CNN-AIDED FACTOR GRAPHS WITH ESTIMATED MUTUAL INFORMATION FEATURES FOR SEIZURE DETECTION

Bahareh Salafian[1], Eyal Fishel Ben-Knaan[2], Nir Shlezinger[2], Sandrine de Ribaupierre[3], Nariman Farsad[1]

[1]Department of Computer Science, Ryerson University,
{bsalafian,nfarsad}@ryerson.ca
[2]School of Electrical and Computer Engineering, Ben-Gurion University of the Negev,
eyalfish@post.bgu.ac.il, nirshl@bgu.ac.il
[3]School of Biomedical Engineering, University of Western Ontario,
sderibau@uwo.ca

ABSTRACT

We propose a convolutional neural network (CNN) aided factor graphs assisted by mutual information features estimated by a neural network for seizure detection. Specifically, we use neural mutual information estimation to evaluate the correlation between different electroencephalogram (EEG) channels as features. We then use a 1D-CNN to extract extra features from the EEG signals and use both features to estimate the probability of a seizure event. Finally, learned factor graphs are employed to capture the temporal correlation in the signal. Both sets of features from the neural mutual estimation and the 1D-CNN are used to learn the factor nodes. We show that the proposed method achieves state-of-the-art performance using 6-fold leave-four-patients-out cross-validation.

1. INTRODUCTION

Epilepsy is a highly common neurological disorder, causing recurrent episodes of the involuntary movement known as epileptic seizures [1]. Based on the place in the brain where seizure starts and the intensity of the abnormal signals, patients with epilepsy may suffer from different symptoms such as auras, repetitive muscle contraction, and loss of consciousness. [2]. Epileptic seizure severely affects the patient's quality of life and can have other social and economic impacts; for instance, some activities, including swimming, bathing, and climbing a ladder, become dangerous as a seizure during that activity might result in unpredictable injuries and even death. Therefore, early detection of epilepsy can notably improve the patient's quality of life. A leading tool to diagnose seizure is based on electroencephalogram (EEG) monitoring, being economical, portable, and non-invasive [3]. However, the review of EEG recordings is a time-consuming expert-dependent process due to contamination by physiological and non-physiological resources [4], and similarity of epileptic spikes to normal EEG waveforms.

The challenges associated with EEG monitoring gave rise to a growing interest in machine learning aided automatic seizure detection. A common approach is to train a model, typically a convolutional neural network (CNN), applied to features extracted from the Wavelet or Fourier transform of the signal [5–10], typically involving careful feature engineering. Other seizure detection methods process the raw EEG signals directly. These include the application of CNNs [11, 12] to the segmented EEGs (e.g., 4-second blocks), providing instantaneous prediction without exploiting temporal correlation between blocks. Prior works have also considered CNN-recurrent neural network (RNN) architectures to capture temporal correlation [13, 14] that lead to high computational complexity during training. The challenges associated with previous works motivate the formulation of a reliable automatic seizure detection algorithm which generalizes to different patients, benefits from both temporal and inter-channel correlation, and is computationally efficient facilitating its application in real-time.

In this work, we propose a data-driven automatic seizure detection system coined Mutual Information-based CNN-Aided Learned factor graphs (MICAL). MICAL combines computationally efficient 1D CNNs with principled methods for benefiting from temporal and inter-channel correlation. Following [15], we exploit the temporal correlation by imposing a Markovian model on the latent seizure activity [16], using the CNN output not as seizure estimates, but as messages conveyed as a form of learned factor graph inference [17–19]. We expand our previous work [15], to exploit the inter-channel correlation during the seizure by estimating the mutual information (MI) between each pair of EEG channels through a neural MI estimator. To the best of our knowledge, this is the first time MI has been used as the feature for seizure detection. The MI features along with features learned by the 1D-CNN are then used for learning the factor nodes for factor graph inference. Our numerical evaluations, which use the CHB-MIT dataset [20], demonstrate how each of the ingredients combined in MICAL contributes to its reliability, allowing it to achieve improved accuracy and generalization performance compared to previous algorithms.

The rest of this paper is organized as follows. In Section 2 we describe the problem statement and review necessary preliminaries. Then, in Section 3 we describe the proposed MICAL algorithm; Section 4 presents a numerical study, while Section 5 provides concluding remarks.

2. PROBLEM STATEMENT AND PRELIMINARIES

2.1. Seizure Detection Problem

In this paper, seizure detection refers to the identification and localization of the ictal (i.e., the seizure) time intervals from EEG recordings of patients with epilepsy [21]. To formulate this mathematically, let $\boldsymbol{X} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_N\}$ be the EEG recordings of a patient, where N represents the number of channels. Each measured channel $\boldsymbol{X}_i$ is comprised of n consecutive blocks, e.g., blocks of 1-second recordings, and we write $\boldsymbol{X}_i = [\boldsymbol{x}_{t_1}^{(i)}, \boldsymbol{x}_{t_2}^{(i)}, \cdots, \boldsymbol{x}_{t_n}^{(i)}]$, where $\boldsymbol{x}_t^{(i)}$ is the signal corresponding to the i-th EEG channel during the t-th

block. The seizure state for each block is represented as a binary vector $\boldsymbol{s} = [s_{t_1}, \ldots s_{t_n}]$, where $s_t \in \{0, 1\}$ models whether or not a seizure occurs in the t-th block. Our goal is to design a system which maps the EEG recordings $\boldsymbol{X}$ into an estimate of $\boldsymbol{s}$, which is equivalent to finding the time indices where seizure occurs.

To model the relationship between the EEG signals $\boldsymbol{X}$ and the seizure states $\boldsymbol{s}$, it is needed to consider both inter-channel correlation as well as temporal correlation that the recordings exhibit. The former stems from the fact that when the seizure starts, the epileptic activity propagates to other areas in the brain [22] which affects the patterns of other channel recordings [23]. This manifests high dependence between different channels, i.e., between $\boldsymbol{x}_t^{(i)}$ and $\boldsymbol{x}_t^{(j)}$, when t is at the beginning and during ictal phase. Temporal correlation results from the fact that seizures typically span multiple recording blocks, and thus the probability of observing a seizure at time instance t depends on the presence of a seizure in the previous block, such that the entries of $\boldsymbol{s}$ can be approximated by a Markovian structure [16]. Our proposed solution, detailed in Section 3, exploits this statistical structure using factor graphs.

2.2. Factor Graph Inference

Factor graphs are a representation of factorizable multi-variable functions, such as probability distributions, as a bipartite graph. These graphical models facilitate inference at reduced complexity via message passing algorithms, such as the sum-product methods [24]. Consider an observed sequence $\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n]$ encapsulating a latent state sequence $\boldsymbol{s} = [s_1, \ldots, s_n]$ whose entries take values in a finite set $\mathcal{S}$, as a form of a hidden Markov model (HMM). In such cases, the joint distribution of $\boldsymbol{y}, \boldsymbol{s}$ obeys

$$P(\boldsymbol{s}, \boldsymbol{y}) = \prod_{k=1}^{n} P(s_k|s_{k-1})P(\boldsymbol{y}_k|s_k), \tag{1}$$

which can be represented as factor graph with variable nodes $\{s_k\}_{k=1}^{n}$ and function nodes $\{f_k\}_{k=1}^{n}$, where $f_k(s_k, s_{k-1}) := P(s_k|s_{k-1})P(\boldsymbol{y}_k|s_k)$.

The factor graph representation allows one to compute the maximum a-posteriori probability (MAP) decision rule with a complexity that only grows linearly with n as opposed to exponentially with n. This is achieved by evaluating the marginal distribution $P(s_k, \boldsymbol{y})$ for each $k \in \{1, \ldots n\}$ via message passing over the factor graph. In this case, the forward messages are recursively updated via

$$\mu_{f_k \to s_k}(s_k) = \sum_{\{s_1, \cdots, s_{i-1}\}} \prod_{j=1}^{k} f_j(s_j, s_{j-1}), \tag{2}$$

and the backward messages via

$$\mu_{f_{k+1} \to s_k}(s_k) = \sum_{\{s_{k+1}, \cdots, s_n\}} \prod_{j=k+1}^{n} f_j(s_j, s_{j-1}). \tag{3}$$

Then, the desired marginal distribution, which is maximized by the MAP rule, is given by

$$P(s_k, \boldsymbol{y}) = \mu_{f_k \to s_k}(s_k) \cdot \mu_{f_{k+1} \to s_k}(s_k). \tag{4}$$

Intuitively (2)-(3) are interpreted as an aggregate of neighboring information. Once all neighbors have communicated (i.e., messages have propagated the entirety of the graph) the product of the forwards and backward messages determines the marginal probability.

3. MICAL SEIZURE DETECTION ALGORITHM

In this section, we present the proposed MICAL algorithm. MICAL is comprised of three main components: neural MI estimator quantifying the instantaneous dependence between different channels at each EEG block to capture the inter-channel correlation (see Subsection 3.1); a 1D CNN which generates a latent representation of the raw EEG block plus a soft estimate of the seizure state using the joint features from the 1D-CNN and the neural MI estimator (see Subsection 3.2); and factor graph inference utilizing the soft estimates as learned function nodes to incorporate temporal correlation (see Subsection 3.3). A high-level illustration of the flow of MICAL is depicted in Fig. 1.

3.1. Neural Mutual Information Estimation

MI is a measure of the statistical dependence between two random variables. While cross-correlation measures linear dependence, MI can capture higher-order statistical dependence [25], and is thus able to capture nonlinear relationship between signals, which is likely to exist between EEG signals during seizure [22, 23]. The MI between the random variables x_1, x_2 taking values in $\mathcal{X} \times \mathcal{X}$ with a joint distribution $P_{X_1 X_2}$ and marginals P_{X_1} and P_{X_2} is defined as

$$I(X_1; X_2) = D_{\mathrm{KL}}(P_{X_1 X_2} \| P_{X_1} P_{X_2}), \tag{5}$$

where D_{KL} is the Kullback-Leibler (KL) divergence.

Using (5) to compute MI as a measure of statistical dependence for EEG samples, with unknwon probability distributions is a challenging task [26]. To address this issue, it was recently shown that neural networks can be trained to estimate MI, based on the Donsker-Varadhan representation

$$D_{KL}(P_{X_1 X_2} \| P_{X_1} P_{X_2}) = \sup_{T : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}} \mathbb{E}_{P_{X_1 X_2}}[T(x_1, x_2)]$$
$$- \log \left(\mathbb{E}_{P_{X_1} P_{X_2}} \left[e^{T(x_1, x_2)} \right] \right), \tag{6}$$

where a neural model with parameters ϕ denoted by T_ϕ is used to represent the function T in (6). To maximize the right hand side of (6) gradient descent can be used to find the maximizing set of parameters ϕ [27]. To overcome the limitations imposed by the estimation variance, [28] proposed the Smoothed MI Lower-bound Estimator (SMILE), which learns to estimate MI by training a neural network to maximize the objective function

$$\hat{I}_\phi(X_1; X_2) = \mathbb{E}_{P_{X_1 X_2}}[T_\phi(x_1, x_2)]$$
$$- \log \mathbb{E}_{P_{X_1} P_{X_2}} \left[\mathrm{clip}(e^{T_\phi(x_1, x_2)}, e^{-\tau}, e^{\tau}) \right], \tag{7}$$

where $\mathrm{clip}(v, l, u) := \max(\min(v, u), l)$ and τ is a hyperparameter. The resulting neural estimator was shown to learn to reliably predict MI under various distributions.

In MICAL, we apply SMILE to estimate $I(\boldsymbol{x}_t^{(i)}; \boldsymbol{x}_t^{(j)})$ at each block t for each channel pair i, j. Since MI is symmetric, i.e., $I(\boldsymbol{x}_t^{(i)}; \boldsymbol{x}_t^{(j)}) = I(\boldsymbol{x}_t^{(j)}; \boldsymbol{x}_t^{(i)})$, we only estimate the MI for $j > i$. We set T_ϕ to be a fully-connected network with two hidden layers and ReLU activations, and train it with $\tau = 0.9$ in the objective (7). The numerical results from neural estimator satisfy the underlying hypothesis of high correlation among recordings during seizure state. This is illustrated in Fig. 2, where it is observed that the trained estimator outputs higher MI values during seizure compared to non-seizure blocks.
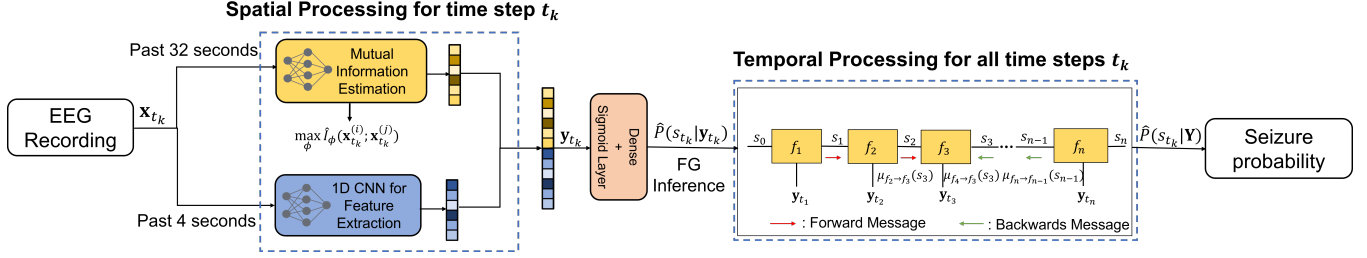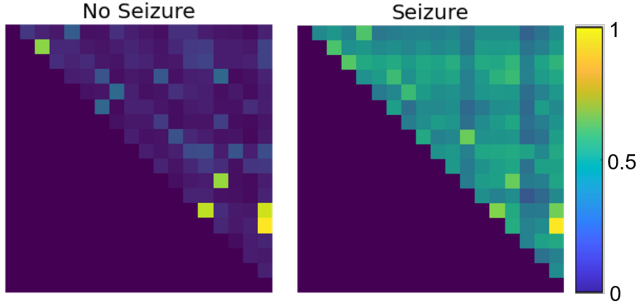
Fig. 1. MICAL illustration.

Fig. 2. Neural MI estimation for seizure and non-seizure.

3.2. 1D CNN

In parallel to the neural MI estimator, each raw EEG signals block is also processed using a dedicated 1D CNN to extract relevant features from the block. The resulting vector $\boldsymbol{y}_t$, representing the stacking of these extracted features and the estimated MI at EEG block t, is used to produce a probabilistic estimate of the presence of a seizure. We develop a 1D CNN architecture to extract meaningful features from raw EEG signals and combine these features with the MI estimation results. Compared to 2D CNNs; specifically, the baseline model proposed by Boonyakitanont et al. [12], 1D CNN can evaluate all EEG channels at a given time instance, but in 2D CNNs only the channel indexes that are close together are processed together.

To have a comparable configuration with the baseline model, we use the same number of filters. Unlike previous studies, we design the kernel size such that our 1D CNN will have a high receptive field of 1 second of the recording, compared to approximately 33 ms in prior works. This feature of the architecture leads to capturing low-frequency components of the signals and long-term temporal correlation within the 4-second blocks in EEG signals. The details of the proposed CNN model is shown in Fig 1.

3.3. Factor Graph Inference

The resulting seizure state probability using only the MI estimates and the features from 1D CNN does not exploit the presence of temporal correlation. Therefore, as proposed in [17] for sleep state tracking, we exploit the presence of temporal correlation by utilizing the block-wise soft decisions not for prediction, but as learned function nodes in a factor graph. We incorporate temporal correlation by assuming that the relationship between the extracted features $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ and the underlying seizure state $s_1, \ldots, s_n$ can be represented as an HMM. Similar modelling was shown to faithfully

capture the temporal statistics in EEG-based seizure detection [16]. For such models, one can compute the MAP rule with linear complexity using sum-product inference over the resulting factor graph, as described in Subsection 2.2. However, to evaluate the messages (2)-(3), one must be able to compute the function nodes $\{f_k\}_{k=1}^n$, given by

$$f_k(s_{t_k}, s_{t_{k-1}}) = P(s_{t_k}|s_{t_{k-1}})P(\boldsymbol{y}_{t_k}|s_{t_k}). \tag{8}$$

In MICAL, we utilize the block-wise soft decisions as estimates of the conditional distribution $P(\boldsymbol{y}_{t_k}|s_{t_k})$. The transition probability $P(s_{t_k}|s_{t_{k-1}})$, which is essentially comprised of two values, can be obtained from histogram, or manually tuned as we do in our numerical study in Section 4. The obtained marginal distributions (4) are compared to a pre-defined threshold T for detection. The resulting seizure detection algorithm is summarized as Algorithm 1.

Algorithm 1: MICAL seizure detection

1 **Inputs:** SMILE and 1D CNN networks, estimated
  $P(s_{t_k}|s_{t_{k-1}})$, EEG signals $\boldsymbol{X}$, threshold T
  Feature extraction:
2 **for** $k = 1, \ldots n$ **do**
3   Apply SMILE to estimate $\hat{I}_\phi(\boldsymbol{x}_{t_k}^{(i)}; \boldsymbol{x}_{t_k}^{(j)})$, $j > i$;
4   Apply 1D CNN to obtain combined features $\boldsymbol{y}_{t_k}$ and
      obtain soft decision;
5 **end**
  Factor graph inference:
6 Compute $\{f_k\}$ from soft decisions via (8);
7 **for** $k = 1, \ldots n$ **do**
8   Compute $\mu_{f_{t_k} \to s_k}(\{0, 1\})$ via (2);
9   Compute $\mu_{f_{t_{n-k+2}} \to s_{n-k+1}}(\{0, 1\})$ via (3);
10 **end**
11 Detect seizure at t_k if $\mu_{f_{t_k} \to s_k}(1)\mu_{f_{t_{k+1}} \to s_k}(1) > T$.

---

4. NUMERICAL RESULTS

We evaluate MICAL[1] using the CHB-MIT dataset [20]. The data is comprised of scalp EEG recordings from 24 pediatric subjects with intractable seizures, sampled at a frequency of 256 Hz where seizure start and end times are labeled. In order to balance and denoise the dataset, few simple pre-processing steps are included. Sample recordings with at least one seizure are selected and a notch filter is applied to remove the noise from power line. Due to the short

---

[1]The source code and hyper-parameters can be found on GitHub.

	AUC-ROC	AUC-PR	F1 score
2D CNN [12]	77.81 ± 0.08	37 ± 0.17	88.3 ± 0.03
Spectrogram [10]	75.4 ± 0.11	37.65 ± 0.10	92.77 ± 0.03
1D CNN	82.12 ± 0.04	42.23 ± 0.12	91.47 ± 0.02
1D CNN-GRU	82.28 ± 0.03	44.43 ± 0.10	90.42 ± 0.06
1D CNN-FG	83.15 ± 0.05	44.50 ± 0.12	92.35 ± 0.02
1D CNN-SMILE	83.10 ± 0.04	48.50 ± 0.11	92.47 ± 0.01
MICAL	$\mathbf{83.8 \pm 0.04}$	$\mathbf{50.38 \pm 0.13}$	$\mathbf{93.42 \pm 0.01}$

Table 1. Summary of results

seizure duration, for each recording file, we reduce non-seizure samples to 10 times before and 10 times after seizure time. Therefore, for every second of seizure data, there are 20 seconds of non-seizure data. The seizures are estimated for every second. To estimate the probability of seizure over the t-th second, the past 32 seconds of recording is used to solve the optimization that estimates MI. This window size has demonstrated the best results over the dataset. The past 4 seconds of recording is used as input to the 1D CNN for estimating the features. The value of 4 seconds is selected to satisfy a good trade-off between the number of samples in a block and the stationarity of the observed signals over a block. In our experiments, seven models are used for comparison. The 2D CNN used in [12] and spectrogram detector of [10] as two baseline models since they reported the best results compared to prior works. For MICAL, we tune the transition probability to $P(s_{t_k} = 1|s_{t_{k-1}} = 1) = 89.54\%$ and $P(s_{t_k} = 1|s_{t_{k-1}} = 0) = 17.90\%$. To evaluate the contribution of each individual component of MICAL, we conduct a complete ablation study. We predict seizure probability based solely on input block through 1D CNN features as well as combined features from MI estimator and CNN. We also add two different structures, including GRU cells and factor graph to the 1D CNN features to exploit temporal correlation without incorporating the inter-channel correlation. All detectors use decision threshold of $T = 0.5$.

For considering variability among patients, a 6-fold leave-4-patients-out evaluation is conducted. To examine the performance of the proposed hybrid algorithm, three metrics are measured: area under ROC curve (AUC-ROC) which shows the capability to distinguish between seizure and non-seizure samples, area under precision recall curve (AUC-PR) that is the indicative of success and failure rates, and F1 score representing the harmonic mean between precision and recall.

The results for three performance measures are summarized in Table 1. The represented values for all metrics show the average across 6 folds. As presented in Table 1, the 1D CNN used by MICAL achieves almost 5% improvement compared with the baseline models, specifically for AUC-ROC and AUC-PR. As indicated, considering only temporal or inter-channel correlation has no significant effect on the model performance. Therefore, the incorporation of MI estimation and factor graph inference by MICAL yields the highest performance measures, 83.8% and 50.38% for AUC-ROC and AUC-PR, respectively and 93.42% for F1 score. The results indicate that our algorithm admits the hypothesis of existing high correlation among signals during seizure states. Furthermore, exploiting temporal correlations in a principled manner through factor graphs is shown to facilitate learning an accurate detector, compared to using a black-box RNNs, at a much reduced computational complexity.

5. CONCLUSION

We proposed MICAL, which is a data-driven EEG-based seizure detector designed to exploit both inter-channel and temporal correla-

tions. For this, MICAL estimates the MI between each pair of EEG channels to capture the non-linear correlation among recordings observed during seizure times. The estimated MI is combined with a carefully designed 1D CNN to provide a soft estimate for each signal block. Instead of using these features for prediction, they are utilized to evaluate the function nodes of an underlying factor graph, allowing it to infer at linear complexity while exploiting temporal features between EEG blocks. We demonstrate that MICAL achieves notable improved performance compared to previously proposed methods.

Acknowledgements

6. REFERENCES

[1] B. Litt and J. Echauz, "Prediction of epileptic seizures," *The Lancet Neurology*, vol. 1, pp. 22–30, 2002.

[2] C. Park *et al.*, "Epileptic seizure detection for multi-channel EEG with deep convolutional neural network," in *2018 International Conference on Electronics, Information, and Communication (ICEIC)*.

[3] A. Subasi, J. Kevric, and M. Abdullah Canbaz, "Epileptic seizure detection using hybrid machine learning methods," *Neural Computing and Applications*, vol. 31, pp. 317–325, Jan. 2019.

[4] M. Golmohammadi *et al.*, "Deep architectures for automated seizure detection in scalp EEGs," *arXiv preprint arXiv:1712.09776*, 2017.

[5] I. B. Slimen *et al.*, "EEG epileptic seizure detection and classification based on dual-tree complex wavelet transform and machine learning algorithms," *Journal of biomedical research*, vol. 34, no. 3, p. 151, 2020.

[6] S. R. Ahmad *et al.*, "Prediction of epileptic seizures using support vector machine and regularization," in *2020 IEEE Region 10 Symposium*, pp. 1217–1220.

[7] S. Raghu *et al.*, "Automated detection of epileptic seizures using successive decomposition index and support vector machine classifier in long-term EEG," *Neural Computing and Applications*, vol. 32, no. 13, pp. 8965–8984, 2020.

[8] C. Li *et al.*, "Seizure onset detection using empirical mode decomposition and common spatial pattern," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 458–467, 2021.

[9] R. Sharan and S. Berkovsky, "Epileptic seizure detection using multi-channel EEG wavelet power spectra and 1-D convolutional neural networks," in *2020 International Conference of the IEEE Engineering in Medicine & Biology Society*, pp. 545–548.

[10] G. C. Jana, R. Sharma, and A. Agrawal, "A 1D-CNN-spectrogram based approach for seizure detection from EEG signal," *Procedia Computer Science*, vol. 167, pp. 403–412, 2020.

[11] S. Khalilpour *et al.*, "Application of 1-D CNN to predict epileptic seizures using EEG records," in *2020 International Conference on Web Research*, pp. 314–318.

[12] P. Boonyakitanont *et al.*, "A comparison of deep neural networks for seizure detection in EEG signals," *bioRxiv*, p. 702654, 2019.

[13] S. Roy, I. Kiral-Kornek, and S. Harrer, "Deep learning enabled automatic abnormal EEG identification," in *International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2756–2759, 2018.

[14] W. Liang *et al.*, "Scalp EEG epileptogenic zone recognition and localization based on long-term recurrent convolutional network," *Neurocomputing*, vol. 396, pp. 569–576, 2020.

[15] B. Salafian *et al.*, "Efficient epileptic seizure detection using CNN-aided factor graphs," *arXiv preprint arXiv:2108.02372*, 2021.

[16] M. Lee *et al.*, "Classification of both seizure and non-seizure based on EEG signals using hidden Markov model," in *2018 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pp. 469–474.

[17] N. Shlezinger *et al.*, "Learned factor graphs for inference from stationary time sequences," *IEEE Trans. Signal Process.*, 2022, early access.

[18] P. Knobelreiter *et al.*, "Belief propagation reloaded: Learning BP-layers for labeling problems," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7900–7909, 2020.

[19] N. Shlezinger *et al.*, "Data-driven factor graphs for deep symbol detection," in *IEEE International Symposium on Information Theory*, pp. 2682–2687, 2020.

[20] 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.

[21] P. D. Emmady and A. C. Anilkumar, "EEG, Abnormal Waveforms," *StatPearls*, 2020.

[22] A. Quintero-Rincón *et al.*, "A new algorithm for epilepsy seizure onset detection and spread estimation from EEG signals," in *Journal of Physics: Conference Series*, vol. 705, p. 012032, 2016.

[23] I. Jemal, A. Mitiche, and N. Mezghani, "A study of eeg feature complexity in epileptic seizure prediction," *Applied Sciences*, vol. 11, no. 4, p. 1579, 2021.

[24] H.-A. Loeliger, "An introduction to factor graphs," *IEEE Signal Process. Mag.*, vol. 21, no. 1, pp. 28–41, 2004.

[25] J. B. Kinney and G. S. Atwal, "Equitability, mutual information, and the maximal information coefficient," *Proceedings of the National Academy of Sciences*, vol. 111, no. 9, pp. 3354–3359, 2014.

[26] L. Paninski, "Estimation of entropy and mutual information," *Neural computation*, vol. 15, no. 6, pp. 1191–1253, 2003.

[27] M. I. Belghazi *et al.*, "Mutual information neural estimation," in *International Conference on Machine Learning*, pp. 531–540, PMLR, 2018.

[28] J. Song and S. Ermon, "Understanding the limitations of variational mutual information estimators," *arXiv preprint arXiv:1910.06222*, 2019.