, pp. 135–142, Mar. 2019, doi: 10.1007/s12553-018-0265-z.

[36] M. Geng, W. Zhou, G. Liu, C. Li, and Y. Zhang, "Epileptic seizure detection based on stockwell transform and bidirectional long short-term memory," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 3, pp. 573–580, Mar. 2020, doi: 10.1109/TNSRE.2020.2966290.

[37] Z. Yu *et al.*, "Automatic seizure detection based on kernel robust probabilistic collab-orative representation," *Med. Biol. Eng. Comput.*, vol. 57, pp. 205–219, Sep. 2019, doi: 10.1007/s11517-018-1881-5.

[38] D. Ma *et al.*, "The automatic detection of seizure based on tensor distance and Bayesian linear discriminant analysis," *Int. J. Neural Syst.*, vol. 31, no. 5, May 2021, Art. no. 2150006, doi: 10.1142/S0129065721500064.

[39] M. Zhou *et al.*, "Epileptic seizure detection based on EEG signals and CNN," *Frontiers Neuroinform.*, vol. 12, p. 95, Dec. 2018, doi: 10.3389/fninf.2018.00095.

[40] W. Hussain, M. T. Sadiq, S. Siuly, and A. U. Rehman, "Epileptic seizure detection using 1 D-convolutional long short-term memory neural networks," *Appl. Acoust.*, vol. 177, Jun. 2021, Art. no. 107941, doi: 10.1016/j.apacoust.2021.107941.

[41] 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, doi: 10.1142/S0129065720500197.