IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 8 February 2023, accepted 23 February 2023, date of publication 6 March 2023, date of current version 10 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3252897

RESEARCH ARTICLE

MICAL: Mutual Information-Based CNN-Aided Learned Factor Graphs for Seizure Detection From EEG Signals

BAHAREH SALAFIAN[1], EYAL FISHEL BEN-KNAAN[2], NIR SHLEZINGER[2], (Member, IEEE), SANDRINE DE RIBAUPIERRE[3,4], AND NARIMAN FARSAD[1], (Senior Member, IEEE)

[1]Department of Computer Science, Toronto Metropolitan University, Toronto, ON M5B 2K3, Canada
[2]School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Be'er Sheva 84105, Israel
[3]Department of Clinical Neurological Sciences, University of Western Ontario, London, ON N6A 5B9, Canada
[4]School of Biomedical Engineering, University of Western Ontario, London, ON N6A 5B9, Canada

Corresponding author: Bahareh Salafian (bsalafian@torontomu.ca)

This work was supported in part by the Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant RGPIN-2020-04926 and in part by the Canada Foundation for Innovation (CFI), John R. Evans Leader Fund under Grant 39767.

ABSTRACT We develop a hybrid model-based data-driven seizure detection algorithm called Mutual Information-based CNN-Aided Learned factor graphs (MICAL) for detection of eclectic seizures from EEG signals. Our proposed method contains three main components: a neural mutual information (MI) estimator, 1D convolutional neural network (CNN), and factor graph inference. Since during seizure the electrical activity in one or more regions in the brain becomes correlated, we use neural MI estimators to measure inter-channel statistical dependence. We also design a 1D CNN to extract additional features from raw EEG signals. Since the soft estimates obtained as the combined features from the neural MI estimator and the CNN do not capture the temporal correlation between different EEG blocks, we use them not as estimates of the seizure state, but to compute the function nodes of a factor graph. The resulting factor graphs allows structured inference which exploits the temporal correlation for further improving the detection performance. On public CHB-MIT database, We conduct three evaluation approaches using the public CHB-MIT database, including 6-fold leave-four-patients-out cross-validation, all patient training; and per patient training. Our evaluations systematically demonstrate the impact of each element in MICAL through a complete ablation study and measuring six performance metrics. It is shown that the proposed method obtains state-of-the-art performance specifically in 6-fold leave-four-patients-out cross-validation and all patient training, demonstrating a superior generalizability.

INDEX TERMS Epilepsy, mutual information, factor graphs, convolutional neural network, deep learning, seizure, EEG, neural mutual information estimator.

I. INTRODUCTION

Epilepsy is a chronic neurological disorder that is accompanied by the sudden and unforeseen occurrence of signs or symptoms resulting from abnormal electrical activity in the brain that may cause seizures [2]. According to World Health Organization, about 50 million people worldwide are diagnosed with epilepsy [3]. The extensive sudden discharges in neural brain activity due to epileptic seizures can lead to life-threatening impacts such as involuntary movements, sensations, and emotions and may cause a temporary loss of awareness and even death [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Gina Tourassi.

A. DIAGNOSTIC TESTS FOR EPILEPSY

There are many different physiological tests as well as imaging and monitoring techniques used to evaluate if a person has a form of epilepsy, and the type of seizure the patient is experiencing. The physiological tests include

reviewing medical history [5], performing blood tests to observe the metabolic or genetic disorders associated with the seizures [6], [7], or monitoring other health conditions that could trigger epileptic seizures [7]. Imaging and monitoring are the most popular tools for detecting epileptic seizures. In this regard, various screening techniques such as magnetoencephalogram (MEG) [8], computed tomography (CT) [9], positron emission tomography (PET) [10], and magnetic resonance imaging (MRI) [11] have been employed. Among these techniques, electroencephalogram (EEG) is considered to be the most powerful method as it shows clear rhythmic electrical activities of the neurons [12].

There are two types of EEG recordings: the invasive Electrocorticography (ECoG) [13], and the non-invasive scalp EEG. ECoG is typically used when a patient is diagnosed with refractory surgery as it provides direct measurement of brain electrical activity by implanting electrodes on the cortex [14]. In scalp EEG, multiple electrodes are placed on the scalp of individuals for recording electrical activity [13]. This technique is widely preferred as it is non-invasive, economical, and portable.

From a clinical point of view, a neurologist can analyze abnormalities in EEG signals through visual inspection to understand the presence or the type of epileptic seizures. However, this diagnosis is time-consuming as it requires careful inspection of data from long recording sessions by a neurologist [15], and is subject to inter-observer variability [16]. Moreover, EEG measurements are usually contaminated by undesired artifacts and noise that can interfere with neural information and cause misdiagnosis of epileptic seizures [17]. To address these issues, an automatic seizure detection algorithm is desirable.

B. AUTOMATED SEIZURE DETECTION FROM EEG SIGNALS

Since scalp EEG signals are non-invasive and pain-less with fewer side effects, they have been broadly used in diagnosing neural disorders such as stroke [18], [19], sleep disorders [20], and epilepsy. In this work, we focus on detecting epileptic seizures.

Several different methods for automatically detecting seizures from EEG recordings have been proposed in the literature. These techniques can be categorized into two different approaches: Machine learning and signal processing approaches based on engineered features extracted from the recordings, and machine learning approaches applied to raw EEG recordings.

1) FEATURE-BASED DETECTION

Spike detection is the most popular feature-based method that aims to identify seizure spikes in the multichannel EEG recording with high sensitivity and selectivity [21]. Generally, spike detection methods are divided into different categories including template matching [22], [23], mimetic analysis [24], [25], power spectral analysis [26], wavelet analysis [27], and techniques based on artificial neural networks [28], [29], [30].

Most of the recent feature-based designs use the extracted features as input to deep learning (DL) algorithms for seizure detection. The authors in [31] first generated three different features based on Fourier, wavelet, and empirical mode decomposition transforms. They then applied a shallow 2D convolutional neural network (CNN) and concluded that Fourier transform achieved the best results. Another wavelet-based deep learning approach was performed in [32]. In this method, discrete wavelet transform (DWT) was used to extract time-frequency domain features in five sub-band frequencies, and then, a 2D CNN architecture was employed to learn the features from predefined coefficients. Applying 1D CNN to time-frequency features was observed in [33], [34]. In the proposed method, first DWT of signals was processed, and a 1D CNN architecture then performed detection.

Feature-based design mainly depends upon the expert definition of EEG characteristics such as slop, duration, height, and sharpness, which are not sufficient enough to represent an epileptic seizure spike, and it can result in high false detection rate [35]. Moreover, using a different transforms such as Fourier or wavelet transform as input to DL models requires careful engineering and considerable domain expertise to design a feature extractor that transforms the raw data into a suitable representation [36]. However, this strategy is difficult since various types of patterns appear when a seizure occurs [37]. Moreover, many interfering artifacts in the signal, for example due to blinking or muscle activity, can have structures similar to the seizure patterns.

2) SIGNAL-BASED DETECTION

In the past decade, different DL models have been investigated and tested in the area of seizure detection and analysis of time series EEG signal. For example, in [38], after using a notch filter, a 1D CNN with few convolutional layers was employed to detect interictal epileptiform spikes due to seizures. Acharya et al. [39] applied a 3-layer CNN architecture to the normalized EEG signals. A 2D CNN architecture was implemented in [40] and applied to a multi-class classification problem, where the EEG was labeled according to different stages of seizure. Boonyakitanont et al. [41] proposed a detection scheme using raw EEG records divided into 4-second blocks followed by a deep 2D CNN architecture to learn the features from EEG signals. They showed state-of-the-art performance based on the detection accuracy when per-patient training was employed.

To capture the temporal dependencies, some prior work have explored recurrent neural networks (RNNs) [42], [43]. Hussein et al. [44] used a deep RNN, particularly an Long-Short Term Memory (LSTM) to the segmented EEG signals to learn the most robust features from recordings. Aristizabal et al. [45] developed another LSTM-based seizure detection technique for six pairs of EEG signals. The performance of GRU was explored in [46]. In this model, the GRU-hidden unites were used to classify EEG into three different classes: healthy, inter-ictal and ictal (i.e.,

seizure) states. Roy et al. [47] proposed an architecture termed ChronoNet by stacking multiple 1D convolution layers followed by GRU layers.

While signal-based designs are widely used in the literature, training models that are generalizable and perform well across different patients requires large networks and very large datasets. This is due to patient-to-patient variability and the fact that EEG recordings are inherently very noisy. Hence, end-to-end training on raw EEG data may not achieve the best performance in practice due to lack of access to large datasets, as well as limitations imposed by computational complexity, both during training or during inference. 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.

C. CONTRIBUTIONS

In this paper, we propose Mutual Information-based CNN-Aided Learned factor graphs (MICAL), which is a hybrid model-based/data-driven approach [48] for automatic seizure detection. This algorithm contains three main novel aspects:

- In contrast to prior works, we carefully design a 1D CNN to process the EEG signals with a higher receptive field and minimal preprocessing. This results in extracting features that capture longer-term dependencies from raw EEG signals.
- We use a neural mutual information (MI) estimator to compute the *inter-channel* dependence between EEG channels during the seizure. When a seizure occurs in one or more EEG channels, the patterns of other channel recordings are affected, and the signals between the channels become correlated at the beginning and during ictal (i.e., seizure) state [49], [50]. Compared to the traditional methods for evaluating levels of dependence, including cross-correlation, MI can capture higher-order statistical dependence between recordings. This is helpful in seizure detection since non-linear relationships often exist between EEG channels during a seizure.
- We propose an inference method that uses the estimates obtained from the extracted features at each time interval to form a learned factor graph [51], which captures the *temporal* correlations in EEG recordings. By applying message passing over the learned factor graph, seizures can be detected more efficiently than deep learning approaches based on RNNs.

The performance of MICAL is comprehensively explored via six performance measures as well as three evaluation strategies. Using an extensive ablation study we systematically show that each component of MICAL contributes positively to its performance. Comparing MICAL with prior works, we demonstrate that our proposed method achieves state-of-the-art results in terms of performance and generalizability, with the gains being most notable when given access to medium sized datasets. It should be noted that a dataset with more than 1000 patients is considered large here, while a dataset with less than 100 patients is considered small. A data set with 100s of patients is considered a medium-sized dataset.

The rest of this paper is organized as follows. In Section II, we discuss the seizure detection problem as well as the challenges related to traditional approaches for seizure detection. Section III describes the proposed MICAL. A comprehensive numerical evaluation of MICAL is reported in Section IV. Finally, we conclude the paper with summary discussions and future research direction in Section V.

II. EEG-BASED SEIZURE DETECTION SYSTEM MODEL

As mentioned in Section I-A, automatic seizure detection can improve the traditional approach, which is based on time-consuming visual inspection by neurologists. Moreover, for some patients, long-term monitoring is required for diagnosis. In these cases, the patients or their families are asked to report the number of seizures that occur during their daily lives. However, this approach has considerable limitations due to inaccurate descriptions of seizures and their frequency. Hence, automatic seizure detection can also provide a more elaborate and accurate technique for quantifying the number/type of seizures during long-term patient monitoring, resulting in better research, diagnosis, and selection of appropriate treatment options.

In this paper, seizure detection refers to the identification and localization of the ictal time intervals from EEG recordings of patients with epileptic seizures [52]. 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 that 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 $\boldsymbol{s}$, one must consider both inter-channel dependence as well as temporal correlations underlying the recordings. The former stems from the fact that when the seizure starts, the epileptic activity propagates to other areas in the brain [50], which affects the patterns of other channel recordings [53]. This manifests as a high dependence between different channels, i.e., between $\boldsymbol{x}_t^{(i)}$ and $\boldsymbol{x}_t^{(j)}$, $i \neq j$, when t is at the beginning and during ictal phase. Fig. 1 demonstrates the signal patterns during seizure vs. no-seizure states. Our proposed solution, detailed in Subsection III-A, uses neural MI estimators to capture this dependency.

Temporal correlation results from the fact that seizures typically span multiple recording blocks. Thus, the probability of observing a seizure at time instance t depends on the presence of a seizure in the previous block, and as a result the entries

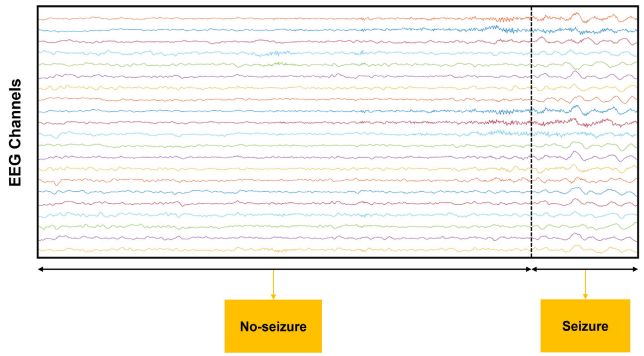


FIGURE 1. Inter-channel correlation during seizure vs. no-seizure.

of $\boldsymbol{s}$ can be approximated by a Markovian structure [54]. Our proposed solution, detailed in Subsection III-C, exploits this statistical structure using factor graphs.

In the next section, we describe our proposed approach to automatic seizure detection in detail.

III. PROPOSED MICAL ALGORITHM

In this section we present the proposed MICAL seizure detector. Our design of MICAL is based on the following considerations:

1) The level of statistical dependence between different channels provides an indication for the presence of a seizure.
2) Direct processing of the signal is preferable as it avoids the need for careful feature engineering.
3) The temporal correlation between different blocks can be approximated as obeying a Markovian structure.
4) The detection algorithm must be operable with low complexity and should not require massive data sets for its training.

Based on these consideration, we propose MICAL, whose structure is illustrated in Fig 2. In the rest of this section, we describe each component of MICAL. To account for the inter-channel dependence 1, we employ a neural MI estimator block described in Subsection III-A; To extract more features from the EEG recordings directly following 2, we design a 1D CNN architecture detailed in Subsection III-B. These two blocks are used to estimate seizure probability for a given block. Finally, to account for temporal correlations following 3 and do so in computationally efficient manner 4, we use estimates over different blocks to form a learned factor graph, as detailed in Subsection III-C.

A. NEURAL MUTUAL INFORMATION ESTIMATION

In order to compute inter-channel correlation among recordings, the most popular approach is cross-correlation, which is a measure of similarity between one signal and the time-delayed version of other signals. However, this method cannot capture the nonlinear relationship between samples, which are likely to occur in EEG signals during the seizure. Unlike cross-correlation, MI represents higher-order joint statistics and is thus able to capture arbitrary statistical dependence between samples, even in the presence of nonlinear relationship between the signals. Mathematically, MI can be formulated as:

$$I(X_1; X_2) = \int_{\mathcal{X}_1 \times \mathcal{X}_2} \log \left(\frac{d\mathbb{P}_{X_1X_2}}{d\mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2}} \right) d\mathbb{P}_{X_1X_2} \quad (1)$$

where $\mathbb{P}_{X_1X_2}$ is the joint probability distribution and $\mathbb{P}_{X_1}$ and $\mathbb{P}_{X_2}$ are marginals. The X_1 and X_2 represent two random variables where in the case of seizure detection, they can be interpreted as the recordings for two different channels.

The MI can be expressed as the Kullback-Leibler (KL) divergence between the joint and the product of the marginals of two random variables X_1 and X_2 [55]:

$$I(X_1, X_2) = D_{KL}(\mathbb{P}_{X_1X_2} \| \mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2}) \quad (2)$$

where D_{KL} is defined as:

$$D_{KL}(\mathbb{P} \| \mathbb{Q}) := \mathbb{E}_{\mathbb{P}} \left[\log \frac{d\mathbb{P}}{d\mathbb{Q}} \right] \quad (3)$$

Although MI is a reliable measure to capture statistical dependence, the exact calculation based on (1) and (3) for finite continuous and non-parametric EEG samples is challenging [56]. To facilitate MI computation we use the Smoothed Mutual Information "Lower-bound" Estimator (SMILE) of [57], which provides further improvements on Mutual Information Neural Estimator (MINE) proposed in [58]. Thus, to describe our MI estimator, we briefly explain the operation of MINE and that of SMILE.

A key technical aspect of MINE is dual representations of the KL-divergence, which is based on Donsker-Varadhan representation [59]. This representation leads to the following lower bound where the supremum is taken over all functions T such that the two expectations are finite.

$$D_{KL}(\mathbb{P} \| \mathbb{Q}) \geq \sup_{T \in \mathcal{F}} \mathbb{E}_{\mathbb{P}}[T] - \log \left(\mathbb{E}_{\mathbb{Q}} \left[e^T \right] \right) \quad (4)$$

Using both (3) and dual representation of KL-divergence, the idea is to choose $\mathcal{F}$ to be the set of functions $T_\theta : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}$ parametrized by a deep neural network with parameters $\theta \in \Theta$, and the apply moving average gradient ascent to find the optimal parameters. This network is called statistics network, where the bound is calculated as:

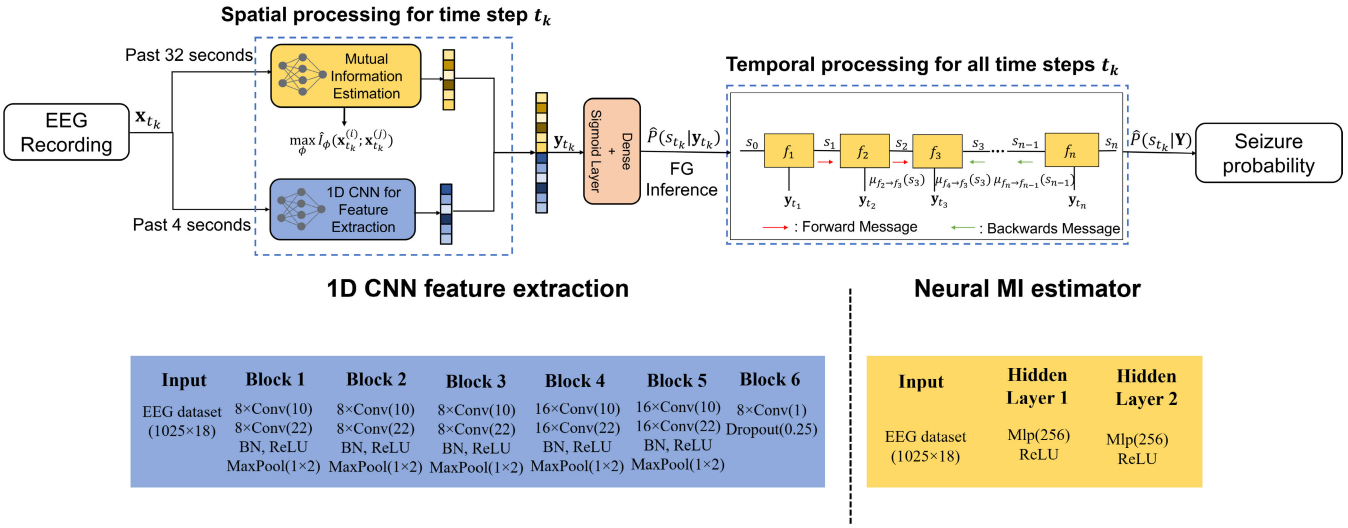
$$I(X_1; X_2) \geq I_\theta(X_1, X_2) \quad (5)$$

and the neural information measure, $I_\theta(X_1, X_2)$ is defined as:

$$I_\theta(X_1, X_2) = \sup_{\theta \in \Theta} \mathbb{E}_{\mathbb{P}_{X_1X_2}}[T_\theta] - \log \left(\mathbb{E}_{\mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2}} \left[e^{T_\theta} \right] \right) \quad (6)$$

An important limitation of MINE is the large variance of the estimator, which can grow exponentially with the ground truth MI value to be estimated from the samples. In order to solve the variance problem, [57] proposes SMILE by introducing a clipping function in (6), resulting in

$$\hat{I}_{\boldsymbol{\theta}}(X_1; X_2) = \mathbb{E}_{P_{X_1X_2}}[T_{\boldsymbol{\theta}}] - \log \mathbb{E}_{P_{X_1}P_{X_2}} \left[clip(e^{T_{\boldsymbol{\theta}}}, e^{-\tau}, e^{\tau}) \right], \quad (7)$$



Input	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6
EEG dataset (1025×18)	8×Conv(10) 8×Conv(22) BN, ReLU MaxPool(1×2)	8×Conv(10) 8×Conv(22) BN, ReLU MaxPool(1×2)	8×Conv(10) 8×Conv(22) BN, ReLU MaxPool(1×2)	16×Conv(10) 16×Conv(22) BN, ReLU MaxPool(1×2)	16×Conv(10) 16×Conv(22) BN, ReLU MaxPool(1×2)	8×Conv(1) Dropout(0.25)

Input	Hidden Layer 1	Hidden Layer 2
EEG dataset (1025×18)	Mlp(256) ReLU	Mlp(256) ReLU

FIGURE 2. MICAL illustration including spatial processing via MI estimation and 1D CNN processing followed by factor graph inference for temporal processing.

where $clip(v, l, u) = \max(\min(v, u), l)$, and τ is a constant parameter that provides a knob for better tuning the bias-variance trade-off. In [57], it was shown that while τ can be tuned to reduce the variance, it may not increase the bias significantly.

In order to ensure MI is a good measure to compute dependency between EEG channels, we compare SMILE with three other popular correlation measures:

- The instantaneous phase synchrony measures the phase similarities between signals at each time-point [60].
- The Pearson correlation measures the strength of the linear relationship between two random variables [61].
- The distance correlation is a measure of association strength between non-linear random variables [62].

Fig. 3 compares these three methods with SMILE. Specifically, we evaluate the degree of dependence between all pairs of channels for time blocks of 4 seconds during seizure as well as during no-seizure regions. We then evaluate average across all blocks for the seizure and no-seizure zones. Based on these mean values, SMILE is the most powerful indicator of highly inter-channel correlation during the seizure compared to the other three correlation measures, which was in agreement with manual inspection of the EEG signals.

Based on the above observation, we use SMILE in MICAL 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 the parametric $T_{\boldsymbol{\theta}}$ to be a fully-connected network with two hidden layers with ReLU activations, and we set $\tau = 0.9$ in the objective (7).

B. 1D CNN

In parallel to MI estimation, we design and employ a 1D CNN in order to produce latent representation of raw EEG signals. For having a similar configuration with the baseline model [41], the same number of layers, including convolutions, pooling, dropout, and fully connected layers, are chosen. As shown in Fig. 2, we design the kernel size to obtain a high receptive field compared to prior works. Our proposed 1D CNN is able to cover almost 1 second of data, while previous studies had a receptive field of only 30 milliseconds. Having a high receptive field leads to capturing long-term correlation as well as capturing low-frequency components of EEG signals. Moreover, compared to 2D CNN, 1D CNN can operate on all EEG channels at a given time instance. Details of the proposed CNN are described in Fig. 2.

The final set of features that are used for estimating the probability of seizure over a given block is obtained by combining the result of the 1D CNN extractor and the estimated MI. Specifically, let $\hat{\mathbf{I}}_{t_k}$ be the estimated MI values between channel pairs at time t_k, and $\mathbf{z}_{t_k}$ be the features extracted by the 1D CNN. The the final set of features used for seizure detection is given by $\mathbf{y}_{t_k} = [\hat{\mathbf{I}}_{t_k}, \mathbf{z}_{t_k}]$. This feature is then used as an input to a logistic regression layer for a soft estimate of the seizure event.

C. FACTOR GRAPH INFERENCE

Predicting seizure solely based on combined features from MI estimator and 1D CNN does not take into account past and future EEG blocks. Therefore, we utilize the block-wise soft decision not as a direct estimate of the corresponding seizure state, but as learned function nodes in a factor graph incorporating the presence of temporal correlation. Our proposed approach follows the methodology of learned factor graphs utilized for sleep stage detection in [63] and for symbol detection in [64]. To describe this operation, we first briefly recall factor graph inference, after which we explain how it is incorporated by MICAL to account for temporal correlation.

Factor graphs are a representation of the factorization of local functions of several variables, typically of joint

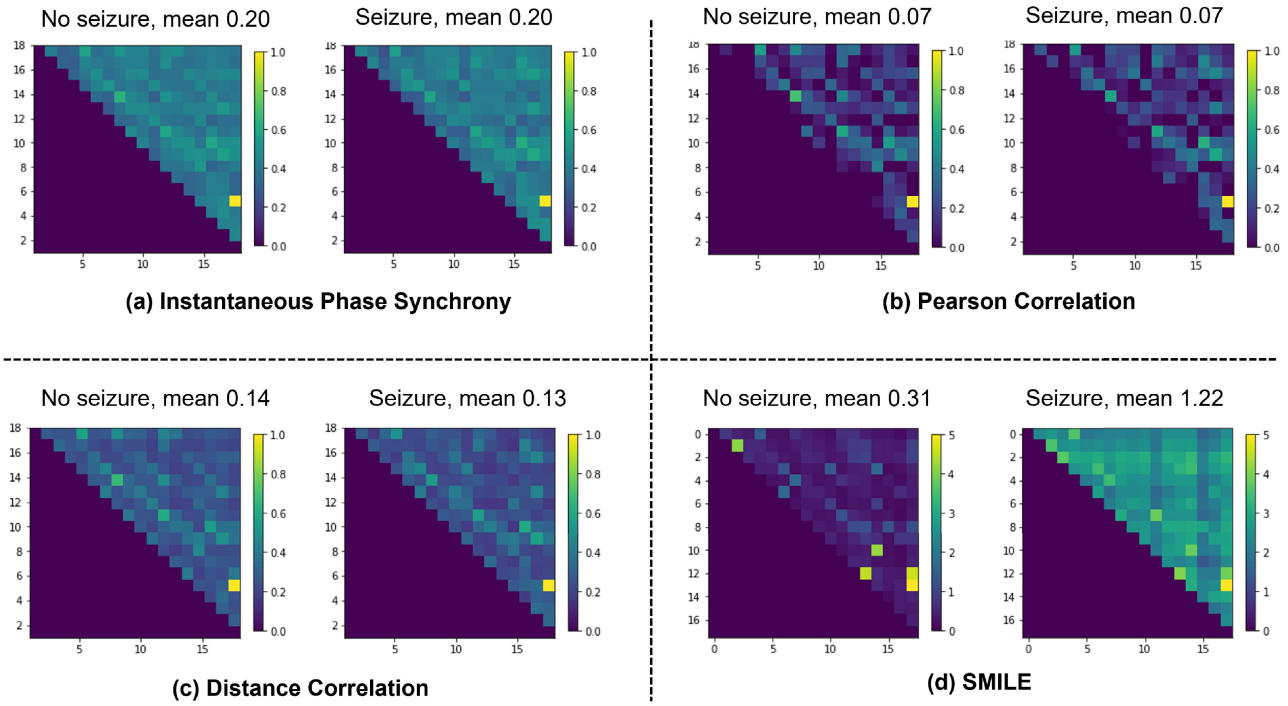


FIGURE 3. Four different correlation measures for seizure and non-seizure. (a) instantaneous phase synchrony, (b) Pearson correlation, (c) distance correlation, and (d) SMILE.

distribution measures, forming a graphical structure [65]. In the Forney-style factor graph (adopted here) of a joint distribution, the random variables correspond to edges, and their statistical dependence is captured as a node in the graph, referred to as the function node. As a result, a factor node is connected to a variable edge if and only if the factor is a function of the variable. The key advantage of this representation is that it facilitates extracting quantities which are typically complex to compute, such as marginal probabilities, with a complexity that only grows linearly with the number of variables via, e.g., the sum-product method [66]. To implement factor graph inference, the first step is to create the structure of the graph, i.e., the interconnection between nodes. For this purpose, we use the underlying property of seizure mechanism where the generation of seizure is closely associated with abnormal synchronization of neurons [67]. To incorporate this feature in our model and following consideration 3, we approximate the temporal relationship as obeying first-order Markovian model.

To formulate this mathematically, the Markovian model implies that the joint distribution of the extracted features $\boldsymbol{y}$ and the latent seizure states $\boldsymbol{s}$ over all N blocks can be factorized as

$$P(\boldsymbol{s}, \boldsymbol{y}) = \prod_{i=1}^{N} P(s_i|s_{i-1})P(y_i|s_i). \tag{8}$$

Here, $P(s_i|s_{i-1})$ represents the seizure state transition probability, which is a control parameter. Based on our numerical experiments, we manually set it to be 89.54% for switching from no-seizure to seizure and 17.9% for opposite situation. The factorization in (8) results in the factor graph representation of the joint distribution $P(\boldsymbol{s}, \boldsymbol{y})$ as the sequential graph illustrated in Fig. 2

The classification of the sleep states requires to compute marginal distribution $P(s_i, \boldsymbol{y})$ from (8). The sum-product algorithm allows to compute this desirable quantity in a recursive manner via forward and backward message exchanges over the factor graph. In particular, the sum-product method computes the marginal probabilities via

$$P(s_k, \boldsymbol{y}) = \mu_{f_j \to s_k}(s_k) \cdot \mu_{f_{j+1} \to s_k}(s_k), \tag{9}$$

for each $k \in \{1, \ldots, N\}$. In (9), $\mu_{f_j \to s_k}(s_k)$ is interpreted as forward message

$$\mu_{f_j \to s_k}(s_k) = \sum_{\{s_1, \cdots, s_{k-1}\}} \prod_{i=1}^{n} f_i(y_i, s_i, s_{i-1}), \tag{10}$$

and $\mu_{f_{j+1} \to s_k}(s_k)$ as the backward message is achieved by

$$\mu_{f_{j+1} \to s_k}(s_k) = \sum_{\{s_{k+1}, \cdots, s_N\}} \prod_{i=n+1}^{N} f_i(y_i, s_i, s_{i-1}), \tag{11}$$

where $f_i(y_i, s_i, s_{i-1})$ is the function node which is given by

$$f_i(y_i, s_i, s_{i-1}) = P(s_i|s_{i-1})P(y_i|s_i). \tag{12}$$

The resultant marginal distributions (9) are compared to a predefined threshold of $T = 0.7$ for detection.

According to (12), implementing sum-product algorithm requires the knowledge of probability distribution $P(y_i|s_i)$. In practice, obtaining this statistical model that relates observations and time series is a highly complex process.

Following [63], we use the joint features of MI estimator and 1D CNN as the soft estimate of probability distributions to learn function nodes in factor graph. Algorithm 1 summarizes the steps in MICAL seizure detection.

Algorithm 1 MICAL Seizure Detection

Inputs: SMILE and 1D CNN networks, estimated state transitions, EEG measurements $\boldsymbol{X}$, threshold T
Feature extraction:
for $k = 1, \ldots n$ **do**
  Apply SMILE to estimate $\hat{I}_{\boldsymbol{\theta}}(\boldsymbol{x}_{t_k}^{(i)}; \boldsymbol{x}_{t_k}^{(j)}), j > i$;
  Apply 1D CNN to obtain combined features $\boldsymbol{y}_{t_k}$;
  Apply dense layer to obtain soft decision $\hat{P}(s_{t_k}|\boldsymbol{y}_{t_k})$;
end
Factor graph inference:
Compute $\{f_i\}$ from soft decisions via (12);
for $k = 1, \ldots n$ **do**
  Compute $\mu_{f_{t_k} \to s_k}(\{0, 1\})$ via (10);
  Compute $\mu_{f_{t_{n-k+2}} \to s_{n-k+1}}(\{0, 1\})$ via (11);
end
Detect seizure at t_k if $\mu_{f_{t_k} \to s_k}(1)\mu_{f_{t_{k+1}} \to s_k}(1) > T$.

IV. RESULTS AND DISCUSSION

In the following subsections, we describe our experimental study of MICAL for seizure detection.[1] We first explain the data used, performance metrics, evaluation methods, and hyperparameter tuning, in Subsections IV-A-IV-C, respectively. We then present the numerical results along with a discussion in Subsection IV-D.

A. DATA DESCRIPTION

1) EEG DATA

The dataset used in this work is publicly available CHB-MIT Database collected at the Children's Hospital Boston. CHB-MIT consists of scalp EEG recordings from pediatric subjects with intractable seizures [68].[2] Recordings belong to 24 cases with ages from 1.5 to 22. Each patient contains 9 to 42 EDF files from a single subject. All signals were sampled at 256 frequency with 16-bit resolution and seizure start and end times are labeled. Table 1 represents the details of the dataset.

2) EEG PRE-PROCESSING

The dataset contains 664 EDF files from all patients. Signals were already annotated by ''seizure'' and ''no-seizure'' labels where each case has at least two seizures. Note that for the CHB-MIT database, seizure types are not specified. Unlike prior works, we consider simple pre-processing steps for our proposed algorithm that makes it well-suited for real-time applications. Since seizures duration (from 7 seconds to 753 seconds) compared to overall recording (from 959 seconds to 14427 seconds) is very short and to have a balanced dataset; first, EDf files that include at least one seizure are selected. Each recording is then shortened to 10 times the seizure duration before and 10 times the seizure duration after the seizure. Therefore, there are 20 seconds of non-seizure data for every second of seizure data. Fig. 4 represetns the electrode placement according to the international 10-20 system. From EEG channels, the 18 bipolar montages are chosen: FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-T3, T3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ, CZ-PZ. In our proposed method compared to previous studies, to reduce complexity, we aim to use minimal pre-processing. Therefore, a notch filter is applied to remove 60 Hz line noise from each EEG signal. 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 is large enough to incorporate the correlation among measurements during the ictal state and demonstrates the best results over the dataset. In addition, the past 4 seconds of recording is used as input to the 1D CNN for estimating meaningful features from EEG blocks. 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.

TABLE 1. CHB-MIT database details.

	Information
Open source database	Yes
Sampling frequency (Hz)	256
Electrode position	10-20
No. of subjects (Female)	18
No. of subjects (Male)	5
Age range (Female)	1.5–19
Age range (Male)	3-22
Total duration (h)	884
No. of channels	23
No. of seizures	182
File format	EDF

B. EVALUATION METHODS AND PERFORMANCE METRICS

To evaluate the performance of the models, six following metrics are measured:

- *AUC-ROC*: is the area under receiver operating characteristics (ROC) curve, which shows the capability of the model to distinguish between seizure and no-seizure samples.
- *AUC-PR*: is the area under the precision-recall curve that represents success and failure rates meaning that a high area under the curve shows a low false positive rate and low false-negative rate.
- *Precision*: intuitively shows the ability of the classifier not to label a sample as positive that is negative.
- *Recall*: represents the capability of the classifier to find all the positive samples.
- *F1 score*: is a harmonic mean of recall and precision.
- *Accuracy*: implies the number of correct predictions over the total number of predictions.

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

[2] This database is available online at PhysioNet (https://physionet.org/physiobank/database/chbmit/)

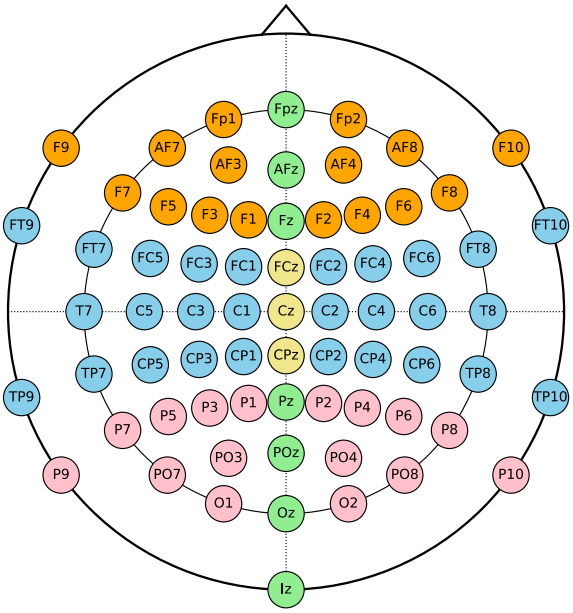


FIGURE 4. International 10-20 electrodes system.

In our implementation, we consider three different strategies for training the models:

- *6-fold-leave-4-patient-out evaluation*: is a generalized evaluation pipeline as it creates all the possible train and test sets. We follow the rules to consider 80% of the patients for the train and 20% for the test. Therefore, in this approach for each fold, 20 patients are considered as the train dataset and 4 different cases are kept out for the test. This approach requires building six models.
- *All patient evaluation*: chooses one EDF file from each patient as the test and trains on the remaining files of all cases. Since one of the patients has only two seizure files and each time a different random EDF session is considered as the test, this algorithm creates two models.
- *Per patient evaluation*: selects an EDF file with maximum length as the test and trains the model on the remaining samples for each case separately. This approach creates 24 models for every patient in the dataset.

C. HYPERPARAMETER TUNING

As mentioned in Subsection IV-A, for every file, the ratio between seizure and no-seizure classes is 1:20. This bias in the training dataset can influence the model performance and result in poor performance over the minority class (seizure samples). One approach to address the problem of class imbalance is to randomly duplicate examples from the minority class, which is called oversampling. To incorporate oversampling, we select half of the samples in the batch from the non-seizure class and randomly sample the other half from the seizure class.

One of the essential hyperparameters in training a neural network is the learning rate. If the learning rate is too low, the weights will be updated slowly. On the other hand, setting the learning rate to significant values will cause undesirable divergent behavior in the loss function. Therefore, we use Learning Rate Schedules that adjust the learning rate during training by reducing the learning rate according to a predefined schedule. Here, we start with 0.0001, and the factor of 0.5 reduces it until it reaches 10^{-5}. We also use ADAM optimizer, and the number of epochs for 6-fold-leave-4-patient-out evaluation and all patient training is 100, and 20 is selected as the number of epochs for per patient training. The batch size for all methods is 256.

D. NUMERICAL RESULTS

In our experiment, we consider eight different configurations for comparison, including two baseline models since they showed the best results compared to previous studies. One model is related to non-feature-based design, and the other is based on a deep learning feature extraction algorithm. The 2D CNN was employed to raw EEG signals in [41] and Jana et al. [69] proposed an architecture comprised of spectrogram features and a 1D CNN; however, we use our proposed CNN model to have the same backbone. To explore the effect of each individual element in MICAL on the performance metrics, a complete ablation study is conducted. Hence, it results in three different situations: 1) ignoring inter-channel and temporal correlations among EEG recordings (1D CNN), 2) Excluding MI estimation results (1D CNN-FG), 3) Removing temporal correlations (1D CNN-SMILE). As another experiment for comparison, we add GRU to 1D CNN and 1D CNN-SMILE to ensure factor graph is strong enough to exploit temporal correlations while it reduces the complexity compared to RNNs.

Table 2 to Table 4 as well as Fig 5 to Fig 8 summarize the average results of all performance metrics for three evaluation pipelines. Since the oversampling approach is not used for evaluation samples, this dataset is imbalanced. Therefore, the accuracy is not a good measure to investigate the performance of the models, and as shown in the tables, the values for this metric are inconsistent. For instance, in Table 2, and Fig 6, MICAL shows the lowest accuracy score while it achieves the best results for AUC-ROC, AUC-PR, F1 score, and recall. As represented, although MICAL can exploit inter-channel and temporal correlations, the accuracy score for our algorithm is less than 1D CNN and 1D CNN-SMILE. As such, in our experiments we mainly observe the performance of other five metrics. As shown in Fig 5 to Fig 8, MICAL indicates the highest state-of-the-art results compared to the other models in ablation study. These results are aligned with our expectation as the proposed method has an efficient 1D CNN unlike baseline models. In addition, considering factor graphs to capture temporal correlations efficiently as well as adding MI estimator for inter-channel correlations result in much higher performance outcome. In the rest of this subsection, we explain the results for each training approach in more details.

The average results for 6-fold leave-4-patients-out evaluation are listed in Table 2, and shown in Fig 6. We find that compared to the baseline models, our proposed CNN architecture improves the AUC-PR and recall by 9% and

TABLE 2. Summary of results for 6-fold leave-4-patients-out validation.

	AUC-ROC	AUC-PR	F1 score	Precision	Recall	Accuracy
2D CNN [41]	75.86 ± 0.08	31.96 ± 0.16	31.25 ± 0.11	27.09 ± 0.1	44.96 ± 0.22	89.08 ± 0.03
Spectrogram [69]	71.13 ± 0.1	30.45 ± 0.14	28.73 ± 0.09	28.68 ± 0.11	36.18 ± 0.19	89.54 ± 0.05
1D CNN (Ours)	74.98 ± 0.07	39.6 ± 0.16	35.28 ± 0.11	32.46 ± 0.13	44.79 ± 0.15	$89.36 \pm 0.0.06$
1D CNN-GRU (Ours)	76.65 ± 0.07	36.99 ± 0.15	36.6 ± 0.11	33.33 ± 0.17	46.91 ± 0.17	89.98 ± 0.04
1D CNN-FG (Ours)	77.13 ± 0.07	42.04 ± 0.15	37.53 ± 0.12	35.45 ± 0.15	46.13 ± 0.14	89.73 ± 0.06
1D CNN-SMILE (Ours)	84.25 ± 0.05	41.8 ± 0.16	37.28 ± 0.09	31.56 ± 0.1	52.65 ± 0.18	89.21 ± 0.05
1D CNN-SMILE-GRU (Ours)	81.65 ± 0.04	40.48 ± 0.15	37.51 ± 0.07	$\mathbf{39.21 \pm 0.12}$	43.71 ± 0.18	$\mathbf{91.43 \pm 0.04}$
MICAL (Ours)	$\mathbf{86.01 \pm 0.05}$	$\mathbf{44.06 \pm 0.16}$	$\mathbf{38.25 \pm 0.1}$	31.05 ± 0.11	$\mathbf{57.88 \pm 0.19}$	88.21 ± 0.06

TABLE 3. Summary of results for all patient training.

	AUC-ROC	AUC-PR	F1 score	Precision	Recall	Accuracy
2D CNN [41]	86.8 ± 0.01	48.35 ± 0.009	43.9 ± 0.02	34.55 ± 0.02	60.19 ± 0.01	91.45 ± 0.006
Spectrogram [69]	84.65 ± 0.02	50.8 ± 0.07	46 ± 0.06	38.85 ± 0.06	56.55 ± 0.06	92.6 ± 0.01
1D CNN (Ours)	90.85 ± 0.01	65.66 ± 0.02	56.2 ± 0.02	45.55 ± 0.03	74.35 ± 0.005	93.55 ± 0.006
1D CNN-GRU (Ours)	88.20 ± 0.01	59.69 ± 0.02	59.19 ± 0.01	53.84 ± 0.01	66.14 ± 0.05	94.95 ± 0.001
1D CNN-FG (Ours)	93.75 ± 0.002	73.1 ± 0.01	64.3 ± 0.02	55.75 ± 0.04	76.3 ± 0.01	95.3 ± 0.007
1D CNN-SMILE (Ours)	93.75 ± 0.0004	68.7 ± 0.01	59.54 ± 0.01	50.15 ± 0.02	73.55 ± 0.01	94.44 ± 0.004
1D CNN-SMILE-GRU (Ours)	91.75 ± 0.01	63.84 ± 0.07	61.1 ± 0.03	55.69 ± 0.03	67.6 ± 0.03	95.19 ± 0.004
MICAL (Ours)	$\mathbf{95.6 \pm 0.001}$	$\mathbf{74 \pm 0.01}$	$\mathbf{65.4 \pm 0.02}$	$\mathbf{56.09 \pm 0.03}$	$\mathbf{78.75 \pm 0.01}$	$\mathbf{95.39 \pm 0.005}$

TABLE 4. Summary of results for per patient training.

	AUC-ROC	AUC-PR	F1 score	Precision	Recall	Accuracy
2D CNN [41]	72.82 ± 0.18	34.72 ± 0.29	31.77 ± 0.24	27.63 ± 0.26	53.77 ± 0.28	79.82 ± 0.19
Spectrogram [69]	74.03 ± 0.19	35.30 ± 0.22	20.81 ± 0.21	25.96 ± 0.3	27.07 ± 0.29	90.93 ± 0.06
1D CNN (Ours)	82 ± 0.19	50.59 ± 0.33	37.53 ± 0.29	42.24 ± 0.34	48.77 ± 0.35	90.16 ± 0.1
1D CNN-GRU (Ours)	85.99 ± 0.12	52.23 ± 0.35	36.92 ± 0.34	46.15 ± 0.4	36.55 ± 0.33	93.69 ± 0.08
1D CNN-FG (Ours)	86.17 ± 0.19	58.07 ± 0.35	46.09 ± 0.31	40.47 ± 0.33	76 ± 0.27	81.1 ± 0.21
1D CNN-SMILE (Ours)	88.53 ± 0.11	61.86 ± 0.3	52.11 ± 0.29	51.87 ± 0.31	61.06 ± 0.32	92.19 ± 0.1
1D CNN-SMILE-GRU (Ours)	$\mathbf{90.83 \pm 0.11}$	$\mathbf{68.87 \pm 0.29}$	52.04 ± 0.34	$\mathbf{67.32 \pm 0.39}$	46.79 ± 0.34	$\mathbf{95.69 \pm 0.05}$
MICAL (Ours)	90.44 ± 0.1	66.77 ± 0.3	$\mathbf{53.46 \pm 0.31}$	50.91 ± 0.32	$\mathbf{69.59 \pm 0.32}$	88.89 ± 0.15

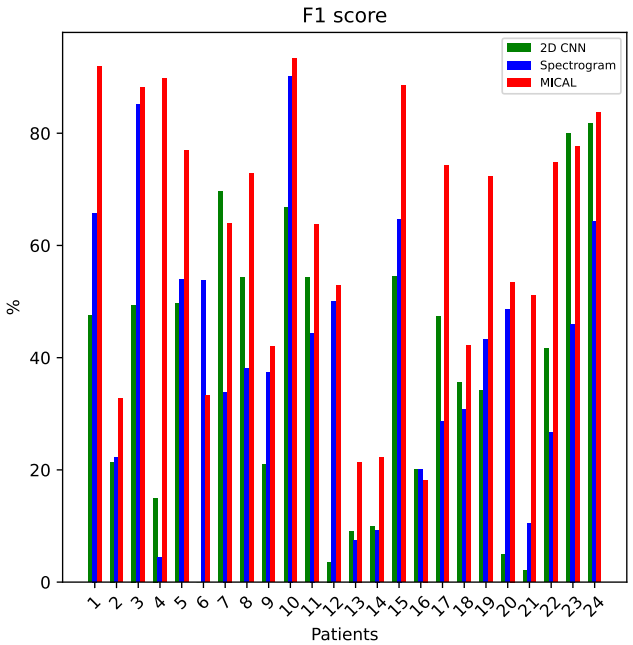


FIGURE 5. F1 score for all patient training.

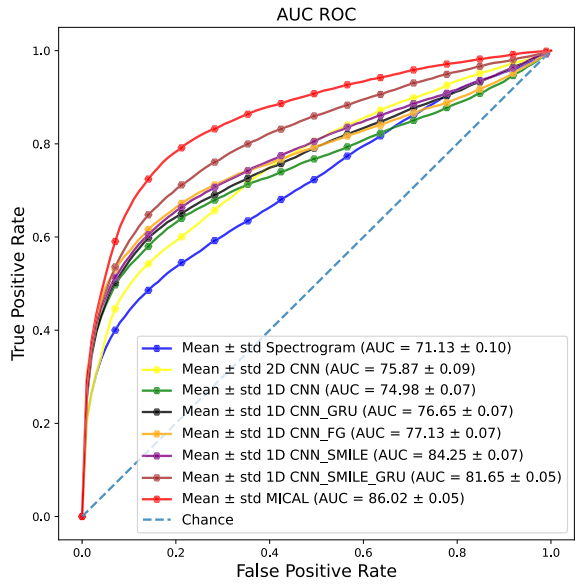


FIGURE 6. AUC-ROC for 6-fold cross validation.

leads to a 5% increase in AUC-ROC, F1 score, and precision. As shown in the table, adding or ignoring the features through the ablation study has not considerably changed the accuracy results. Incorporating temporal correlation for evaluation is conducted via 1D CNN-GRU and 1D CNN-FG. Unlike 1D CNN-GRU that decreases the AUC-PR by 3%, 1D CNN-FG causes 2% improvement for all performance metrics compared to our CNN architecture. This, in fact, implies the strength of the proposed factor graph inference for capturing temporal correlations. The reduction in AUC-PR and F1 score for 1D CNN-SMILE compared to 1D CNN-FG admits that incorporating only raw EEG features and MI estimations is not sufficient for detecting the seizure. Compared to baseline results, this is also proved by MICAL

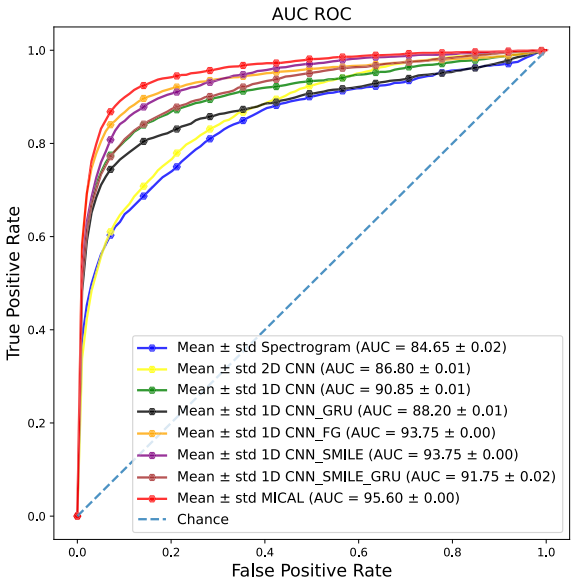


FIGURE 7. AUC-ROC for all patients training.

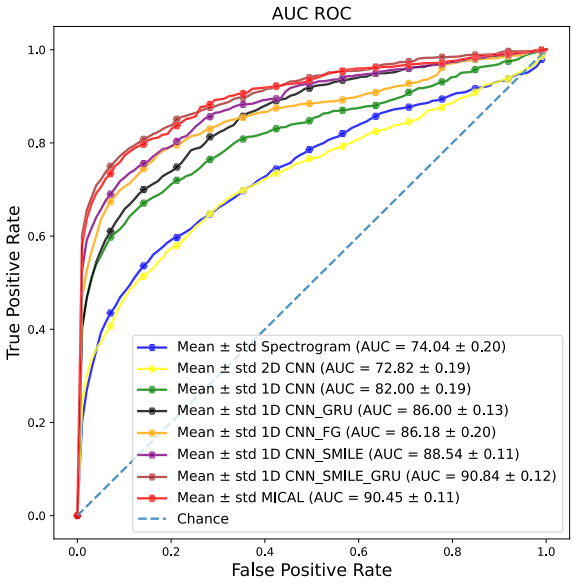


FIGURE 8. AUC-ROC for per patient training.

that it shows overall the best performance results of 15% increase in AUC-ROC and AUC-PR, up to 10% for F1 score and 3% and 20% improvement for precision and recall, respectively. Note that 1D CNN-SMILE-GRU represents a higher precision value than MICAL as there is a trade-off between precision and recall. Therefore, MICAL shows the best evaluation result for recall while it achieves the lowest precision score among other models in this study.

All patient training results in Table 3, Fig 5, and Fig 7 demonstrate that our proposed CNN design achieves almost 15% improvement compared to the baselines. MICAL in comparison to 1D CNN-FG and 1D CNN-SMILE, where only inter-channel or temporal correlation is captured indicates the best results. MICAL enhances the performance by almost 10% for AUC-ROC and 20% for remaining metrics. We also find that adding GRU to the 1D CNN-SMILE leads to decreasing the values for some of the performance metrics. Since the models are training on the limited dataset from only 24 cases, including extra layers could result in overfitting. The results for per patient training in Table 4, and Fig 8 also emphasize that considering both temporal and inter-channel correlations leads to the best performance. Since we average the results across evaluation set from all 24 trained models, and there might be some cases as outliers, GRU-based design works slightly better than MICAL in some of the performance measures.

V. CONCLUSION

In this paper, we have developed MICAL, which is a hybrid model-based/data-driven seizure detection algorithm. MICAL enables capturing two essential features embedded in the EEG recordings: inter-channel dependency during seizures and temporal correlations. The former is extracted through a neural MI estimator, and the latter is achieved via factor graph inference. To implement MICAL, we first carefully design a 1D CNN to extract features from raw EEG signals. Then, soft estimates of common features from CNN and MI estimator are used as the learned factor graph nodes to capture temporal correlation at reduced complexity. In this study, we also conduct a comprehensive evaluation strategy and an ablation study. Although MICAL is not purely based on DL and it is a hybrid and data-efficient approach, it achieves state-of-the-art performance compared to the previous seizure detection studies.

However, there are still a few challenges that need to be addressed. While the neural MI estimator is a powerful tool to capture the inter-channel correlation during seizure times, it increases the computational complexity of the algorithm. Although compared to prior work, our proposed approach generalizes well on the 24-patient dataset, using more patients for training and testing is required for understanding the generalization properties of our algorithm across the population. Therefore, future work will include optimizing the neural MI estimator to address the issue of complexity as well as incorporating large datasets such as TUH from Temple University to improve over-fitting and a better understanding of algorithmic generalization.

ACKNOWLEDGMENT

An earlier version of this paper was presented at the 2022 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) [DOI: 10.1109/ICASSP43922.2022.9746730].

REFERENCES

[1] B. Salafian, E. F. Ben-Knaan, N. Shlezinger, S. D. Ribaupierre, and N. Farsad, "CNN-aided factor graphs with estimated mutual information features for seizure detection," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 8677–8681.

[2] R. S. Fisher, W. V. E. Boas, W. Blume, C. Elger, P. Genton, P. Lee, and J. Engel Jr., "Epileptic seizures and epilepsy: Definitions proposed by the international league against epilepsy (ILAE) and the international bureau for epilepsy (IBE)," *Epilepsia*, vol. 46, no. 4, pp. 470–472, 2005.

[3] *Epilepsy*. [Online]. Available: https://communitymedicine4all.com/2019/02/14/who-updates-fact-sheet-on-epilepsy/

[4] T. T. Sajobi, C. B. Josephson, R. Sawatzky, M. Wang, O. Lawal, S. B. Patten, L. M. Lix, and S. Wiebe, "Quality of life in epilepsy: Same questions, but different meaning to different people," *Epilepsia*, vol. 62, no. 9, pp. 2094–2102, Sep. 2021.

[5] K. Malmgren, M. Reuber, and R. Appleton, "Differential diagnosis of epilepsy," in *Oxford Textbook of Epilepsy and Epileptic Seizures*. Oxford, U.K.: Oxford Univ. Press, 2012, pp. 81–94.

[6] D. K. Pal, A. W. Pong, and W. K. Chung, "Genetic evaluation and counseling for epilepsy," *Nature Rev. Neurol.*, vol. 6, no. 8, pp. 445–453, Aug. 2010.

[7] H. Kandar, S. K. Das, L. Ghosh, and B. K. Gupta, "Epilepsy and its management: A review," *J. Pharma*, vol. 1, no. 2, pp. 20–26, 2012.

[8] N. van Klink, A. Mooij, G. Huiskamp, C. Ferrier, K. Braun, A. Hillebrand, and M. Zijlmans, "Simultaneous MEG and EEG to detect ripples in people with focal epilepsy," *Clin. Neurophysiol.*, vol. 130, no. 7, pp. 1175–1183, Jul. 2019.

[9] T. M. Salmenpera and J. S. Duncan, "Imaging in epilepsy," *J. Neurol., Neurosurg. Psychiatry*, vol. 76, no. 3, pp. 2–10, 2005.

[10] S. S. Spencer, "The relative contributions of MRI, SPECT, and PET imaging in epilepsy," *Epilepsia*, vol. 35, pp. S72–S89, Dec. 1994.

[11] S. Kulaseharan, A. Aminpour, M. Ebrahimi, and E. Widjaja, "Identifying lesions in paediatric epilepsy using morphometric and textural analysis of magnetic resonance images," *NeuroImage, Clin.*, vol. 21, Jan. 2019, Art. no. 101663.

[12] A. Subasi, J. Kevric, and M. A. Canbaz, "Epileptic seizure detection using hybrid machine learning methods," *Neural Comput. Appl.*, vol. 31, no. 1, pp. 317–325, Jan. 2019, doi: 10.1007/s00521-017-3003-y.

[13] J. Yang and M. Sawan, "From seizure detection to smart and fully embedded seizure prediction engine: A review," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 5, pp. 1008–1023, Oct. 2020.

[14] P. van Mierlo, M. Papadopoulou, E. Carrette, P. Boon, S. Vandenberghe, K. Vonck, and D. Marinazzo, "Functional brain connectivity from EEG in epilepsy: Seizure prediction and epileptogenic focus localization," *Prog. Neurobiol.*, vol. 121, pp. 19–35, Oct. 2014.

[15] Ö. Türk and M. S. Özerdem, "Epilepsy detection by using scalogram based convolutional neural network from EEG signals," *Brain Sci.*, vol. 9, no. 5, p. 115, 2019.

[16] U. Asif, S. Roy, J. Tang, and S. Harrer, "SeizureNet: Multi-spectral deep feature learning for seizure type classification," in *Machine Learning in Clinical Neuroimaging and Radiogenomics in Neuro-Oncology*. Peru: Springer, 2020, pp. 77–87.

[17] X. Jiang, G.-B. Bian, and Z. Tian, "Removal of artifacts from EEG signals: A review," *Sensors*, vol. 19, no. 5, p. 987, 2019.

[18] I. Hussain and S. J. Park, "HealthSOS: Real-time health monitoring system for stroke prognostics," *IEEE Access*, vol. 8, pp. 213574–213586, 2020.

[19] I. Hussain and S.-J. Park, "Quantitative evaluation of task-induced neurological outcome after stroke," *Brain Sci.*, vol. 11, no. 7, p. 900, Jul. 2021.

[20] I. Hussain, M. A. Hossain, R. Jany, M. A. Bari, M. Uddin, A. R. M. Kamal, Y. Ku, and J.-S. Kim, "Quantitative evaluation of EEG-biomarkers for prediction of sleep stages," *Sensors*, vol. 22, no. 8, p. 3079, Apr. 2022.

[21] A. T. Tzallas, M. G. Tsipouras, D. G. Tsalikakis, E. C. Karvounis, L. Astrakas, S. Konitsiotis, and M. Tzaphlidou, "Automated epileptic seizure detection methods: A review study," in *Epilepsy: Histological, Electroencephalographic and Psychological Aspects*. Rijeka, Croatia: InTech, 2012, pp. 75–98.

[22] R. Sankar and J. Natour, "Automatic computer analysis of transients in EEG," *Comput. Biol. Med.*, vol. 22, no. 6, pp. 407–422, Nov. 1992.

[23] K. Vijayalakshmi and A. M. Abhishek, "Spike detection in epileptic patients EEG data using template matching technique," *Int. J. Comput. Appl.*, vol. 2, no. 6, pp. 5–8, Jun. 2010.

[24] J. Gotman and P. Gloor, "Automatic recognition and quantification of interictal epileptic activity in the human scalp EEG," *Electroencephalogr. Clin. Neurophysiol.*, vol. 41, no. 5, pp. 513–529, Nov. 1976.

[25] J. Gotman, "Automatic recognition of epileptic seizures in the EEG," *Electroencephalogr. Clin. Neurophysiol.*, vol. 54, no. 5, pp. 530–540, 1982.

[26] T. P. Exarchos, A. T. Tzallas, D. I. Fotiadis, S. Konitsiotis, and S. Giannopoulos, "EEG transient event detection and classification using association rules," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 3, pp. 451–457, Jul. 2006.

[27] K. P. Indiradevi, E. Elias, P. Sathidevi, S. D. Nayak, and K. Radhakrishnan, "A multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram," *Comput. Biol. Med.*, vol. 38, no. 7, pp. 805–816, 2008.

[28] C.-W. Ko and H.-W. Chung, "Automatic spike detection via an artificial neural network using raw EEG data: Effects of data preparation and implications in the limitations of online recognition," *Clin. Neurophysiol.*, vol. 111, no. 3, pp. 477–481, Mar. 2000.

[29] A. J. Gabor and M. Seyal, "Automated interictal EEG spike detection using artificial neural networks," *Electroencephalogr. Clin. Neurophysiol.*, vol. 83, no. 5, pp. 271–280, Nov. 1992.

[30] N. Acir, I. Oztura, M. Kuntalp, B. Baklan, and C. Guzelis, "Automatic detection of epileptiform events in EEG by a three-stage procedure based on artificial neural networks," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 1, pp. 30–40, Jan. 2005.

[31] R. San-Segundo, M. Gil-Martín, L. F. D'Haro-Enríquez, and J. M. Pardo, "Classification of epileptic EEG recordings using signal transforms and convolutional neural networks," *Comput. Biol. Med.*, vol. 109, pp. 148–158, Jun. 2019.

[32] R. Akut, "Wavelet based deep learning approach for epilepsy detection," *Health Inf. Sci. Syst.*, vol. 7, no. 1, pp. 1–9, Dec. 2019.

[33] X. Chen, J. Ji, T. Ji, and P. Li, "Cost-sensitive deep active learning for epileptic seizure detection," in *Proc. ACM Int. Conf. Bioinf., Comput. Biol., Health Informat.*, Aug. 2018, pp. 226–235.

[34] K. Fukumori, H. T. T. Nguyen, N. Yoshida, and T. Tanaka, "Fully data-driven convolutional filters with deep learning models for epileptic spike detection," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 2772–2776.

[35] S. B. Wilson and R. Emerson, "Spike detection: A review and comparison of algorithms," *Clin. Neurophysiol.*, vol. 113, no. 12, pp. 1873–1881, Dec. 2002.

[36] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[37] T. Kim, P. Nguyen, N. Pham, N. Bui, H. Truong, S. Ha, and T. Vu, "Epileptic seizure detection and experimental treatment: A review," *Frontiers Neurol.*, vol. 11, p. 701, Jul. 2020.

[38] 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. (ICASSP)*, Mar. 2016, pp. 754–758.

[39] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Comput. Biol. Med.*, vol. 100, pp. 270–278, Sep. 2018.

[40] J. Liu and B. Woodson, "Deep learning classification for epilepsy detection using a single channel electroencephalography (EEG)," in *Proc. 3rd Int. Conf. Deep Learn. Technol.*, Jul. 2019, pp. 23–26.

[41] P. Boonyakitanont, A. Lek-Uthai, K. Chomtho, and J. Songsiri, "A comparison of deep neural networks for seizure detection in EEG signals," *bioRxiv*, Jul. 2019, Art. no. 702654.

[42] L. Vidyaratne, A. Glandon, M. Alam, and K. M. Iftekharuddin, "Deep recurrent neural network for seizure detection," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 1202–1207.

[43] X. Yao, Q. Cheng, and G.-Q. Zhang, "Automated classification of seizures against nonseizures: A deep learning approach," 2019, *arXiv:1906.02745*.

[44] R. Hussein, H. Palangi, Z. J. Wang, and R. Ward, "Robust detection of epileptic seizures using deep neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 2546–2550.

[45] D. Ahmedt-Aristizabal, C. Fookes, K. Nguyen, and S. Sridharan, "Deep classification of epileptic signals," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 332–335.

[46] S. S. Talathi, "Deep recurrent neural networks for seizure detection and early seizure detection systems," 2017, *arXiv:1706.03283*.

[47] S. Roy, I. Kiral-Kornek, and S. Harrer, "ChronoNet: A deep recurrent neural network for abnormal EEG identification," in *Proc. Conf. Artif. Intell. Med. Eur.* Springer, 2019, pp. 47–56.

[48] N. Shlezinger, J. Whang, Y. C. Eldar, and A. G. Dimakis, "Model-based deep learning," 2020, *arXiv:2012.08405*.

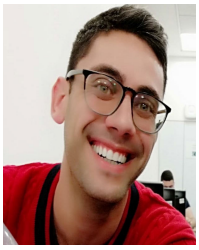
[49] M. A. B. Brazier, "Spread of seizure discharges in epilepsy: Anatomical and electrophysiological considerations," *Exp. Neurol.*, vol. 36, no. 2, pp. 263–272, Aug. 1972.

[50] A. Quintero-Rincón, M. Pereyra, C. D'Giano, H. Batatia, and M. Risk, "A new algorithm for epilepsy seizure onset detection and spread estimation from EEG signals," *J. Phys., Conf. Ser.*, vol. 705, Apr. 2016, Art. no. 012032.

[51] N. Shlezinger, N. Farsad, Y. C. Eldar, and A. J. Goldsmith, "Learned factor graphs for inference from stationary time sequences," *IEEE Trans. Signal Process.*, vol. 70, pp. 366–380, 2022.

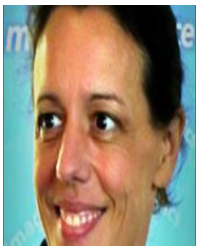
[52] P. D. Emmady and A. C. Anilkumar, *EEG Abnormal Waveforms*. St. Petersburg, FL, USA: StatPearls, 2020.
[53] I. Jemal, A. Mitiche, and N. Mezghani, "A study of EEG feature complexity in epileptic seizure prediction," *Appl. Sci.*, vol. 11, no. 4, p. 1579, Feb. 2021.
[54] M. Lee, I. Youn, J. Ryu, and D.-H. Kim, "Classification of both seizure and non-seizure based on EEG signals using hidden Markov model," in *Proc. IEEE Int. Conf. Big Data Smart Comput. (BigComp)*, Jan. 2018, pp. 469–474.
[55] S. Kullback, *Information Theory and Statistics*. Chelmsford, MA, USA: Courier Corporation, 1997.
[56] L. Paninski, "Estimation of entropy and mutual information," *Neural Comput.*, vol. 15, no. 6, pp. 1191–1253, 2003.
[57] J. Song and S. Ermon, "Understanding the limitations of variational mutual information estimators," 2019, *arXiv:1910.06222*.
[58] M. I. Belghazi, A. Baratin, S. Rajeshwar, S. Ozair, Y. Bengio, A. Courville, and D. Hjelm, "Mutual information neural estimation," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 531–540.
[59] M. D. Donsker and S. R. S. Varadhan, "Asymptotic evaluation of certain Markov process expectations for large time, I," *Commun. Pure Appl. Math.*, vol. 28, no. 1, pp. 1–47, 1975.
[60] T. S. Zarghami, G.-A. Hossein-Zadeh, and F. Bahrami, "Deep temporal organization of fMRI phase synchrony modes promotes large-scale disconnection in schizophrenia," *Frontiers Neurosci.*, vol. 14, p. 214, Mar. 2020.
[61] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise Reduction in Speech Processing*. Berlin, Germany: Springer, 2009, pp. 1–4.
[62] D. Edelmann, T. F. Móri, and G. J. Székely, "On relationships between the Pearson and the distance correlation coefficients," *Statist. Probab. Lett.*, vol. 169, Feb. 2021, Art. no. 108960.
[63] N. Shlezinger, N. Farsad, Y. C. Eldar, and A. J. Goldsmith, "Learned factor graphs for inference from stationary time sequences," 2020, *arXiv:2006.03258*.
[64] N. Shlezinger, N. Farsad, Y. C. Eldar, and A. J. Goldsmith, "Data-driven factor graphs for deep symbol detection," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2020, pp. 2682–2687.
[65] H.-A. Loeliger, "An introduction to factor graphs," *IEEE Signal Process. Mag.*, vol. 21, no. 1, pp. 28–41, Jan. 2004.
[66] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
[67] F. Mormann, T. Kreuz, R. G. Andrzejak, P. David, K. Lehnertz, and C. E. Elger, "Epileptic seizures are preceded by a decrease in synchronization," *Epilepsy Res.*, vol. 53, no. 3, pp. 173–185, Mar. 2003.
[68] A. L. Goldberger, L. A. N. 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: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, Jun. 2000.
[69] G. C. Jana, R. Sharma, and A. Agrawal, "A 1D-CNN-Spectrogram based approach for seizure detection from EEG signal," *Proc. Comput. Sci.*, vol. 167, pp. 403–412, Jan. 2020.

BAHAREH SALAFIAN was born in Isfahan, Iran, in 1996. She received the B.S. degree in electrical engineering from the Isfahan University of Technology, Isfahan, in 2018, and the M.Sc. degree in biomedical engineering from Western University, London, ON, Canada, in 2021. She is currently a Research Assistant with the Learning and Inference Algorithms (LIA) Research Group, Toronto Metropolitan University, Toronto, ON, Canada, where her work focuses on developing automated approach for epileptic seizure detection and prediction. She was also the Data Scientist Intern at Legion Technologies, Redwood City, CA, USA, where she worked on implementing time series forecasting models for labor optimization and demand prediction.



EYAL FISHEL BEN-KNAAN received the M.Sc. degree in physics from the Ben-Gurion University of the Negev, Israel, in 2021, where he is currently pursuing the Ph.D. degree with the School of Electrical and Computer Engineering.

NIR SHLEZINGER (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in electrical and computer engineering from the Ben-Gurion University of the Negev, Israel, in 2011, 2013, and 2017, respectively. He is currently an Assistant Professor with the School of Electrical and Computer Engineering, Ben-Gurion University of the Negev. From 2017 to 2019, he was a Postdoctoral Researcher with Technion. From 2019 to 2020, he was a Postdoctoral Researcher with the Weizmann Institute of Science, where he was awarded the FGS Prize for Outstanding Research Achievements. His research interests include communications, information theory, signal processing, and machine learning.



SANDRINE DE RIBAUPIERRE received the M.D. degree from the University of Geneva, Switzerland. After a Neurosurgery Residency in Lausanne, Switzerland, she completed an Epilepsy Fellowship with Fondation Rothschild, Paris, France, and a Pediatric Neurosurgery Fellowship with the Hospital for Sick Children, Toronto. She is currently a Professor in clinical neurological sciences with Western University, where she is also working as a Pediatric Neurosurgeon with a subspecialty in epilepsy surgery. She holds a research chair in paediatric neurosurgery and neurosciences. Her research interest includes neurodevelopment using functional imaging particularly in children at risk, such as preterm infants or children with early onset of neurological conditions, such as epilepsy or hydrocephalus.

NARIMAN FARSAD (Senior Member, IEEE) was a Senior Machine Learning and Algorithms Researcher with Apple Inc. From 2015 to 2018 and 2018 to 2020, he was a Postdoctoral Research Fellow and a Visiting Research Fellow with Stanford University, respectively, where he was a recipient of the Natural Sciences and Engineering Research Council of Canada Postdoctoral Fellowship. He is currently an Assistant Professor with the Department of Computer Science, Ryerson University, Toronto, Canada. His research laboratory, Learning and Inference Algorithms (LIA) Laboratory, focuses on the general areas of signal processing, machine learning, and information theory. Specifically, the group investigates problems within biomedical sensing, multimodal sensing, and AR/VR. He has received the Second Prize from the 2014 IEEE ComSoc Student Competition, has received the Best Demo Award at INFOCOM2015, and was recognized as a Finalist for the 2014 Bell Laboratories Prize.