2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)
4-7 March 2018
Las Vegas, Nevada, USA

A Novel Channel-aware Attention Framework for Multi-channel EEG Seizure Detection via Multi-view Deep Learning

Ye Yuan[1], Guangxu Xun[2], Fenglong Ma[2], Qiuling Suo[2], Hongfei Xue[2], Kebin Jia[1] and Aidong Zhang[2]

***Abstract*— Epileptic seizure detection using multi-channel scalp electroencephalogram (EEG) signals has gained increasing attention in clinical therapy. Recently, researchers attempt to employ deep learning techniques with channel selection to determine critical channels. However, existing models with such hard selection procedure do not take dynamic constraints into account, since the irrelevant channels vary significantly across different situations. To address these issues, we propose ChannelAtt, an end-to-end multi-view deep learning model with channel-aware attention mechanism, to express multi-channel EEG signals in a high-level space with interpretable meanings. ChannelAtt jointly learns both multi-view representation and its contribution scores. We propose two attention mechanisms to learn the attentional representations of multi-channel EEG signals in time-frequency domain. Experimental results show that the proposed ChannelAtt model outperforms the baselines in detecting epileptic seizures. Analytical results of a case study demonstrate that the learned attentional representations are meaningful.**

I. INTRODUCTION

Epilepsy is a serious chronic neurological disorder caused by sudden abnormal neuronal discharges in the brain. The major symptom of epilepsy is associated with recurrent and unpredictable epileptic seizures [1]. According to previous studies, there has been a growing interest in detecting multi-channel scalp electroencephalogram (EEG) epileptic seizure through deep learning methods [2], [3]. These deep learning based methods are proposed to explore inherent EEG representations to observe brain electrical activities. The challenge is that most of the channels in multi-channel EEG signals are irrelevant to brain related activities, including seizure onset [4]. These irrelevant channels introduce lots of noise to data and could significantly reduce the performance of the learning methods. To address this, researchers attempt to employ deep learning techniques with channel selection procedure to determine the critical features by the partition of channels [5], [6]. Yuan et al. [4] proposed an SSDA-based EEG channel selection procedure considering multiple information to jointly determine the critical channel features. These methods result in a marked mitigation of the noise effect, since the irrelevant channels are filtered out. However, they do not always yield good performance with such hard selection strategy, since the irrelevant channels change over different situations.

[1]Y. Yuan and K. Jia are with the Beijing Laboratory of Advanced Information Network, College of Information and Communication Engineering, Beijing University of Technology, Beijing 100124, China. E-mail: yuanye91@emails.bjut.edu.cn, kebinj@bjut.edu.cn.

[2]G. Xun, F. Ma, Q. Suo, H. Xue and A. Zhang are with the Department of Computer Science and Engineering, State University of New York at Buffalo, Buffalo, NY 14260. E-mail: guangxux, fenglong, qiulings, hongfeix, azhang@buffalo.edu.

To tackle the aforementioned issues, we propose an innovative channel-aware attention framework (ChannelAtt) to achieve dynamic soft channel selection in multi-channel EEG signals, and employ the learned low-dimensional features in seizure detection. Our framework is described in Fig. 1. Specifically, as the biosignals expressed in time-frequency domain are more meaningful than time domain [7]. In our model, we first slice the signals into pieces with a sliding window and adopt short-time Fourier transform (STFT) [8] as preprocessing to describe the frequency content in signals over time, denoted as spectrogram. We then train an unified model, which consists of a multi-view representation layer with two encoders and a channel-aware attention layer, to further express multi-channel signals in a high-level space with interpretable meanings. Note that the described framework also caters for additional multi-channel time-series tasks from other sensor types, such as human activity recognition and arrhythmia detection. In summary, our contributions are as follows:

- We propose ChannelAtt, an end-to-end multi-view deep learning model to accurately detect seizure onset from multi-channel EEG signals, without depending on any expert medical knowledge.
- We develop two channel-aware attention mechanisms to dynamically calculate scores for each channel and achieve soft channel selection. To the best of our knowledge, this is the first work using attention mechanism for biosignal channel selection in healthcare.
- We empirically show that the proposed ChannelAtt outperforms existing seizure detection methods on a benchmark dataset. The results of a case study indicate that the proposed channel-aware attention mechanism is able to identify the influential clinical concepts of seizure onset.

II. METHODOLOGY

In this section, we show the details of our proposed channel-aware attention (ChannelAtt) methodology that contains two models, namely ChannelAtt$_{loc}$ and ChannelAtt$_{glo}$, respectively.

A. Multi-view Representation

In the task of multi-channel data mining, simply concatenating the raw input features may not be enough for deep learning models to yield robust and accurate results. The importance of extracting features from different perspectives,

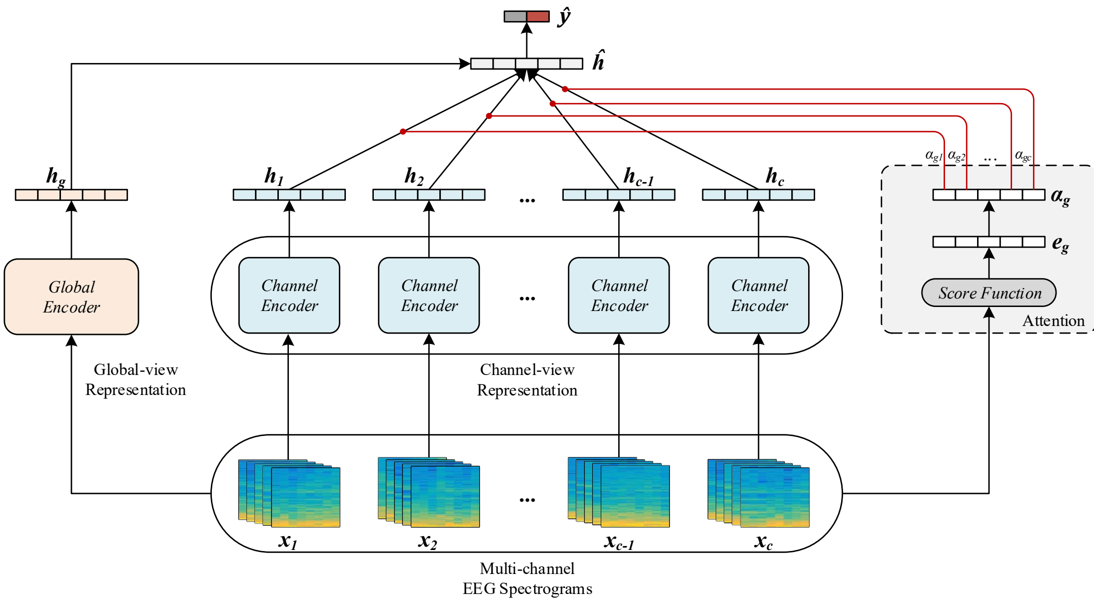

Fig. 1. Schematic illustration of the overall approach pipeline.

denoted as multi-view deep learning, has proven to be effective in bioinformatics [4]. Inspired by this work, we further extend this strategy by adopting two feature encoders to learn latent representations from global and channel-specific views, i.e., global-encoder and channel-encoder, respectively. Specifically, given a C-channel EEG spectrogram segment $\boldsymbol{x} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_C)$ where $\boldsymbol{x}_i \in \mathbb{R}^n$, the global-view representation $\boldsymbol{h}_g \in \mathbb{R}^p$ can be expressed by a global-encoder based on forward-propagation:

$$\boldsymbol{h}_g = f(\boldsymbol{W}_g x_{1:C} + \boldsymbol{b}_g), \tag{1}$$

where $f(\cdot)$ is the activation function, $\boldsymbol{W}_g \in \mathbb{R}^{p \times Cn}$ and $\boldsymbol{b}_g \in \mathbb{R}^p$ denote the weight matrix and bias vector, respectively. For each $\boldsymbol{x}_i$, we can obtain its channel-view representation $\boldsymbol{h}_i \in \mathbb{R}^p$ by feeding a channel-encoder as follows:

$$\boldsymbol{h}_i = f(\boldsymbol{W}_c x_i + \boldsymbol{b}_c), \tag{2}$$

where $W_c \in \mathbb{R}^{p \times n}$ and $b_c \in \mathbb{R}^p$ are the weight matrix and bias vector, respectively. Note that the encoder could be parameterized by any multi-layer deep learning models. To demonstrate the basic idea of the proposed method, both global and channel encoders are parameterized by two-layer stacked autoencoders (SAEs) [9].

B. Channel-aware Attention Mechanism

In health domain, interpreting the learned representations is important to understand the clinical meaning [10]. Since the input of our model is multi-channel data, we focus on discovering the contribution scores of channels and interpreting which ones are crucial to clinical tasks. To achieve this, we propose channel-aware attention mechanism. The main idea is to derive a global context vector $\boldsymbol{c}_g \in \mathbb{R}^p$ that captures relevant channel information to help detect the seizure onset. Specifically, we build a score function to calculate the energy e_{gi} of each channel i separately and obtain a normalized attention score vector $\boldsymbol{\alpha}_g$ using softmax function:

$$\boldsymbol{\alpha}_g = softmax([e_{g1}, e_{g2}, ..., e_{gC}]). \tag{3}$$

Then the context vector can be calculated based on the scores obtained from Eq.(3) and the channel-view representation $\boldsymbol{h}_i (1 \le i \le C)$ as follows:

$$\boldsymbol{c}_g = \sum_{i=1}^{C} \alpha_{gi} \boldsymbol{h}_i. \tag{4}$$

In our framework, based on the learned multi-view representation, we employ two mechanisms to learn attention scores.

Local-based Attention: An easy way to calculate the energy is scoring solely from each channel-view representation:

$$e_{gi} = \boldsymbol{W}_e^\top \boldsymbol{h}_i + b_e, \tag{5}$$

where $\boldsymbol{W}_e \in \mathbb{R}^p$ and $b_e \in \mathbb{R}$ are the weight vector and bias value, respectively. Note that the local-based attention mechanism does not capture the relationships among channels, since it only considers individual channel information. Moreover, it is time-consuming to directly measure the correlation of any two channels. Thus, to utilize the information from all channels, we adopt a global-based attention mechanism in the proposed ChannelAtt.

Global-based Attention: We use a multi-layer perceptron (MLP) [11] to calculate the energy by considering the relationships between global-view and channel-view representations:

$$e_{gi} = \boldsymbol{v}_e^\top f(\boldsymbol{W}_e[\boldsymbol{h}_g; \boldsymbol{h}_i]), \tag{6}$$

where $\boldsymbol{W}_e \in \mathbb{R}^{q \times 2p}$ and $\boldsymbol{v}_e \in \mathbb{R}^q$ is the parameter to be learned, and we can obtain the context vector $\boldsymbol{c}_g$ with Eq.(3) and Eq.(4).

C. Seizure Detection

Given the context vector, we concatenate it with the global-view vector to generate an attentional hidden representation defined as:

$$\hat{\boldsymbol{h}} = f(\boldsymbol{W}_h[\boldsymbol{c}_g; \boldsymbol{h}_g]), \tag{7}$$

where $\boldsymbol{W}_h \in \mathbb{R}^{r \times 2p}$ and $\boldsymbol{b}_g \in \mathbb{R}^r$ denote the weight matrix and bias vector, respectively. The attentional vector $\hat{\boldsymbol{h}}$ is then fed through the softmax layer for seizure detection:

$$\hat{\boldsymbol{y}} = softmax(\boldsymbol{W}_s \hat{\boldsymbol{h}} + \boldsymbol{b}_s). \tag{8}$$

To train an unified model on m samples, the final cost function of our end-to-end ChannelAtt model measured by the cross-entropy loss is defined as:

$$\begin{aligned} & J_{ChannelAtt}(\boldsymbol{W}_{g,c,e,h,s}, \boldsymbol{b}_{g,c,e,h,s}) \\ = & -\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{|class|}\left[y_j^{(i)}\log \hat{y}_j^{(i)} + (1-y_j^{(i)})\log(1-\hat{y}_j^{(i)})\right]. \end{aligned} \tag{9}$$

III. Experiments

A. Dataset

The EEG dataset we use is the CHB-MIT dataset [12] captured from the PhysioNet [13]. In this dataset, the multi-channel EEG signals are recorded from pediatric subjects with intractable seizures. The data is obtained from 23 patients, including 18 females and 5 males from age 2 to age 22, to characterize their seizures and assess the necessity of surgery for them. The beginning and end of each seizure are both annotated in the ground truth. Specifically, the EEG signals of each patient contains 23 channels (24 or 26 in a few cases), and the data of each channel is recorded at 256 Hz with 16-bit resolution.

B. Experimental Setup

1) Baseline Approaches: We compare our model with several state-of-the-art methods to validate the performance of seizure detection. First, following the handcrafted feature engineering, we use PCA as preprocessing and then adopt SVM, denoted as PSVM. We select top-r related components as features from signals, in order to reduce the number of dimensions. Second, SAEs is one of the most popular deep learning algorithms [9]. For the sake of fairness, we extend this model by employing our multi-view strategies, namely Global-SAEs, Channel-SAEs and Fusion-SAEs, respectively. Finally, STFT-mSSDA [4] is a multi-view deep learning model with hard channel selection. This approach can extract multiple latent features from multi-channel EEG signals and is effective for EEG seizure detection.

2) Evaluation Strategies: Since the evaluation task is a classification problem, we use F1-score and Accuracy to evaluate our model. Moreover, area-under-the-curve of receiver operator characteristic (AUC-ROC) and precision-recall (AUC-PR) are also used to numerically evaluate the quality of each method.

TABLE I
DETECTION PERFORMANCE COMPARISONS

	CHB-MIT Dataset			
Method	AUC-ROC	AUC-PR	F1-score	Accuracy
PSVM	0.7535	0.6356	0.9342	0.8850
Global-SAEs	0.9515	0.8482	0.9317	0.8973
Channel-SAEs	0.9637	0.8972	0.9384	0.9053
Fusion-SAEs	0.9656	0.9054	0.9509	0.9222
STFT-mSSDA [4]	0.9833	0.9545	0.9605	0.9382
ChannelAtt$_{loc}$	0.9643	0.9421	0.9781	0.9651
ChannelAtt$_{glo}$	**0.9847**	**0.9651**	**0.9785**	**0.9661**

3) Implementation Details: To evaluate our method, we combine 4302 EEG fragments from nine different patients as our experiment dataset. Considering the computational expense, we then conduct hold-out validation throughout our experiments as [3]. During the whole training step, we use Adam [14] with mini-batch to minimize the cost function. We also use normalization, regularization, and drop-out strategies for all the approaches. Moreover, since the data is huge due to channels, to fit the input of our task, we set the same $p = [100, 50]$ and $q = r = 50$ for baselines and our approaches.

C. Results of Seizure Detection

The experimental results on the benchmark dataset are listed in Table I. We can easily observe that both our ChannelAtt$_{loc}$ and ChannelAtt$_{glo}$ methods outperform the baseline methods. Our proposed ChannelAtt$_{glo}$ method not only achieves better Accuracy and F1-score, but also obtains higher AUC-ROC and AUC-PR than baselines. Compared with the SAEs-based methods, we can see that the performance of Fusion-SAEs is better than that of Global-SAEs and Channel-SAEs. This results from the high-quality features extracted from EEG signals using multi-view representation. The results of STFT-mSSDA which utilizes hard channel selection reveal that using the most significant features, the performance increases in terms of all four different evaluation criteria. From the results of ChannelAtt-based methods which consider both multi-view representation and channel-aware attention mechanism, we can conclude that the proposed methods can successfully extract meaningful features from multi-channel EEG signals.

Moreover, the comparison between ChannelAtt$_{glo}$ and ChannelAtt$_{loc}$ demonstrate that the global-based attention mechanism is more suitable for seizure detection. We arrive at a conclusion that channel-aware attention plays an important role to identify seizure patterns, and the attentional representation provides complementary information, which is crucial for successful seizure detection.

D. Case Study

To demonstrate the benefits of our proposed channel-aware attention mechanisms in seizure detection, we analyze the attention scores learned from one of the proposed approach ChannelAtt$_{glo}$, which uses global-based mechanism. Table II

TABLE II

MULTI-CHANNEL EEG SEIZURE DETECTION OF A PATIENT IN THE CASE STUDY (VALUE FORMAT: 00.0%, F: NON-ICTAL, T: ICTAL)

		Channel $No.$																						
Time	State	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	F	3.6	0.9	0.6	0.2	9.0	5.4	2.7	1.0	8.6	**11.0**	4.6	2.1	9.7	3.5	2.3	3.2	9.5	7.1	0.9	1.4	7.6	2.9	2.3
3	F	5.9	0.4	0.8	0.3	12.7	**12.8**	0.2	0.1	3.8	12.3	10.6	6.5	1.8	1.8	0.9	3.4	9.4	7.0	1.0	0.2	5.6	1.4	0.9
5	F	2.2	2.0	3.6	3.0	**19.7**	7.2	1.7	0.6	12.3	6.5	1.3	1.6	3.6	5.9	0.9	2.2	7.7	3.1	2.3	1.5	8.1	2.0	0.9
7	T	0.0	0.0	0.0	0.0	11.4	1.2	0.1	0.1	7.4	6.9	1.3	3.8	0.1	0.0	0.0	2.6	**33.6**	31.0	0.0	0.0	0.2	0.0	0.0
9	T	0.0	0.0	0.0	0.0	11.8	0.8	0.2	0.1	3.6	5.0	0.4	1.5	0.1	0.0	0.0	2.2	**38.1**	35.8	0.0	0.0	0.2	0.0	0.0
11	T	0.0	0.0	0.0	0.0	18.5	0.6	0.2	0.2	2.2	8.0	1.0	1.9	0.1	0.0	0.0	2.4	**39.7**	24.4	0.0	0.0	0.4	0.0	0.0
13	F	1.0	0.1	0.3	1.0	11.8	6.4	2.0	1.8	6.7	7.2	10.6	8.5	3.0	0.3	0.5	4.7	**16.5**	15.3	0.3	0.0	1.3	0.0	0.5
15	F	3.4	0.6	0.3	1.5	5.6	4.9	2.1	2.1	9.9	6.6	7.4	**12.9**	5.7	0.9	1.5	7.9	6.8	12.0	0.6	0.1	5.7	0.2	1.5
17	F	5.1	1.0	0.1	0.6	3.4	5.2	3.8	3.6	5.6	9.2	4.4	8.6	5.2	1.7	0.3	11.4	9.1	8.9	0.2	0.7	**11.2**	0.5	0.3

shows a case study for multi-channel EEG seizure detection of a patient on the CHB-MIT dataset. To provide a clear visualization, we calculate the mean values of every 5 segments and interlaced display. We also highlight the max score of all channel attention scores for each row.

From Table II, we can observe that for different timestamps, the attention scores learned by the attention mechanism are different. Specifically, when the patient is in non-ictal state, the distribution of attention scores are relatively uniform. This is because no such ictal pattern is found within whole channel views and hence the attention scores make evenly contribution to the seizure detection. During the seizures, we can see that the attentional representations in the 17th and 18th channels significantly contribute to the final detection. According to the dataset description on the International 10-20 System [13], this seizure onset is located on the central area. The larger attention score means the probability of seizure onset on this area is higher. In summary, the case study indicates that we can learn accurate attention scores with interpretable representations by our ChannelAtt models which not only improve the detection performance, but also identify the influential clinical concepts of seizure onset in healthcare.

IV. CONCLUSION

In this paper, we proposed a novel channel-aware attention framework, named ChannelAtt, to express multi-channel signals in a high-level space with interpretable meanings. ChannelAtt is an unified model that jointly learns both multi-view representation and its contribution scores at the same time. Two attention mechanisms allow us to interpret the contributions of each channel on seizure onset detection. Experimental results indicate the effectiveness of our proposed ChannelAtt model in detecting epileptic seizures. Throughout a case study, we demonstrate that the learned attentional representations are meaningful.

ACKNOWLEDGMENT

This paper is supported by the National Science Foundation of China (No. 61672064, 81370038), the Beijing Science Foundation (No. 4172001), the China Postdoctoral Science Foundation (No. 2016T90022, 2015T80030, 2015M580029), the Science and Technology Project of Beijing Municipal Education Commission (No. KZ201610005007), and the China Scholarship Council Fund (No. 201606540008).

REFERENCES

[1] R. S. Fisher, W. v. E. Boas, W. Blume, C. Elger, P. Genton, P. Lee, and J. Engel, "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.

[2] Y. Yuan, G. Xun, K. Jia, and A. Zhang, "A novel wavelet-based model for eeg epileptic seizure detection using multi-context learning," in *Bioinformatics and Biomedicine (BIBM), 2017 IEEE International Conference on*. IEEE, 2017, pp. 694–699.

[3] G. Xun, X. Jia, and A. Zhang, "Detecting epileptic seizures with electroencephalogram via a context-learning model," *BMC Medical Informatics and Decision Making*, vol. 16, no. 2, p. 70, 2016.

[4] 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 *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. ACM, 2017, pp. 213–222.

[5] K. Li, X. Li, Y. Zhang, and A. Zhang, "Affective state recognition from eeg with deep belief networks," in *Bioinformatics and Biomedicine (BIBM), 2013 IEEE International Conference on*. IEEE, 2013, pp. 305–310.

[6] X. Jia, K. Li, X. Li, and A. Zhang, "A novel semi-supervised deep learning framework for affective state recognition on eeg signals," in *Bioinformatics and Bioengineering (BIBE), 2014 IEEE International Conference on*. IEEE, 2014, pp. 30–37.

[7] Y. Yuan, G. Xun, Q. Suo, K. Jia, and A. Zhang, "Wave2vec: Learning deep representations for biosignals," in *Data Mining (ICDM), 2017 IEEE 16th International Conference on*. IEEE, 2017, pp. 1159–1164.

[8] L. Cohen, "Time-frequency distributions-a review," *Proceedings of the IEEE*, vol. 77, no. 7, pp. 941–981, 1989.

[9] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, no. 5786, pp. 504–507, 2006.

[10] F. Ma, R. Chitta, J. Zhou, Q. You, T. Sun, and J. Gao, "Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2017, pp. 1903–1911.

[11] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[12] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Massachusetts Institute of Technology, 2009.

[13] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[14] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.