A Multi-View Deep Learning Framework for EEG Seizure Detection

Ye Yuan, Guangxu Xun, Kebin Jia, *Member, IEEE*, and Aidong Zhang, *Fellow, IEEE*

***Abstract*—The recent advances in pervasive sensing technologies have enabled us to monitor and analyze the multi-channel electroencephalogram (EEG) signals of epilepsy patients to prevent serious outcomes caused by epileptic seizures. To avoid manual visual inspection from long-term EEG readings, automatic EEG seizure detection has garnered increasing attention among researchers. In this paper, we present a unified multi-view deep learning framework to capture brain abnormalities associated with seizures based on multi-channel scalp EEG signals. The proposed approach is an end-to-end model that is able to jointly learn multi-view features from both unsupervised multi-channel EEG reconstruction and supervised seizure detection via spectrogram representation. We construct a new autoencoder-based multi-view learning model by incorporating both inter and intra correlations of EEG channels to unleash the power of multi-channel information. By adding a channel-wise competition mechanism in the training phase, we propose a channel-aware seizure detection module to guide our multi-view structure to focus on important and relevant EEG channels. To validate the effectiveness of the proposed framework, extensive experiments against nine baselines, including both traditional handcrafted feature extraction and conventional deep learning methods, are carried out on a benchmark scalp EEG dataset. Experimental results show that the proposed model is able to achieve higher average accuracy and f1-score at 94.37% and 85.34%, respectively, using 5-fold subject-independent cross validation, demonstrating a powerful and effective method in the task of EEG seizure detection.**

***Index Terms*—Deep learning, epileptic seizure, electroencephalogram, multi-view learning, feature extraction.**

I. Introduction

EPILEPSY is the fourth most common neurological disorder, and there are approximately 50 million people affected by epilepsy worldwide [1]. Epilepsy is characterized by unprovoked seizures associated with sudden irregular neuronal discharges in the brain [2] and can result in other health problems, even death [3], [4]. Electroencephalogram (EEG) is one of the commonly used methods to monitor, detect, and diagnose epileptic seizures due to its affordability and high temporal resolution [5]. EEG signals often contain multiple channels which measure brain electrical activities through several electrodes placed in different locations of the brain. According to previous work, EEG signals are usually collected with electrodes placed on the scalp or exposed surface of the brain, referred to as scalp EEG and intracranial EEG, respectively. Most studies of EEG seizure detection prefer to utilize the scalp EEG since it can be acquired with non-invasive technologies [6]. In clinical practice, EEG readings require highly-trained professionals in neurology. However, the long-term EEG visual inspection is extremely time-consuming and laborious for physicians to identify characteristic patterns of seizures from multiple channels [7]. Thus, automatic EEG seizure detection has garnered great interest among researchers in neuro-informatics.

Recently, numerous studies have been conducted on automatic EEG seizure detection. In general, the task of automatic EEG seizure detection can be divided into two main stages: feature extraction and classification [8]. In particular, feature extraction is key, since its aim is to characterize distinctive EEG patterns, which directly affect the performance of the classifier. Towards this end, most of the existing work employs expert handcrafted features to detect EEG seizures, which can be broadly classified into two categories, namely temporal-domain features and spectral-domain features [9]. The design of these handcrafted features, though different in various aspects, often involves a significant amount of expert knowledge to derive interpretable representations from data. For instance, Fourier analysis [10] is a typical signal processing tool to extract frequency contents from signals. Since brain abnormality is often reflected by increased amplitude and frequency changes in EEG signals, the results of different Fourier transform modifications could be used as features for EEG seizure detection [11], [12]. In addition, principal component analysis (PCA) [13], a statistical procedure to explore distinctive features using orthogonal transformation, has also been frequently employed for reducing the redundancy in feature space to improve the accuracy in EEG seizure detection [14], [15]. More recently, deep learning algorithms have shown promising results in medical image and signal processing, due to its strong capability of learning high-level representations from natural signals [16]. In EEG seizure

Manuscript received January 23, 2018; revised August 7, 2018; accepted September 15, 2018. Date of publication September 23, 2018; date of current version January 2, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61672064, in part by the Beijing Natural Science Foundation under Grant 4172001, in part by the Science and Technology Project of Beijing Municipal Education Commission under Grant KZ201610005007, in part by the Beijing Laboratory of Advanced Information Networks under Grant 040000546617002, and in part by the China Scholarship Council Fund under Grant 201606540008. *(Corresponding author: Ye Yuan.)*

Y. Yuan and K. Jia are with the College of Information and Communication Engineering, Beijing University of Technology, Beijing 100124, China (e-mail: yuanye91@emails.bjut.edu.cn; kebinj@bjut.edu.cn).

G. Xun and A. Zhang are with the Department of Computer Science and Engineering, State University of New York at Buffalo, Buffalo, NY 14260 USA (e-mail: guangxux@buffalo.edu; azhang@buffalo.edu).

Digital Object Identifier 10.1109/JBHI.2018.2871678

detection, the features extracted by deep learning models, such as convolutional neural networks (CNN) and stacked autoencoders (SA), have proven to be more robust than the handcrafted features to achieve better detection performance [17], [18]. Thus, we adopt deep learning-based EEG seizure detection as the central topic of this study.

Despite the promising results reported by previous deep EEG seizure detection approaches, there are still several challenges to be addressed. First, most deep learning-based methods fail to explicitly incorporate the inherent relationships among EEG channels in their objective functions. On one hand, there exist some irrelevant channels since the underlying region of some channels might not pick up seizure activity [19]. On the other hand, the features extracted from multiple channels may contain redundant information that does not help discriminate between patterns, known as the curse of dimensionality [20]. Although the standard deep learning structure can relatively reduce such influence of the redundant and irrelevant features, it may not be enough to unleash the power of multi-channel information. Moreover, in practice, both imbalanced and unlabeled data issues would potentially undermine the effectiveness of deep learning approaches, due to the limited number of seizure events and the high costs of labeling the data [18], [21].

To handle the aforementioned challenges, in our previous work [22], we presented a deep learning method to extract brain abnormalities associated with seizures based on multi-channel scalp EEG signals. Specifically, we segmented EEG signals into fragments, and adopted short-time Fourier transform (STFT) as preprocessing to express the EEG fragments in the time-frequency domain, namely EEG spectrogram representation. Then, we trained a cross-patient EEG seizure detector using deep learning technologies to extract hidden features from the preprocessed spectral images. The generated 2-D spectrogram fragments are suitable for latent feature extraction using deep learning, since they can provide more detailed contextual information in the time-frequency domain than the time domain.

In this paper, we further extend our previous work [22] to jointly learn multi-view deep features from both unsupervised multi-channel EEG reconstruction and supervised seizure detection. Our framework is described in Fig. 1. Firstly, we propose a multi-view autoencoder to unsupervisedly learn features by incorporating both intra and inter correlations of EEG channels, referred to as intra-channel and cross-channel features, respectively. We conduct experiments with two types of autoencoder: denoising autoencoder and convolutional autoencoder. Secondly, to reach a better understanding of the seizure patterns among EEG channels, we propose a channel-aware seizure detection module that guides our model to focus on important and relevant EEG channels via a channel-wise competition mechanism. Finally, we train an end-to-end model consisted of both multi-view reconstruction and detection objectives, to provide complementary information towards each other.

In summary, the main contributions of this paper, compared with our preliminary model [22], are as follows:

- We propose a unified deep learning framework to jointly extract multi-view features from both unsupervised multi-channel EEG reconstruction and supervised seizure detection. Our model can also be easily extended to a semi-supervised manner.
- We construct a new autoencoder-based multi-view learning model by incorporating both inter and intra correlations of EEG channels to unleash the power of multi-channel information.
- We develop a channel-aware seizure detection module that guides our model to focus on important and relevant EEG channels via a channel-wise competition mechanism.
- We empirically demonstrate that the proposed model outperforms nine existing seizure detection methods on different evaluation criteria.

The rest of the paper is organized as follows: In the next section, we discuss the connection of our proposed method to related work. Our proposed methodology is then described in Section III. Experimental results of our framework are presented and discussed in Section IV. Section V gives the conclusions.

II. Related Work

The problem and methodologies presented in this paper are related to the following two research areas: EEG seizure detection and multi-view learning.

A. EEG Seizure Detection

Though the task of automatic EEG seizure detection has been extensively studied, there are still many challenges to address [6], [8], [9], [17], [23]. Two different trends are mainly included in this task, named online seizure prediction and offline seizure detection. Online prediction systems tend to predict the occurrence of seizures for prompt treatment and intervention options, while offline detection aims at analyzing recordings to replace the need for laborious visual inspection. For the purpose of offline seizure detection, the literature mainly considers the EEG fragment as being ictal (containing seizure) or non-ictal (normal). To differentiate between events, the acquired EEG fragment is represented as a function of features, increasing the discriminative power of the classifier [24], [25]. This study focuses on the problem of offline seizure detection. Below we briefly discuss some representative EEG seizure detection methods in this area.

1) Handcrafted Feature Extraction: Various methods for handcrafted feature extraction for EEG seizure detection have been proposed. Several researchers propose to combine expert handcrafted features with a classifier, e.g., support vector machine (SVM) [26] or neural networks (NN) [27], to build a multi-stage EEG seizure detection system. Firstly, different types of features have been extracted in the time domain, such as phase [28], [29], entropy [30], and other statistical measurements [31]. Secondly, various spectral features have also been widely used. Both the Fourier transform [11], [12] and wavelet transform [32]–[34] can be exploited for this purpose. Thirdly, there has been a growing interest in introducing time-frequency image descriptors to EEG seizure detection [35]–[37]. These methods attempt to capture the non-stationary behavior of EEG signals by integrating both image and signal related features.

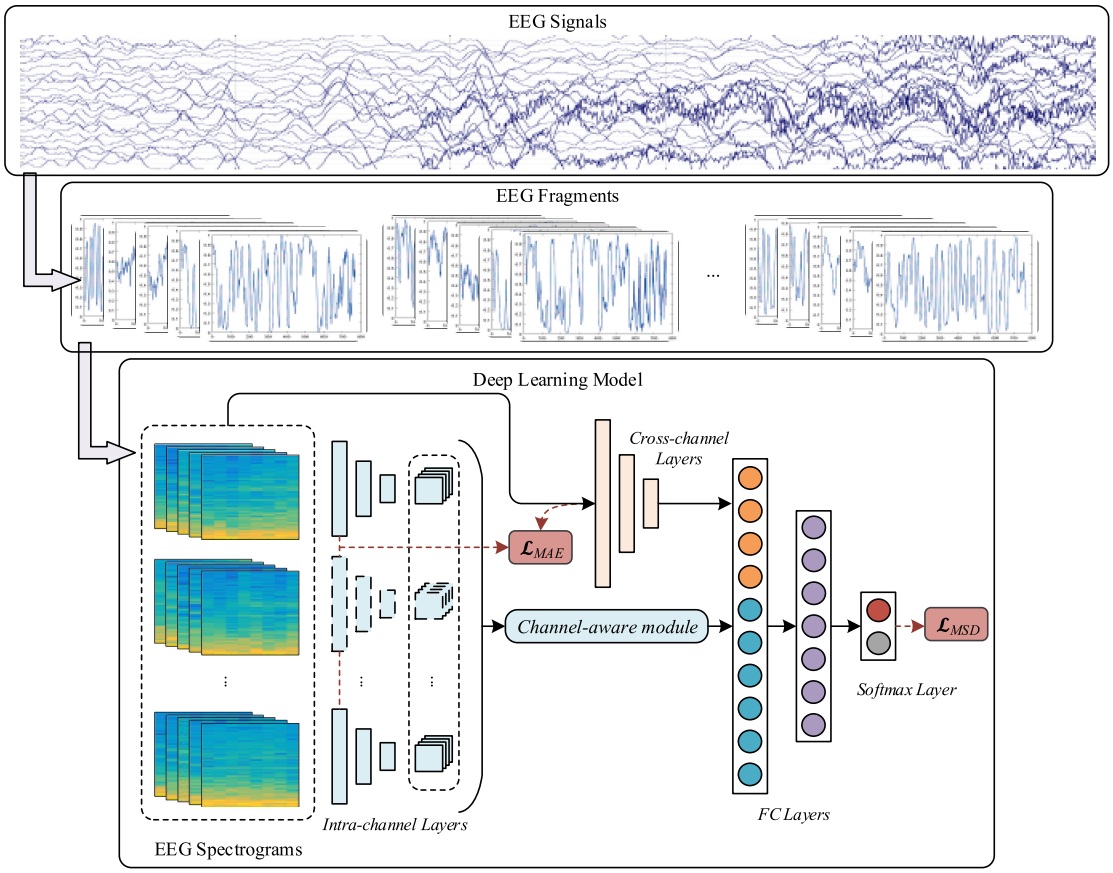


Fig. 1. Schematic illustration of the overall approach pipeline. In this framework, we propose to jointly learn multi-view deep features from both unsupervised multi-channel EEG reconstruction $\mathcal{L}_{MAE}$ and supervised seizure detection $\mathcal{L}_{MSD}$.

Considering the fact that a typical EEG system often contains multiple electrode channels, the dimension of features may grow substantially with the increase in the number of channels [38]. To avoid the curse of dimensionality, a number of feature selection (or channel selection) methods have been proposed to reduce feature dimensions by eliminating the redundant and irrelevant features before training the classifier. Numerous techniques exploiting this idea have been applied for multi-channel EEG detection, such as PCA [14], [15], individual component analysis (ICA) [39], Laplacian eigen maps [40], and multi-dimensional scaling [41]. By selecting a smaller subset of useful features, the classifier performance could be improved.

The design of these handcrafted features, though different in various aspects, typically requires a significant amount of domain knowledge and assumes that the data are clean and artifact free [9]. These features also would not guarantee consistent good performance using multi-stage training procedures to make all the components work together.

2) Deep Learning: More recently, deep learning technologies have been adopted to automatically extract features from raw high-dimensional EEG data. By constructing multi-layer neural networks, deep learning is able to reduce the influence of irrelevant features while increasing the influence of relevant features during end-to-end training. Several researchers have studied the potential of deep learning in EEG seizure detection and validated that deep learning can achieve better classification performance than handcrafted feature engineering. Some researchers apply stacked autoencoders utilizing unsupervised EEG features to detect seizures [42]–[45]. [46] present a holistic solution of seizure detection using stacked autoencoders in a semi-supervised manner. Second, CNN have also been extensively employed to analyze EEG seizure signals [47]–[49]. Different from the above approaches, which apply standard deep learning models, in this paper, we modify the deep learning structure by incorporating multi-view information to enhance the performance of EEG seizure detection. Moreover, [50] propose a wavelet-based context learning approach for EEG seizure detection. The authors generate EEG scalograms as preprocessing and then present a multi-feature fusion approach using a denoising autoencoder. This method combines deep learning with handcrafted engineering, but cannot be trained end-to-end. In contrast, we propose a channel-aware seizure detection module to train a unified deep learning model in effectiveness.

B. Multi-View Learning

With the recent advancement in artificial intelligence, multi-view learning has drawn significant attention in a wide range of applications. The goal of multi-view learning is to learn features from multiple perspectives to improve the generalization

performance [51]. Traditionally, canonical correlation analysis (CCA) and its kernel extensions have been commonly adopted in early studies of multi-view learning technologies [52]–[54]. Several advanced algorithms have been incorporated to improve the learning performance, including topic learning [55], sparse coding [56], [57], non-negative matrix factorization [58], [59], and Markov networks [60], [61]. Recently, inspired by the success of deep neural networks, some hybrid models that combine multi-view learning with deep learning have been proposed to learn shared representations. Specifically, some variants of the deep Boltzmann machine (DBM) have been proposed to model the joint distribution over visual, auditory, and text input data [62]–[64]. Autoencoders have been extended to discover correlations between hidden representations of image and text [65]–[67]. CNN-based architectures that combine information from multiple sensors have also been presented for the application of 3-D shape recognition [68] and human activity recognition [69]. Although these methods integrate deep learning in the multi-view learning framework, new chances to develop multi-view deep learning still exist in terms of methodologies and applications [51]. Different from the above models, we propose to learn multi-view deep representations from multi-channel EEG spectrograms by considering both inter and intra correlations of EEG channels. Furthermore, we extract multi-view features by explicitly considering both the EEG reconstruction and seizure detection errors to unleash the power of multi-channel information.

III. Methodology

A. EEG Spectrogram Representation

In biosignal processing, the waveform pattern associated with physiological meanings is related to an interval rather than a certain data point [22], [70]. In practice, a moving window analysis is often performed to split the raw EEG data into segments of smaller duration for feature extraction. Towards this end, we segment EEG signals into several fixed-length slots by sliding a l-length window. In order to further express EEG signals, we employ spectrogram representation to transform segments into the time-frequency domain. The generated 2-D spectrogram can be regarded as an image shown in Fig. 1, which is suitable for latent feature extraction using deep learning, since the detailed contextual characteristics in the time-frequency domain would be learned more easily by deep neural networks than those in the time domain [22], [50], [71].

Formally, given a single-channel EEG fragment $x(t)$, the spectrogram $x_s \in \mathbb{R}^{n_f \times n_t}$ is obtained by squaring the magnitude of complex-valued STFT [72], as follows:

$$\begin{aligned} x_s(\tau, w) &= |STFT_x(\tau, w)|^2 \\ &= |\int_{-\infty}^{\infty} x(t) w(t-\tau) e^{-jwt} dt|^2, \end{aligned} \tag{1}$$

where τ is the time index to localize signal and take Fourier transform with the zero-centered window function $w(t)$. Here in Eq. (1), the Blackman window [73] is adopted as the STFT window function to derive EEG spectrograms. In this way, we can obtain the time-varied frequency content of epileptic EEG signals, and further extract 2-D features using deep learning methods. It is worth noting that there are different settings of the window length l in the literature. Typically, the duration of seizure for clinical purposes is usually less than 10 seconds [74], whereas the segments of 1 to 30 seconds are usually adopted for different EEG seizure detection methods [17]. Thus, we set $l = 3$ seconds as default to be the length of the sliding window. The detailed discussions of different window lengths are provided in Section IV-H. Moreover, to simplify the notification, in the context of our model, we denote $x \in \mathbb{R}^{n_f \times n_t}$ as the input spectrogram.

B. EEG Multi-View Feature Learning

In the task of learning representations from multi-channel signals, simply concatenated raw features may not be enough to reach accurate and robust performance. To unleash the power of multi-channel information, we unsupervisedly learn latent seizure characteristics from different perspectives through an autoencoder-based model. Taking the advantage of the autoencoders, the proposed model is able to recognize instances rather than discriminating among classes, which has proved to work well for the class imbalance problem [75], [76]. In our model, we attempt to extract a set of deep features from EEG spectrograms considering both inter and intra correlation of EEG channels, i.e., cross-channel and intra-channel features. These extracted features are then used to jointly reconstruct the input. Formally, given a C-channel EEG spectrogram $x = \{x_1, \ldots, x_C\}$, the cross-channel autoencoder can be expressed by learning an encoder and a decoder based on forward-propagation:

$$\hat{x}_r = Decoder_{cross}(Encoder_{cross}(x_{1:C})), \tag{2}$$

where $\hat{x}_r \in \mathbb{R}^{C(n_f \times n_t)}$ is the reconstructed spectrogram. Similarly, for each x_c, we can obtain its reconstruction vector $\hat{x}_c \in \mathbb{R}^{n_f \times n_t}$ by feeding an intra-channel autoencoder, as follows:

$$\hat{x}_c = Decoder_{intra}(Encoder_{intra}(x_c)). \tag{3}$$

In general, the training objective of our multi-view feature extraction is to minimize the reconstruction error of input data. Given an unlabeled training dataset $\{x^{(i)}, i = 1, 2, \ldots, m\}$, the cost function of the multi-view autoencoder (MAE) is defined as:

$$J_{MAE} = \frac{1}{m} \sum_{i=1}^{m} \left[\mathcal{L}_{\mathbb{H}}(x^{(i)}, (\hat{x}_r^{(i)})) + \sum_{c=1}^{C} \mathcal{L}_{\mathbb{H}}(x_c^{(i)}, (\hat{x}_c^{(i)})) \right], \tag{4}$$

where $\mathcal{L}_{\mathbb{H}}$ is measured by cross-entropy loss as follows:

$$\begin{aligned} \mathcal{L}_{\mathbb{H}}(x^{(i)}, \hat{x}^{(i)}) = -\sum_{j=1}^{n} \Big[& x_j^{(i)} \log(\hat{x}_j^{(i)}) \\ & + (1 - x_j^{(i)}) \log(1 - \hat{x}_j^{(i)}) \Big]. \end{aligned}$$

In our model, we employ two types of autoencoder to learn our multi-view model.

1) Denoising Autoencoder: One easy way to capture multi-view features is adopting a fully-connected (FC) autoencoder. Specifically, both the cross-channel and intra-channel operators, i.e., Eq. (2) and Eq. (3), can be parameterized by two denoising autoencoders (DA) [77], as follows:

$$\hat{x}_r = f(W_r^{'} f(W_r \bar{x}_{1:C} + b_r) + b_r^{'})$$
$$\hat{x}_c = f(W_c^{'} f(W_c \bar{x}_c + b_c) + b_c^{'}),$$

where $W^{'} = W^{\top}$ is referred to as tied weights, $f(\cdot)$ is the activation function, $W_r \in \mathbb{R}^{p \times Cn}$, $W_c \in \mathbb{R}^{p \times n}$, $b_r, b_c \in \mathbb{R}^p$, and $b_r^{'}, b_c^{'} \in \mathbb{R}^n$ are the learnable parameters. Here we reshape the spectrogram of each channel into a vector space, i.e., $n = n_f n_t$, to fit the input of DA. In order to uncover robust representations, DA randomly corrupts input data by sampling $\bar{x} \sim P_{corr}(\bar{x} \mid x)$ before feature encoding. In our model, we adopt a multi-layer autoencoder with a sparsity constraint, denoted as stacked sparse denoising autoencoder (SSDA) [78], to unleash the power of multi-view feature extraction.

2) Convolutional Autoencoder: In order to further preserve spatial locality of the spectrogram, a convolutional autoencoder (ConvA) [79], can be employed. The benefit of applying ConvA is to utilize layers with convolving filters to share weights among all locations in the input. Specifically, given a C-channel spectrogram input x, the latent representation of k-th feature map is obtained by:

$$h_r^k = f\left(\sum_{c=1}^{C} x_c * W_{rc}^k + b_r^k\right),$$

where $*$ denotes the 2D convolution, W_{rc}^k denotes the filter of k-th feature map in c-channel, and b_r^k denotes the bias vector. All the extracted features are then flattened to represent the cross-channel representation $h_r \in \mathbb{R}^p$. The reconstruction of the input x can be passed into a convolutional layer with C feature maps. Specifically, the c-th feature map $\hat{x}_r^c$ is obtained using:

$$\hat{x}_r^c = f\left(\sum_{k \in H_r} h_r^k * \tilde{W}_{rc}^k + c_r^c\right),$$

where c_r^c is the bias vector of c-th feature map, $\tilde{W}$ denotes the flip operation over both dimensions of the weights, H_r identifies the group of latent feature maps. Similarly, for each channel spectrogram input x_c, the latent representation of k-th feature map is obtained by:

$$h_c^k = f(x_c * W_{cc}^k + b_{cc}^k),$$

W_{cc}^k and b_{cc}^k are the parameters to be learned. We then flatten all the extracted features to represent the intra-channel representation $h_c \in \mathbb{R}^p$. The reconstruction of input x_c is then obtained using:

$$\hat{x}_c = f\left(\sum_{k \in H_c} h_c^k * \tilde{W}_{cc}^k + c_{cc}\right),$$

where c_{cc} is the bias vector of channel c, H_c denotes the group of latent feature maps. Note that the dimensions of all the learnable parameters rely on the structure of ConvA. In our model, as shown in Fig. 2, we adopt a multi-layer ConvA with pooling and FC operators. The detailed structure configuration is given in Section IV-D.

C. Channel-Aware Seizure Detection

In order to guide our model to focus on important and relevant EEG channels during seizure detection, we propose a channel-aware module to enhance the capability of hidden feature learning among all the channels. In particular, we propose to incorporate sparsity and selectivity constraints in the end-to-end training phase via a channel-wise mutual competition. In the feedforward procedure, after computing the intra-channel representation $\{h_1, h_2, \ldots, h_C\}$, we identify the $q = \alpha C$ largest activated channels by sorting the defined response energy R_c of each channel c, as follows:

$$R_c = ||h_c - \hat{\rho_c}||^2, \tag{5}$$

where h_c denotes the response stimulated by sample vector x_c, $\hat{\rho_c}$ is the average activation vector of the c-th channel, and α is the predefined channel-aware rate. Intuitively, Eq. (5) represents the strength of response given the input data, where the salience activation of hidden neurons in the c-th channel results in the high value of R_c.

Subsequently, we keep the neurons of the identified q channels activated, while setting the activations of all the neurons within other channels to zero. It means that no gradients will flow directly from the unidentified channels in the back-propagation training step, since they are made inactive. We can finally obtain a competitive intra-channel vector, denoted as $\bar{h}_c \in \mathbb{R}^{Cp}$, for the next feedforward step. Note that once the channel-aware networks have been trained, no ranking and selecting (i.e., the training penalty) is required in the testing phase since the hidden neurons are well trained to be distinctive among channels. Different from the handcrafted feature selection strategy which manually selects specific features (or channels) to conduct multi-stage training, our module incorporates channel information in an end-to-end training manner. In addition, our channel-aware module also acts as a regularizer that prevents the useful hidden neurons when concatenating channel features. This training strategy is closely related to Dropout [80], a widely used regularization technique in deep learning. Unlike Dropout that sets neurons to zero randomly, our module utilizes channel patterns and sets neurons to zero via channel-wise ranking. Given the competitive intra-channel vector $\bar{h}_c$, we concatenate it with the cross-channel vector h_r to derive a fusional hidden representation $h_u \in \mathbb{R}^u$ defined as:

$$h_u = f(W_u[h_r; \bar{h}_c] + b_u),$$

where $W_u \in \mathbb{R}^{u \times (C+1)p}$ and $b_u \in \mathbb{R}^u$ are the weight matrix and bias vector, respectively. The fusional vector h_u is then fed through the softmax layer for seizure detection:

$$\hat{y} = softmax(W_s h_u + b_s),$$

where $W_s \in \mathbb{R}^{|class| \times u}$ and $b_s \in \mathbb{R}^{|class|}$ are the learnable parameters. To minimize the detection error over training samples based on the extracted features, given a labeled training

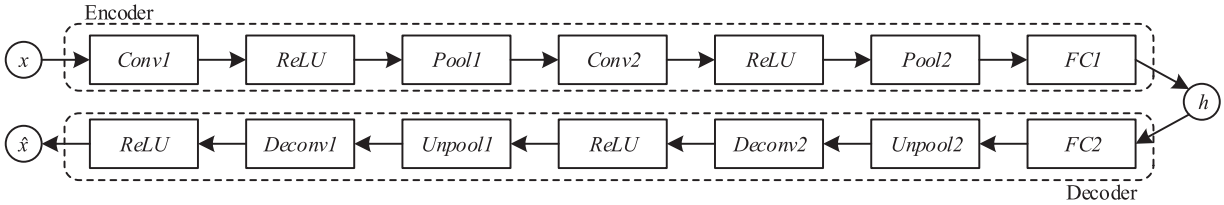


Fig. 2. Structure of multi-layer convolutional autoencoders (ConvA). In general, ConvA can be constructed by a series of convolutional-nonlinear-pooling cells with multiple filter maps.

dataset $\{x^{(i)}, y^{(i)}, i = 1, 2, \ldots, m\}$, the cost function of multi-view seizure detection (MSD) is defined as:

$$J_{MSD} = \frac{1}{m}\sum_{i=1}^{m} \mathcal{L}_{\text{IH}}(y^{(i)}, \hat{y}^{(i)}), \tag{6}$$

where $\mathcal{L}_{\text{IH}}$ is measured by cross-entropy loss as follows:

$$\mathcal{L}_{\text{IH}}(y^{(i)}, \hat{y}^{(i)}) = -\sum_{j=1}^{|class|} [y_j^{(i)} \log(\hat{y}_j^{(i)}) + (1 - y_j^{(i)}) \log(1 - \hat{y}_j^{(i)})].$$

D. Unified Training Procedure

The training objective of our end-to-end framework can be obtained by integrating the two cost functions J_{MAE} in Eq. (4) and J_{MSD} in Eq. (6). The final cost function of our multi-view model $J_{\text{Multi-View}}$ is defined as:

$$J_{\text{Multi-View}}(W_{r,c,p,u,s}, b_{r,c,p,u,s}) = \frac{1}{m}\sum_{i=1}^{m}\Big\{\beta_1 \mathcal{L}_{\text{IH}}(y^{(i)}, \hat{y}^{(i)}) + \beta_2\Big[\mathcal{L}_{\text{IH}}(x^{(i)}, (\hat{x}_r^{(i)})) + \sum\nolimits_{c=1}^{C} \mathcal{L}_{\text{IH}}(x_c^{(i)}, (\hat{x}_c^{(i)}))\Big]\Big\}, \tag{7}$$

where β_1 and β_2 denote the hyper-parameters to adjust the contribution of each cost function. In our model, we use $ReLU(z) = \max(z, 0)$ as the activation function in our model. By combining the two cost functions, the unsupervised objective is able to impact the network training when supervised learning took place. Moreover, due to the flexibility of our objective function in Eq. (7), it is easy to augment more unlabeled training data as auxiliary information in a semi-supervised learning manner.

IV. Experimental Results

A. Dataset Description

The scalp EEG dataset we use is the CHB-MIT dataset collected from the Children's Hospital Boston [81]. This dataset is open access available and can be downloaded from the PhysioNet [82]. In the CHB-MIT dataset, the multi-channel EEG signals are recorded from 23 patients with intractable seizures, including 5 males and 18 females from age 2 to age 22. The data of each channel is recorded at 256 Hz with 16-bit resolution. The beginning and end of each seizure are both manually annotated by clinical experts after visual inspection. From the data, most patients contain 23 bipolar EEG signals derived from electrodes placed according to the International 10-20 position system. In our experiments, only the recordings that contain the same 23 EEG channels (FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ, CZ-PZ, P7-T7, T7-FT9, FT9-FT10, FT10-T8, and T8-P8) are utilized.

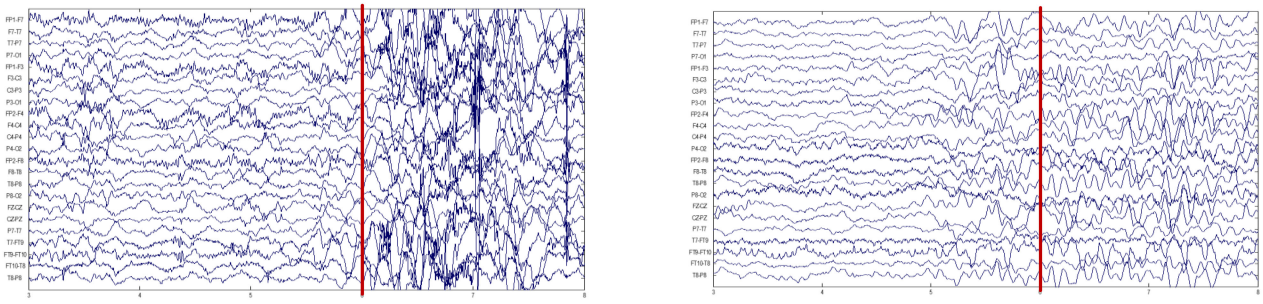
Fig. 3 illustrates two examples of EEG seizures within two different patients on the CHB-MIT dataset. From the figures, we can observe that the EEG seizure reflects distinctive rhythmic patterns in different channels and varies significantly across patients. The seizure in EEG signals comes with significant frequency changes from patient A shown in Fig. 3a, while with relative weak changes from patient B shown in Fig. 3b. Thus, the variability of rhythmic patterns makes the EEG seizure detection problem even more difficult.

B. Baseline Methods

To validate the performance of our proposed framework in EEG seizure detection, we compare it with nine popular existing algorithms as baselines, including both traditional handcrafted feature extraction and conventional deep learning methods.

We employ some widely used handcrafted feature extraction baselines combined with classifiers, including standard SVM and NN. In order to reduce the number of feature dimensions to the same as ours, for the sake of fairness, we first adopt PCA to select top-u components as temporal-domain features from EEG signals, and then feed them to train classifiers, named PSVM and PNN. Assuming that the EEG signal in a short interval is relatively stationary, we utilize fast Fourier transform (FFT) as spectral-domain features to train a SVM, namely FFT-PSVM. We also adopt these methods in the STFT domain, referred to as STFT-PSVM and STFT-NN. Note that PCA is still combined with SVM to select distinctive features from multiple channels and reduce the curse of dimensionality.

Since our framework is based on deep learning, we choose standard CNN [83], SSDA [78], and ConvA [79], named STFT-CNN, STFT-sSSDA and STFT-sConvA, respectively, to compare detection performance via EEG spectrogram representation. Different from STFT-CNN, both STFT-sSSDA and STFT-sConvA integrate reconstruction and detection errors during training. Moreover, WT-CtxFusionEEG [50], a wavelet-based multi-stage context learning method, is also conducted for our comparisons.



(a) Scalp EEG signals of patient A

(b) Scalp EEG signals of patient B

Fig. 3. Two samples of multi-channel scalp EEG signals on the CHB-MIT dataset. The red bar marks the beginning of a EEG seizure, and both patients A and B start seizure at the 6-th second. The EEG seizure reflects distinctive rhythmic patterns in different channels and varies significantly across patients.

C. Evaluation Criteria

Since the evaluation task belongs to a classification problem, we use F1-score and Accuracy to evaluate our model. Moreover, the receiver operator characteristic (ROC) curve and the precision-recall (PR) curve are plotted to illustrate the diagnostic ability of the seizure detector. In the experiments, we also employ the area-under-the-curve (AUC) of both curves to numerically validate the quality of each method, referred to as AUC-ROC and AUC-PR.

D. Experiment Setup

To evaluate our method as a general seizure detecting algorithm, we combine EEG fragments from all the patients. In order to enlarge the number of data samples, we employ data augmentation by adding a 1-second overlap during segmentation based on the experience gained in previous research [45], [50], [84]. Finally, the total number of spectrogram we generated is 252,862. We label the class as ictal and non-ictal states for each EEG fragment according to the ground truth. Due to the scarcity of ictal events, the generated experimental dataset is imbalanced where the ratio of ictal and non-ictal segments for each patient is around $6:1$. Considering the computational expense, we perform 5-fold subject-independent cross validation and report the average testing performance with standard deviation ($\mu \pm \sigma$) for each method. Note that, the models are never trained on data from the testing subjects, and we adopt the same subject combination for all the models in each fold in order to fairly compare the performance. As the original data contain 23 channels with 256 Hz sampling rate, a 3-second long EEG fragment consists of 17,664 raw data points.

Furthermore, we implement our models with PyTorch [85]. The training process is done locally using NVIDIA Titan Xp GPU. During the training phase, we minimize the cost function by utilizing the Adadelta optimization algorithm [86]. Some training strategies including batch normalization, momentum ($\rho = 0.95$), weight decay ($L2$ penalty with the coefficient 0.001), and Dropout (the dropout rate is 0.5) are also adopted for all the approaches. Since the data is huge due to channels, we adopt a 2-layer SSDA for both the cross-channel and intra-channel autoencoders. We set 256 as the hidden size of the first layer and 128 for the second layer, i.e., $p = 128$. The ConvA structure of the proposed model for both autoencoders is shown in Table I. We set $u = 128$ as the feature dimension for all the approaches.

TABLE I
CONVA STRUCTURE OF BOTH THE CROSS-CHANNEL AND INTRA-CHANNEL AUTOENCODERS IN THE PROPOSED MODEL

Encoder No.	Conv	Non-linear	Pooling
1	$3 \times 3 \times 16$	ReLU	2×2
2	$3 \times 3 \times 8$	ReLU	2×2
Decoder No.	Deconv	Non-linear	Unpooling
1	$3 \times 3 \times 8$	ReLU	2×2
2	$3 \times 3 \times 16$	ReLU	2×2

E. Seizure Detection Performance

In this subsection, we compare the EEG seizure detection performance of our proposed multi-view deep learning framework with the aforementioned baseline methods. Both our previous model STFT-mSSDA [22] and the new version STFT-mConvA are included for evaluation. We also implement the reduced models for comparisons, which are purely based on the intra-channel features, referred to as STFT-cSSDA and STFT-cConvA. The experimental results on the benchmark dataset are listed in Table II. We can easily observe that feature representation is key for EEG seizure detection, and our models outperform the baselines on all the four evaluation criteria.

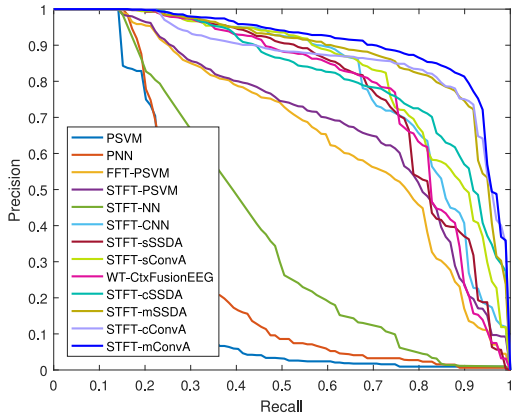
From the results of baselines, the SVM-based models perform better than NN-based models both in the time and time-frequency domain. This is because the raw feature space is unsuitable for neural networks which there is a high probability to reach local minimum using gradient decent optimization algorithm. The performance using the spectral features is better than that using the temporal features. The results of methods in the time-frequency domain also indicate that all the classifiers take advantage of EEG spectrogram representation. The reason is that representing the signal in spectrogram would provide various detailed energy shapes and distributions, and classifiers can hence capture more powerful information. Given the results of deep learning baselines, we can see that both the STFT-sSSDA

TABLE II
DETECTION PERFORMANCE COMPARISONS USING 5-FOLD SUBJECT-INDEPENDENT CROSS VALIDATION

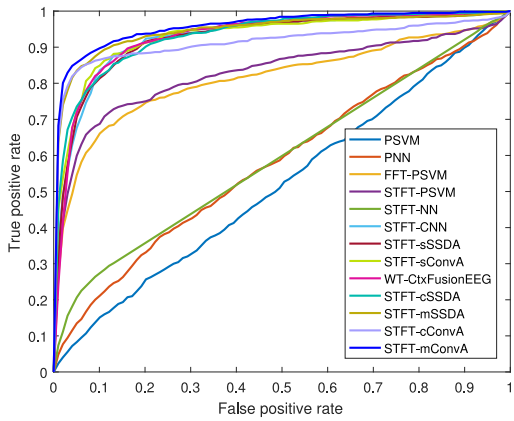
Method	CHB-MIT Dataset			
	AUC-ROC	AUC-PR	F1-score	Accuracy
PSVM	0.5200 ± 0.0527	0.2470 ± 0.0591	0.0077 ± 0.0109	0.7946 ± 0.0336
PNN	0.5789 ± 0.0191	0.3025 ± 0.0606	0.2932 ± 0.0653	0.7268 ± 0.0408
FFT-PSVM	0.8134 ± 0.0467	0.6672 ± 0.0910	0.6049 ± 0.0903	0.8659 ± 0.0226
STFT-PSVM	0.8291 ± 0.0434	0.7021 ± 0.0872	0.6421 ± 0.0758	0.8768 ± 0.0223
STFT-NN	0.5934 ± 0.0377	0.4180 ± 0.1189	0.0668 ± 0.0415	0.7987 ± 0.0309
STFT-CNN	0.9255 ± 0.0295	0.8054 ± 0.0618	0.7506 ± 0.0293	0.8849 ± 0.0268
STFT-sSSDA	0.9244 ± 0.0202	0.7866 ± 0.0873	0.7462 ± 0.0455	0.8895 ± 0.0212
STFT-sConvA	0.9308 ± 0.0132	0.8277 ± 0.0489	0.7560 ± 0.0394	0.8927 ± 0.0189
WT-CtxFusionEEG	0.9287 ± 0.0306	0.7833 ± 0.1147	0.7202 ± 0.1485	0.9025 ± 0.0099
STFT-cSSDA	0.9320 ± 0.0282	0.8269 ± 0.0982	0.7593 ± 0.1105	0.9151 ± 0.0152
STFT-cConvA	0.9140 ± 0.0440	0.8706 ± 0.0206	0.8411 ± 0.0173	0.9381 ± 0.0108
STFT-mSSDA	0.9450 ± 0.0228	0.8801 ± 0.0406	0.8186 ± 0.0801	0.9364 ± 0.0079
STFT-mConvA	**0.9572 ± 0.0182**	**0.9057 ± 0.0367**	**0.8534 ± 0.0335**	**0.9437 ± 0.0119**

and STFT-sConvA methods get higher AUC-ROC and F1-score than STFT-NN and STFT-CNN, respectively. This means that autoencoders can perform auto-associative mapping based on reconstruction errors to handle the imbalanced seizure data. The comparison between them demonstrates that the performance of STFT-sConvA is better than that of STFT-sSSDA. This results from the high-quality features learned by convolutional operators. WT-CtxFusionEEG, which combines deep learning with wavelet features, gets the best average Accuracy of 90.25%, but the worst deviation of 0.1485 in F1-score among baselines. This means that the multi-stage training procedure cannot guarantee consistently good performance across patients. Furthermore, given the results of our methods which consider multi-view representation, the proposed STFT-mConvA model is able to achieve higher average Accuracy and F1-score at 94.37% and 85.34%, respectively, demonstrating a powerful and effective cross-patient method in the task of EEG seizure detection.

Fig. 4 illustrates the mean PR and ROC curves of all the testing folds on the CHB-MIT dataset, respectively. On one hand, our STFT-mConvA model gets the best 0.9572 of AUC-ROC and 0.9057 of AUC-PR in terms of the ability of cross-patient EEG seizure detection. On the other hand, the true positive rate of the STFT-mConvA model increases fast at the beginning, which indicates that STFT-mConvA is able to obtain critical information to separate data effectively and hence results in the best of F1-score and Accuracy. Moreover, compared with our proposed methods, we can see that the intra-channel view provides more representative features than the cross-channel view, and the convolutional-based model is more suitable for seizure detection in EEG spectrograms. We arrive at a conclusion that our multi-view feature learning plays an important role to identify critical EEG patterns, and the combination objective training provides complementary information towards each other, which is crucial for detection performance.



(a) PR curves



(b) ROC curves

Fig. 4. Mean PR and ROC curves of the baselines and proposed methods on the CHB-MIT dataset. The proposed STFT-mConvA model achieves the best values of AUC-PR and AUC-ROC.

F. Channel-Aware Module Analysis

In this subsection, we present the performance improvements of our models when using the proposed channel-aware module. Specifically, we compare with the results using the conventional "full-channel" training strategy, referred to as cSSDA(Full), cConvA(Full), mConvA(Full), and mSSDA(Full), respectively. The comparison results are reported in Fig. 5. From the figure, we can see that the performance of all the models increases

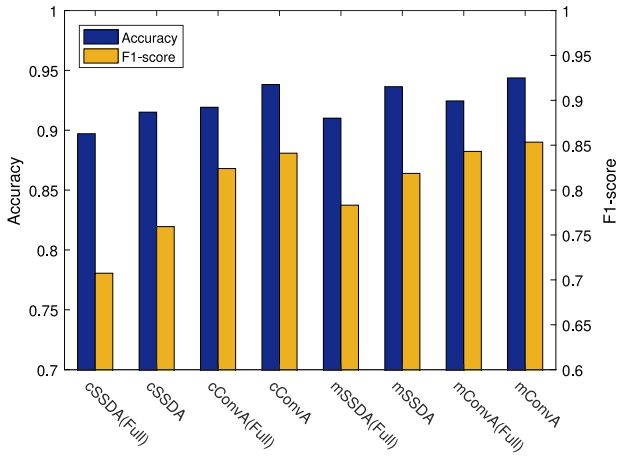


Fig. 5. Performance evaluation of our proposed channel-aware module on the CHB-MIT dataset. The performance of all the models increases on average two percentage points in terms of Accuracy and F1-score when incrementally adopting the channel-aware module.

on average two percentage points in terms of Accuracy and F1-score when adopting the channel-aware module. It means that by adding channel-wise sparsity and selectivity constraints, the models can be trained to focus on important and relevant channels, and thus enhance the feature representation of multi-channel data input. As shown in Table II, the proposed channel-aware module supports the effectiveness of different multi-view deep learning structures, which achieve good performance in EEG seizure detection.

G. Semi-Supervised Seizure Detection Performance

In clinical settings, supervised information is limited for many diagnosis applications, since it requires expensive expert efforts in labeling patient records [84]. Thus, semi-supervised learning schema is in great demand to be designed. Based on the objective function of our proposed model in Eq. (7), it is easy to augment unlabeled training data in a semi-supervised learning manner. To evaluate the performance of our proposed models with a semi-supervised learning setting, we treat half of the training samples as unlabeled fragments during the 5-fold subject-independent cross validation. We then conduct additional experiments on our STFT-mSSDA and STFT-mConvA models, denoted as STFT-mSSDA$_{semi}$ and STFT-mConvA$_{semi}$, respectively. We choose the standard SSDA and ConvA as baselines, denoted as STFT-sSSDA$_{semi}$ and STFT-sConvA$_{semi}$, respectively, since both of them are strong and robust in a semi-supervised manner.

We summarize the semi-supervised detection performance in Table III. We can observe that our models consistently beat the baselines despite the influence of unlabeled information. Specifically, STFT-mConvA$_{semi}$ achieves the best of 93.97% on Accuracy and 84.58% on F1-score, compared with 87.70% and 65.86% for the ConvA baseline. The results of this experiment justify that the proposed model not only makes use of labeled data, but also takes advantage of the multi-channel information contained in the unlabeled fragments. Moreover, compared with the results within the supervised learning setting, our STFT-mConvA model can still achieve comparable results, which shows the strong power of multi-view feature extraction.

H. Sensitivity Analysis

Having demonstrated the effectiveness of our proposed framework compared to other methods, we conduct sensitivity analysis to study the impact of various hyper-parameter settings, such as the dimensionality of multi-view representation p and fusional representation u, channel-aware rate α, and the length of time window l. The F1-score and Accuracy results with different hyper-parameter settings are plotted, as shown in Fig. 6. The default configuration is mentioned in Section IV-D, and we vary one hyper-parameter while keeping others fixed to the default setting.

Both the dimensions of multi-view representation p and fusional representation u are the main factors of the inner structure of our deep learning model. We report the influence of p and u in Fig. 6a and Fig. 6b, respectively. From the figures, the common trend is that when we increase the number of hidden neurons, the accuracy and F1-score continue to rise, but eventually drop off. We set 128 as the default value since it is enough to capture the hidden features while offers the appropriate trade-off in complexity and performance; only relatively minor gains are achieved in increasing the number of neurons beyond 128. We can also see that STFT-mSSDA is more sensitive to the structure variations than STFT-mConvA, which indicates that too many fully-connected layers in mSSDA would lead to be relatively unstable to extract representative features from the EEG spectrogram.

Regarding the channel-aware rate α shown in Fig. 6c, we can observe that the performance degrades when we make α larger, which is expected since larger α implies lesser channel-wise competition. In practice, when tuning α, we find that starting by a value close to around a half of the number of channels is a good strategy. When $\alpha = 1$, which means there is no channel-aware constraints during training, the performance is the worst. This once again reflects the effectiveness of our proposed channel-aware module.

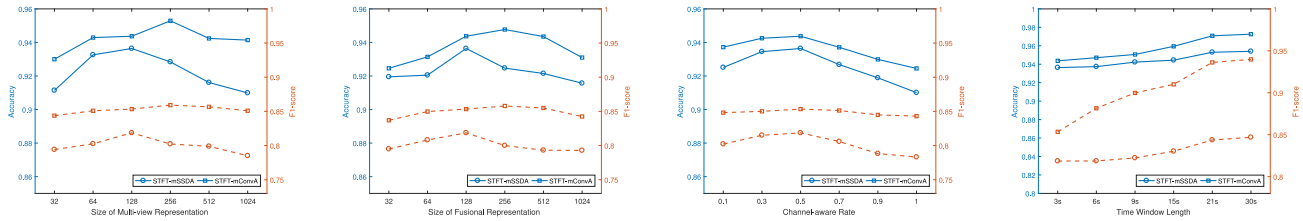
Finally, Fig. 6d illustrates the results for different performance metrics when the window length l takes the values from 3 seconds to 30 seconds. From the figure, as l increases, the performance of both models increases, because more sufficient data are included in the spectrogram. Furthermore, we can see that the increase of STFT-mSSDA is not as significant as that of STFT-mConvA. The reason is still related to the different nature of two autoencoders, where ConvA utilizing locally-connected neural networks is able to preserve more spatial contents from the spectrogram than fully-connected SSDA.

I. Domain Adaptation Analysis

To further validate the proposed framework, we analyze the adaptation in different domain choices, including the time domain and several widely used time-frequency domains with different STFT window functions. Fig. 7 illustrates the performance change of the STFT-mSSDA and STFT-mConvA models

TABLE III
SEMI-SUPERVISED DETECTION PERFORMANCE COMPARISONS USING 5-FOLD SUBJECT-INDEPENDENT CROSS VALIDATION

Method	CHB-MIT Dataset			
	AUC-ROC	AUC-PR	F1-score	Accuracy
STFT-sSSDA$_{semi}$	0.9184 ± 0.0356	0.7738 ± 0.1411	0.6151 ± 0.1688	0.8556 ± 0.0603
STFT-sConvA$_{semi}$	0.8679 ± 0.0318	0.7633 ± 0.0858	0.6586 ± 0.1229	0.8770 ± 0.0276
STFT-mSSDA$_{semi}$	0.9199 ± 0.0404	0.8445 ± 0.0744	0.7666 ± 0.0908	0.9115 ± 0.0143
STFT-mConvA$_{semi}$	**0.9264 ± 0.0465**	**0.8813 ± 0.0215**	**0.8458 ± 0.0123**	**0.9397 ± 0.0114**



(a) Performance variations under different sizes of multi-view representation p

(b) Performance variations under different sizes of fusional representation u

(c) Performance variations under different channel-aware rates α

(d) Performance variations under different time window lengths l

Fig. 6. Performance variations under different hyper-parameter settings. Values represent mean Accuracy and F1-score across 5 testing folds of the CHB-MIT dataset.

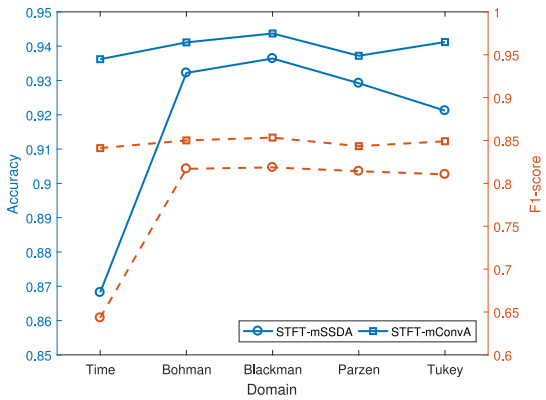


Fig. 7. Performance of model adaptation tested in different domains. Values represent mean Accuracy and F1-score across 5 testing folds of the CHB-MIT dataset.

in different domains based on the same hyper-parameter settings. Note that the first item in Fig. 7 is the result tested directly in time domain (i.e., without spectrogram representation). From the figure, we can see that both of our proposed STFT-mSSDA and STFT-mConvA models perform stable under different STFT window functions. The comparison among window functions demonstrates that the Blackman window is more suitable for EEG seizure detection. This is because Blackman has a lower speak side lobe while containing a relatively narrow transition bandwidth when compared with the others [87], [88]. It means that Blackman is able to preserve more detailed information of amplitude changes, which is one of the main characteristics in EEG seizure signals. Moreover, our STFT-mConvA model still achieves comparable results in the time domain, whereas STFT-mSSDA does not. We conjecture that ConvA has a similar convolution procedure as STFT, which is the basic operator in Fourier transform. Taking advantage of this, STFT-mConvA can still learn representative features directly from the time domain and outperform the baselines.

V. CONCLUSION

In this paper, we propose and evaluate our unified multi-view deep learning framework for automatic EEG seizure detection on a clinical scalp multi-channel EEG epilepsy dataset. The proposed method is an end-to-end framework which is capable of learning multi-view hidden representations by incorporating both inter and intra correlations of EEG channels. We transform EEG fragments into the time-frequency domain via STFT, and obtain the EEG spectrogram representation. The generated 2-D spectrogram fragments are suitable for latent feature extraction using deep learning. Two autoencoder mechanisms, named SSDA and ConvA, are employed to construct multi-view architecture to unleash the power of multi-channel information. By adding sparsity and selectivity constraints, we propose a channel-aware seizure detection module to guide our multi-view model to focus on important and relevant EEG channels. To train the cross-patient seizure detector, we jointly learn multi-view deep features from both unsupervised multi-channel EEG reconstruction and supervised seizure detection. Due to the flexible integrated objective function, it is easy for our model to augment more unlabeled auxiliary data in a semi-supervised manner. Experimental results on the CHB-MIT scalp EEG dataset justify the effectiveness of our proposed deep learning framework in the task of multi-channel EEG seizure detection. As the proposed multi-view deep learning framework is an end-to-end task-oriented model, it is applicable to other medical tasks with similar data structures, especially in health sensing where channel awareness is still a major challenge.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers, and NVIDIA Corporation for the donation of the Titan Xp GPU.

REFERENCES

[1] W. H. Organization, "Epilepsy fact sheet," 2017. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs999/en/
[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] S. Tong and N. V. Thakor, *Quantitative EEG Analysis Methods and Clinical Applications*. Norwood, MA, USA: Artech House, 2009.
[4] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure prediction: The long and winding road," *Brain*, vol. 130, no. 2, pp. 314–333, 2006.
[5] 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.
[6] T. N. Alotaiby, S. A. Alshebeili, T. Alshawi, I. Ahmad, and F. E. A. El-Samie, "EEG seizure detection and prediction algorithms: A survey," *EURASIP J. Adv. Signal Process.*, vol. 2014, no. 1, 2014, Art. no. 183.
[7] L. J. Greenfield, J. D. Geyer, and P. R. Carney, *Reading EEGs: A Practical Approach*. Philadelphia, PA, USA: Lippincott Williams & Wilkins, 2012.
[8] A. T. Tzallas *et al.*, "Automated epileptic seizure detection methods: A review study," in *Epilepsy-Histological, Electroencephalographic and Psychological Aspects*. London, U.K.: InTech, 2012.
[9] M. Ahmad, M. Saeed, S. Saleem, and A. M. Kamboh, "Seizure detection using EEG: A survey of different techniques," in *Proc. IEEE Int. Conf. Emerg. Technol.*, 2016, pp. 1–6.
[10] P. Bloomfield, *Fourier Analysis of Time Series: An Introduction*. New York, NY, USA: Wiley, 2004.
[11] P. E. McSharry, L. A. Smith, and L. Tarassenko, "Prediction of epileptic seizures: Are nonlinear methods relevant?" *Nature Med.*, vol. 9, no. 3, pp. 241–242, 2003.
[12] K. Samiee, P. Kovacs, and M. Gabbouj, "Epileptic seizure classification of EEG time-series using rational discrete short-time fourier transform," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 541–552, Feb. 2015.
[13] I. Jolliffe, "Principal component analysis," in *International Encyclopedia of Statistical Science*. New York, NY, USA: Springer, 2011, pp. 1094–1096.
[14] U. R. Acharya, S. V. Sree, A. P. C. Alvin, and J. S. Suri, "Use of principal component analysis for automatic classification of epileptic EEG activities in wavelet framework," *Expert Syst. Appl.*, vol. 39, no. 10, pp. 9072–9078, 2012.
[15] S. Xie and S. Krishnan, "Signal decomposition by multi-scale PCA and its applications to long-term EEG signal classification," in *Proc. IEEE/ICME Int. Conf. Complex Med. Eng.*, 2011, pp. 532–537.
[16] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.
[17] Z. Mei, X. Zhao, H. Chen, and W. Chen, "Bio-signal complexity analysis in epileptic seizure monitoring: A topic review," *Sensors*, vol. 18, no. 6, 2018, Art. no. 1720.
[18] A. Supratak, C. Wu, H. Dong, K. Sun, and Y. Guo, "Survey on feature extraction and applications of biosignals," in *Machine Learning for Health Informatics*. New York, NY, USA: Springer, 2016, pp. 161–182.
[19] A. Temko, G. Lightbody, E. M. Thomas, G. B. Boylan, and W. Marnane, "Instantaneous measure of EEG channel importance for improved patient-adaptive neonatal seizure detection," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 3, pp. 717–727, Mar. 2012.
[20] E. Keogh and A. Mueen, "Curse of dimensionality," in *Encyclopedia of Machine Learning*. New York, NY, USA: Springer, 2011, pp. 257–258.
[21] M. Längkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognit. Lett.*, vol. 42, pp. 11–24, 2014.
[22] Y. Yuan, G. Xun, K. Jia, and A. Zhang, "A multi-view deep learning method for epileptic seizure detection using short-time fourier transform," in *Proc. 8th ACM Int. Conf. Bioinform., Comput. Biol., Health Inform.*, 2017, pp. 213–222.
[23] K. M. Tsiouris, A. T. Tzallas, S. Markoula, D. Koutsouris, S. Konitsiotis, and D. I. Fotiadis, "A review of automated methodologies for the detection of epileptic episodes using long-term EEG signals," in *Handbook of Research on Trends in the Diagnosis and Treatment of Chronic Conditions*. Hershey, PA, USA: IGI Global, 2016, pp. 231–261.
[24] L. Wang *et al.*, "Automatic epileptic seizure detection in EEG signals using multi-domain feature extraction and nonlinear analysis," *Entropy*, vol. 19, no. 6, 2017, Art. no. 222.
[25] A. H. Shoeb and J. V. Guttag, "Application of machine learning to epileptic seizure detection," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 975–982.
[26] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[27] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biophys.*, vol. 5, no. 4, pp. 115–133, 1943.
[28] J.-L. Song, W. Hu, and R. Zhang, "Automated detection of epileptic EEGs using a novel fusion feature and extreme learning machine," *Neurocomputing*, vol. 175, pp. 383–391, 2016.
[29] R. Sharma and R. B. Pachori, "Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions," *Expert Syst. Appl.*, vol. 42, no. 3, pp. 1106–1117, 2015.
[30] U. R. Acharya, H. Fujita, V. K. Sudarshan, S. Bhat, and J. E. Koh, "Application of entropies for automated diagnosis of epilepsy using EEG signals: A review," *Knowl.-Based Syst.*, vol. 88, pp. 85–96, 2015.
[31] A. Dalton *et al.*, "Development of a body sensor network to detect motor patterns of epileptic seizures," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 11, pp. 3204–3211, Nov. 2012.
[32] 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.
[33] U. R. Acharya, S. V. Sree, G. Swapna, R. J. Martis, and J. S. Suri, "Automated EEG analysis of epilepsy: A review," *Knowl.-Based Syst.*, vol. 45, pp. 147–165, 2013.
[34] 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, 2015.
[35] A. Şengür, Y. Guo, and Y. Akbulut, "Time–frequency texture descriptors of EEG signals for efficient detection of epileptic seizure," *Brain Inform.*, vol. 3, no. 2, pp. 101–108, 2016.
[36] P. Kovacs, K. Samiee, and M. Gabbouj, "On application of rational discrete short time fourier transform in epileptic seizure classification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2014, pp. 5839–5843.
[37] K. Fu, J. Qu, Y. Chai, and Y. Dong, "Classification of seizure based on the time-frequency image of EEG signals using HHT and SVM," *Biomed. Signal Process. Control*, vol. 13, pp. 15–22, 2014.
[38] N. D. Truong, L. Kuhlmann, M. R. Bonyadi, J. Yang, A. Faulks, and O. Kavehei, "Supervised learning in automatic channel selection for epileptic seizure detection," *Expert Syst. Appl.*, vol. 86, pp. 199–207, 2017.
[39] Y.-H. Shih *et al.*, "Hardware-efficient EVD processor architecture in FastICA for epileptic seizure detection," in *Proc.Asia Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, 2012, pp. 1–4.
[40] B. Erem, D. E. Hyde, J. M. Peters, F. H. Duffy, D. H. Brooks, and S. K. Warfield, "Combined delay and graph embedding of epileptic discharges in EEG reveals complex and recurrent nonlinear dynamics," in *Proc. IEEE 12th Int. Symp. Biomed. Imag.*, 2015, pp. 347–350.
[41] J. L. Illingworth, P. Watson, and H. Ring, "Why do seizures occur when they do? Situations perceived to be associated with increased or decreased seizure likelihood in people with epilepsy and intellectual disability," *Epilepsy Behav.*, vol. 39, pp. 78–84, 2014.
[42] Q. Lin *et al.*, "Classification of epileptic EEG signals with stacked sparse autoencoder based on deep learning," in *Proc. Int. Conf. Intell. Comput*. Springer, 2016, pp. 802–810.
[43] A. Supratak, L. Li, and Y. Guo, "Feature extraction with stacked autoencoders for epileptic seizure detection," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2014, pp. 4184–4187.
[44] Y. Qi, Y. Wang, J. Zhang, J. Zhu, and X. Zheng, "Robust deep network with maximum correntropy criterion for seizure detection," *BioMed Res. Int.*, vol. 2014, 2014, Art. no. 703816.
[45] G. Xun, X. Jia, and A. Zhang, "Detecting epileptic seizures with electroencephalogram via a context-learning model," *BMC Med. Inform. Decis. Making*, vol. 16, no. 2, 2016, Art. no. 70.
[46] A. Gogna, A. Majumdar, and R. Ward, "Semi-supervised stacked label consistent autoencoder for reconstruction and analysis of biomedical signals," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2196–2205, Sep. 2017.
[47] A. R. Johansen, J. Jin, T. Maszczyk, J. Dauwels, S. S. Cash, and M. B. Westover, "Epileptiform spike detection via convolutional neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2016, pp. 754–758.
[48] A. Antoniades, L. Spyrou, C. C. Took, and S. Sanei, "Deep learning for epileptic intracranial EEG data," in *Proc. IEEE 26th Int. Workshop Mach. Learn. Signal Process.*, 2016, pp. 1–6.

[49] P. Thodoroff, J. Pineau, and A. Lim, "Learning robust features using deep learning for automatic seizure detection," in *Proc. Mach. Learn. Healthcare Conf.*, 2016, pp. 178–190.
[50] Y. Yuan, G. Xun, K. Jia, and A. Zhang, "A novel wavelet-based model for EEG epileptic seizure detection using multi-context learning," in *Proc. IEEE Int. Conf. Bioinformatics Biomed.*, 2017, pp. 694–699.
[51] J. Zhao, X. Xie, X. Xu, and S. Sun, "Multi-view learning overview: Recent progress and new challenges," *Inform. Fusion*, vol. 38, pp. 43–54, 2017.
[52] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3/4, pp. 321–377, 1936.
[53] F. R. Bach and M. I. Jordan, "Kernel independent component analysis," *J. Mach. Learn. Res.*, vol. 3, pp. 1–48, 2002.
[54] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural Comput.*, vol. 16, no. 12, pp. 2639–2664, 2004.
[55] D. M. Blei and M. I. Jordan, "Modeling annotated data," in *Proc. 26th Annu. Int. ACM SIGIR Conf. Develop. Inform. Retrieval.*, 2003, pp. 127–134.
[56] Y. Jia, M. Salzmann, and T. Darrell, "Factorized latent spaces with structured sparsity," in *Proc. Adv. Neural Inform. Process. Syst.*, 2010, pp. 982–990.
[57] W. Liu, D. Tao, J. Cheng, and Y. Tang, "Multiview hessian discriminative sparse coding for image annotation," *Comput. Vision Image Understanding*, vol. 118, pp. 50–60, 2014.
[58] H. Shen, F. Ma, X. Zhang, L. Zong, X. Liu, and W. Liang, "Discovering social spammers from multiple views," *Neurocomputing*, vol. 225, pp. 49–57, 2017.
[59] H. Xiao, J. Gao, D. S. Turaga, L. H. Vu, and A. Biem, "Temporal multi-view inconsistency detection for network traffic analysis," in *Proc. 24th Int. Conf. World Wide Web*. ACM, 2015, pp. 455–465.
[60] N. Chen, J. Zhu, and E. P. Xing, "Predictive subspace learning for multi-view data: A large margin approach," in *Proc. Adv. Neural Inform. Process. Syst.*, 2010, pp. 361–369.
[61] E. P. Xing, R. Yan, and A. G. Hauptmann, "Mining associated text and images with dual-wing harmoniums," 2012, arXiv:1207.1423.
[62] H. Hu, B. Liu, B. Wang, M. Liu, and X. Wang, "Multimodal DBN for predicting high-quality answers in CQA portals," in *Proc. 51st Annu. Meeting Assoc. Comput. Linguistics (Volume 2: Short Papers)*, vol. 2, 2013, pp. 843–847.
[63] L. Ge, J. Gao, X. Li, and A. Zhang, "Multi-source deep learning for information trustworthiness estimation," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2013, pp. 766–774.
[64] N. Srivastava and R. R. Salakhutdinov, "Multimodal learning with deep boltzmann machines," in *Proc. Adv. Neural Inform. Process. Syst.*, 2012, pp. 2222–2230.
[65] W. Wang, R. Arora, K. Livescu, and J. Bilmes, "On deep multi-view representation learning," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1083–1092.
[66] F. Feng, X. Wang, and R. Li, "Cross-modal retrieval with correspondence autoencoder," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 7–16.
[67] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 689–696.
[68] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view convolutional neural networks for 3D shape recognition," in *Proc. IEEE Int. Conf. Comput. Vision*, 2015, pp. 945–953.
[69] S. Yao, S. Hu, Y. Zhao, A. Zhang, and T. Abdelzaher, "Deepsense: A unified deep learning framework for time-series mobile sensing data processing," in *Proc. 26th Int. Conf. World Wide Web*, 2017, pp. 351–360.
[70] J. Gotman, D. Flanagan, J. Zhang, and B. Rosenblatt, "Automatic seizure detection in the newborn: Methods and initial evaluation," *Electroencephalography Clin. Neurophysiol.*, vol. 103, no. 3, pp. 356–362, 1997.
[71] D. Ravi, C. Wong, B. Lo, and G.-Z. Yang, "Deep learning for human activity recognition: A resource efficient implementation on low-power devices," in *Proc. IEEE 13th Int. Conf. Wearable Implantable Body Sensor Netw.*, 2016, pp. 71–76.
[72] L. Cohen, "Time-frequency distributions-a review," *Proc. IEEE*, vol. 77, no. 7, pp. 941–981, Jul. 1989.
[73] A. V. Oppenheim, *Discrete-Time Signal Processing*. Noida, India: Pearson Education India, 1999.
[74] E. Niedermeyer and F. L. da Silva, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Philadelphia, PA, USA: Lippincott Williams & Wilkins, 2005.
[75] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," 2017, arXiv:1710.05381.
[76] H.-J. Lee and S. Cho, "The novelty detection approach for different degrees of class imbalance," in *Proc. Int. Conf. Neural Inform. Process.*, Springer, 2006, pp. 21–30.
[77] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Mach. Learn.*, ACM, 2008, pp. 1096–1103.
[78] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Proc. Adv. Neural Inform. Process. Syst.*, 2012, pp. 341–349.
[79] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *Proc. Int. Conf. Artif. Neural Netw.*, Springer, 2011, pp. 52–59.
[80] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[81] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, 2009.
[82] 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.
[83] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[84] Y. Yuan, G. Xun, Q. Suo, K. Jia, and A. Zhang, "Wave2vec: Learning deep representations for biosignals," in *Proc. IEEE Int. Conf. Data Mining*, 2017, pp. 1159–1164.
[85] A. Paszke, S. Gross, and S. Chintala, "Pytorch," 2017. [Online]. Available: https://github.com/pytorch/pytorch
[86] M. D. Zeiler, "Adadelta: An adaptive learning rate method," 2012, arXiv:1212.5701.
[87] A. Kumar, "A comparative study of performance of Blackman window family for designing cosine-modulated filter bank," in *Proc. Int. Conf. Circuits, Syst. Simul.*, 2011.
[88] A. Datar, A. Jain, and P. Sharma, "Performance of Blackman window family in M-channel cosine modulated filter bank for ECG signals," in *Proc. Int. Multimedia, Signal Process. Commun. Technol.*, 2009, pp. 98–101.