

Biomimetic Deep Learning Networks With Applications to Epileptic Spasms and Seizure Prediction

Christopher Lucasius^{ID}, Vasily Grigorovsky^{ID}, Member, IEEE, Hiroki Nariai^{ID}, Aristea S. Galanopoulou^{ID}, Jonathan Gursky^{ID}, Solomon L. Moshé^{ID}, and Berj L. Bardakjian^{ID}, Life Member, IEEE

Abstract—Objective: In this study, we present a novel biomimetic deep learning network for epileptic spasms and seizure prediction and compare its performance with state-of-the-art conventional machine learning models. Methods: Our proposed model incorporates modular Volterra kernel convolutional networks and bidirectional recurrent networks in combination with the phase amplitude cross-frequency coupling features derived from scalp EEG. They are applied to the standard CHB-MIT dataset containing focal epilepsy episodes as well as two other datasets from the Montefiore Medical Center and the University of California Los Angeles that provide data of patients experiencing infantile spasm (IS) syndrome. Results: Overall, in this study, the networks can produce accurate predictions (100%) and significant detection latencies (10 min). Furthermore, the biomimetic network outperforms conventional ones by producing no false positives. Significance: Biomimetic neural networks utilize extensive knowledge about processing and learning in the electrical networks of the brain. Predicting seizures in adults can improve their quality of life. Epileptic spasms in infants are part of a particular seizure type that needs identifying when suspicious behaviors are noticed in babies. Predicting epileptic spasms within a

given time frame (the prediction horizon) suggests their existence and allows an epileptologist to flag an EEG trace for future review.

Index Terms—Biomimetic neural networks, deep learning, principal dynamic modes, bidirectional recurrent networks, electroencephalogram, infantile spasms.

I. INTRODUCTION

OME of the most common building blocks of current deep learning networks have biological correlates. Individual units of multilayer perceptrons, for example, behave as simplified point neurons; convolutional neural networks (CNN), on the other hand, show similarity in their organization to the visual cortex [1]. However, these, and other neural network structures often fail to incorporate complex, yet vital, interactions observed in biological neural networks. One such fundamental complexity is neural coding which is observed in the brain's networks. One example of such coding is the coupling between different frequencies of the electroencephalogram (EEG) and its role in memory consolidation [2]. A more fundamental example relevant to this work is the observation that neuron firing patterns in the brain directly correlate with macroscopic activity recorded in cats [3].

Advances in machine learning algorithms have highlighted them as a potential tool for early seizure detection in patients with epilepsy. These algorithms have been used with EEG data for detection and prediction of seizure states before they become visually apparent [4]. One of the measures of success for an automated seizure prediction system (SPS) is detection latency. In order to capture the neural coding from a feature engineering approach, an index of global phase-amplitude cross-frequency coupling (GPAC) has been used to (a) localize epileptogenic zones for surgical resections in patients with drug resistant epilepsy [5], and (b) predict seizure states [6], [7], [8], by identifying increased coupling between the delta and high frequency oscillation (HFO) ranges. Our lab has previously published encouraging preliminary results using GPACs as input features to random forest classifiers, to predict seizures in 12 adult patients with drug resistant epilepsy [6], [7]. Since GPACs can be depicted as pixel heatmap images, a CNN is an appropriate choice for the task and can replace simpler classifiers such as support vector machines and random forests.

Manuscript received 17 February 2023; revised 9 June 2023 and 29 August 2023; accepted 9 October 2023. Date of publication 18 October 2023; date of current version 26 February 2024. The work of Aristea S. Galanopoulou was supported by NIH under Grants U54 NS100064 and R01 NS091170, in part by the U.S. Department of Defense under Grant W81XWH-18-1-0612, in part by the American Epilepsy Society seed grant, in part by the Heffer Family and the Segal Family Foundations, and in part by the Abbe Goldstein/Joshua Lurie and Laurie Marsh/Dan Levitz families. The work of Solomon L. Moshé was supported in part by NIH under Grants U54 NS100064 and NS43209, in part by the U.S. Department of Defense under Grant W81XWH-18-1-0612, in part by the Heffer Family and the Segal Family Foundations, and in part by the Abbe Goldstein/Joshua Lurie and Laurie Marsh/Dan Levitz families. Solomon L. Moshé is the Charles Frost Chair in Neurosurgery and Neurology. (*Corresponding author: Berj L. Bardakjian*.)

Christopher Lucasius and Vasily Grigorovsky are with the Department of Electrical and Computer Engineering, University of Toronto, Canada.

Hiroki Nariai is with the Department of Pediatrics, UCLA Mattel Children's Hospital, USA.

Aristea S. Galanopoulou, Jonathan Gursky, and Solomon L. Moshé are with the Department of Neurology and the Comprehensive, Einstein/Montefiore Epilepsy Center, Albert Einstein College of Medicine, USA.

Berj L. Bardakjian is with the Institute of Biomedical Engineering, University of Toronto, Toronto, ON M5S 3G9, Canada (e-mail: berj.bardakjian@utoronto.ca).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TBME.2023.3325762>, provided by the authors.

Digital Object Identifier 10.1109/TBME.2023.3325762

Seizure state evolution is a dynamic activity, and an algorithm that is sensitive to a temporal aspect makes a good choice for machine learning. Recurrent neural networks (RNNs) such as the long short-term memory cell (LSTM) learn and extract meanings in time-series data by keeping track of a hidden state [9], [10]. The LSTM uses several additional feedback weights to regularize the gradients [11], [12]. These algorithms have been used in seizure prediction using simpler EEG signal processing methods than phase amplitude cross-frequency coupling (PAC) analyses [13]. Although this work produced promising results with LSTMs, it still gave several false positives among its subjects.

In this study, in addition to using feature engineering to capture neuronal coding, we have incorporated biological codes directly into a CNN-based deep learning network. We have achieved this by adding modular Volterra filter based CNN to a bi-directional LSTM (MVNN-BiLSTM), which we then compared with a standard CNN-LSTM configuration. This novel network was applied to the problem of epileptic seizure prediction, both in adults and babies with infantile spasms (IS). Being able to predict seizures in adults is important because this can allow them to improve their quality of life by taking appropriate precautions. Regarding IS, approximately one in every 2500 babies develop this syndrome and 90% of them develop IS within the first year of life [14]. It is important to quickly diagnose their spasms since it is quite plausible for an epileptologist to miss a spasm given all of the EEGs they must review. By utilizing a model that can predict a spasm within a reasonable timeframe (the prediction horizon), the system can automatically mark traces containing spasms for future review by the epileptologist.

Epileptic spasms in IS can be subtle, and they are a particular seizure type that needs identifying when suspicious behaviors are noticed in babies. Clinicians have difficulty in determining if a seizure has occurred when they only observe the EEG traces. Consequently, a tool that accurately predicts epileptic spasm events based solely on EEG could support medical professionals in making timely IS diagnoses, potentially enhancing an infant's clinical outcomes through early intervention. Furthermore, algorithms that predict EEG-captured epileptic spasms with high accuracy and specificity will be an essential unbiased tool in clinical trials where the effects of treatments on spasms are being tested.

II. RELATED WORK

A. Convolutional and Recurrent Neural Networks

Two key aspects of EEG data are that it is spatially distributed and evolves in time. This makes the CNN and RNN obvious choices for machine learning tool design. CNNs use learnable kernels which take advantage of spatial patterns to extract meaningful features from images. RNNs, on the other hand, use temporal information to make a classification decision. Both CNNs [4], [15], [16] and LSTMs [13] have been used in seizure prediction.

In a CNN-LSTM model, a CNN would reduce input images into an embedding representation which is then fed into

a network of LSTM cells to analyze the images' temporal variation. Previous work used publicly available ictal and interictal datasets (e.g., CHB MIT) to evaluate these networks in seizure prediction problems [13], [15], [17]. Another study used a CNN-LSTM to analyze two-dimensional images of EEG signals [18]. However, since EEG signal amplitude is prone to measurement artifact, the interaction among brain rhythms is more closely associated with the coding mechanisms of the brain. This makes PAC an excellent choice for use as an input feature to the CNN-LSTM model for biomimetic learning.

B. Biological Inspirations

Over the past decade, deep learning neural networks have made significant contributions in a variety of classification and prediction problems, such as decoding and processing complex, hierarchical datasets. Artificial neural networks (ANNs) are largely neural-inspired, drawing parallels in their hierarchical structure, spatiotemporal integration, and activation thresholds [19]. However, they still tend to perform poorly in higher level tasks that require reasoning [20]. Given their limitations in performance compared to the adaptive nature of the human brain, it is desirable to apply biological learning and coding mechanisms to deep learning models.

Several works by neural engineers proposed successful models to capture the dynamics of nonlinear physiological systems such as the brain [21], [22]. They made use of principal dynamic modes (PDMs) to break down the complexities after modelling them as input-output Volterra systems. Previous studies have succeeded in applying this technique to model physiological phenomena such as the effects of infusing insulin on glucose [23], and pathology characterization in rodent epilepsy models [24]. The latter study identified both differentiating and integrating PDMs in low magnesium high potassium models of epilepsy.

Our group has previously used PDMs as inputs to an oscillator-based model of epilepsy to simulate neuronal dynamics for neural coding purposes [25], neuromodulation for the prevention of seizure-like events [26], and hyperexcitable neuroglial networks [27]. These models produced similar complex responses to those seen in real neurons and neural assemblies, highlighting their accuracies in capturing the complex processes that occur within brain networks. Therefore, incorporating these models in existing deep learning networks can be a key step towards making these networks more biomimetic. In a previous study [28], it was shown that a Volterra Series with an arbitrary degree can be equivalent to an ANN with polynomial activations. This suggests that a well-established solution to dynamical systems can be modelled by a neural network assuming that its kernels match those of the Volterra model. As explained in [25], neuronal modes can be shown to have differentiating and integrating functionalities indicating how a deep learning network model can be modified to be more biomimetic.

We aim to embed neural coding features directly into our proposed model and take advantage of the spatio-temporal benefits provided by combining the CNN and LSTM architectures.

TABLE IA
TRAINING DATA

Subject Label	Sampling Rate (Hz)	Training			
		Interictal		Ictal	
		Data (s)	Number of Windows	Data (s)	Number of Windows
1	1024	415	264-810	127	72-234
2	1024	1797	1185-3574	799	520-1578
3	1024	0	0	1005	657-1990
8	1024	1197	785-2374	227	139-434
9	1024	316	198-612	1797	1185-3574
12	2000	2100	694-2090	2100	694-2090
13	2000	1620	534-1610	1620	534-1610
4	2000	0	0	269	84-259
5	2000	0	0	1887	623-1877
6	2000	0	0	3113	1032-3103
Totals		7445	11098	12944	5540-16749
Control Subjects					
		Data (s)	Number of Windows		
17	2000	1305	429-1295		
18	2000	2031	671-2021		
Totals		3336	1100-3316		

The aim is to combine the classical neural network structures as described in II.A with biomimetic ones described in II.B.

III. METHODS

A. Data Acquisition

The study population consisted of 14 infants (4.8-13.7 months old) who had IS: seven from the Montefiore Medical Center (Montefiore) and seven from the University of California Los Angeles (UCLA). Inclusion criteria were subjects with epileptic spasms with scalp EEGs that demonstrated electrodecremental responses, hypsarrhythmia or its variants, with or without psychomotor delay or arrest at presentation, and had high quality scalp EEG and electromyogram (EMG) measurements. The control group included scalp EEG recordings from four babies without IS/seizures (3-6 months old). Two neurologists board certified in clinical neurophysiology confirmed the presence of spasms in IS subjects or movements associated with arousals in control subjects in video recordings, and the onsets were marked by EMG deflections. The study was approved by the Institutional Review Boards at Montefiore (IRB#13-02-073, initial approval date 02/01/2016) and UCLA Medical Centers (IRB#18-001599, initial approval date 11/02/2018).

Scalp EEG recordings obtained for clinical indication were used. At Montefiore, an XLTEK EEG acquisition system (Natus Medical Inc., Pleasanton, CA) with a digital sampling frequency of 1024 Hz and a high pass filter at 0.05 Hz was used. At the UCLA, A Nihon Kohden system (Irvine, California, USA) with a digital sampling frequency of 2000 Hz, which defaulted to a proprietary setting of a low frequency filter of 0.016 Hz and a high frequency filter of 600 Hz, provided EEG acquisition. For each patient, scalp EEG signals were obtained from 19 electrodes placed according to the International 10-20 system, including Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, Fz, Cz, and Pz as well as an EMG lead over the anterior chest wall. The common reference (Ref) was the FCz electrode in the UCLA EEGs and the midpoint of either the Fz-C3 or Fz-C4 electrodes in the Montefiore EEGs. Recordings were grounded at Fpz.

TABLE IB
TESTING DATA

Subject Label	Sampling Rate (Hz)	Testing			
		Interictal		Ictal	
		Data (s)	Number of Windows	Data (s)	Number of Windows
10	1024	496	318-972	98	53-176
11	1024	1797	1185-3574	387	245-754
7	2000	0	0	1952	645-1942
14	2000	269	84-259	269	84-259
Totals		2562	4819	2706	1027-3131
Control Subjects					
		Data (s)	Number of Windows		
15	2000	2270	751-2260		
16	2000	1687	556-1677		
Totals		3957	1307-3937		

Both at Montefiore and UCLA, continuous EEG data from each subject were clipped and merged to form a single EEG file containing several representative interictal or normal background segments (including wakefulness, drowsiness, and arousal) and at least one ictal EEG segment with spasms (IS) and/or arousal per subject. Interictal EEGs from IS subjects were at least one hour away from a spasm cluster. The selected segments containing spasms included up to 52 minutes of recordings before the spasms in IS subjects. The clipped EEG data were analyzed without any clinical information known in advance. A total of 16 video EEGs containing spasms (first spasm of a cluster or individual spasms) and four arousals in control subjects were collected for this study, using 10.25h as shown in Table IA and IB). For further analyses, a finite impulse response 4000th order notch filter was applied to cutoff frequencies of 60 Hz and its harmonics.

B. Seizure Prediction System (SPS)

Our proposed SPS is depicted in Fig. 1(a). It includes a pre-processing stage, a machine learning processor, a refractory threshold trigger (RTT), and a post-processing stage.

C. Pre-Processing Stage

PAC features from individual EEG channels were extracted from 5s, 10s, and 15s windows of scalp EEG by applying the PyWavelets toolbox in Python [29]. The window size was kept consistent between training and testing, (e.g., the classifier trained on 10s windows was tested on 10s ones).

The EEG was first de-noised by removing the 60 Hz line noise along with its harmonics. The PAC computation was then carried out by first using the continuous wavelet transform (CWT) to extract time-frequency spectrograms of the de-noised EEG. Note that this method was used as an alternative to the Hilbert-Huang Transform and Empirical Mode Decomposition as described by Tort et al.'s computation of the PAC [30] in order to avoid the assumption of narrowband signals [31]. A complex Morlet wavelet was used as the mother wavelet to extract the time-varying PAC as similarly done in previous work [6], [7]. The computation of the CWT was performed as follows:

$$W(s, t) = \int_{\tau} x(\tau) \psi_{s,t}^*(\tau) d\tau \quad (1)$$

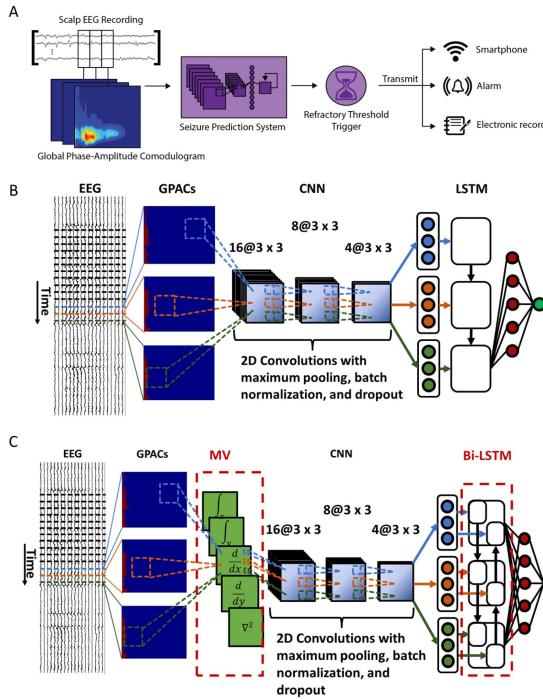


Fig. 1. (a) System diagram of Seizure Prediction System. (b) CNN-LSTM architecture. (c) MVNN-BILSTM architecture. Similar to the CNN-LSTM with the addition of two components, the modular volterra (MV) layer consists of a filter bank of differentiating and integrating kernels, similar to the PDMs observed in the brain. The kernels operate in the horizontal and vertical directions in image space. The LSTM is also modified to mimic bidirectional communication more closely in biological neural networks.

where τ varies across the given window, $\psi_{s,t}^*(\tau)$ is the conjugate of the Morlet basis function, $x(t)$ is the input EEG signal, and $W(s,t)$ is the CWT defined in the s and t domain. The Morlet basis function is given by:

$$\psi_{s,t}^*(\tau) = \frac{1}{\sqrt{s}} \psi_0 \left(\frac{t - \tau}{s} \right) \quad (2)$$

where ψ_0 is the Mother wavelet:

$$\psi_0(t) = \frac{1}{\sqrt{\pi f_b}} e^{j2\pi f_c t - \frac{t^2}{f_b}} \quad (3)$$

where the centre frequency is $f_c = 0.8125$ Hz and the time-decay parameter is $f_b = 6$. The desired frequencies to analyze are converted to CWT scales by normalizing them based on the centre and sampling frequencies. Since calculating the CWT coefficients through direct integration can be extremely time consuming and hence impractical to calculate the PACs, a Fast Fourier Transform was used to compute the convolution by first multiplying in the frequency domain [32]. This effectively improved the time complexity from $O(n^2)$ to $O(n \log n)$ where n is the size of the window. The complex valued CWT coefficients are given by:

$$W(f, t) = w(f, t) + j\tilde{w}(f, t) \quad (4)$$

where w and \tilde{w} represent the real and imaginary parts of the wavelet function's coefficients, respectively. The frequencies chosen for this work are included in the low range $f_L \in$

$\{1, 1.1, \dots, 10\}$ Hz and high range $f_H \in \{20, 21, \dots, 150\}$ Hz. The lower frequencies correspond to the delta and theta bands while the higher ones correspond to the HFO bands; the resulting PAC covers the delta-HFO range. The CWTs were computed with a buffer of 2.873s on both sides of each window to consider the edge effects at low frequencies.

The amplitude $A(t, f_H)$ and phase $\phi(t, f_L)$ signals were computed using the following formulas:

$$A(t, f_H) = \sqrt{w^2(t, f_H) + \tilde{w}^2(t, f_H)} \quad (5)$$

$$\phi(t, f_L) = \arctan \frac{\tilde{w}(t, f_L)}{w(t, f_L)} \quad (6)$$

The PAC is computed from these signals by binning the sums of the amplitudes across the given time window over the phases. The phases are divided into $N = 18$ bins of 20° each, and the amplitudes are normalized over the sum of all bins to convert them into probability values:

$$p_j(t, f_L, f_H) = \frac{\sum_{\tau=1}^T A_{j,t}(\tau, f_H)}{\sum_{k=1}^N \sum_{\tau=1}^T A_{k,t}(\tau, f_H)} \quad (7)$$

where τ is varied across the window of length T and t represents the time point of the beginning of the given window. Each p_j represents a probability value of the j^{th} bin. The probability distribution is then used to compute an entropy measure $H(t, f_L, f_H)$:

$$H(t, f_L, f_H) = - \sum_{j=1}^N p_j(t, f_L, f_H) \log(p_j(t, f_L, f_H)) \quad (8)$$

The PACs are computed by normalizing the entropy measures with respect to the maximum possible entropy given by a uniform distribution with an entropy of $H_{\max} = \log N$:

$$PAC(t, f_L, f_H) = \frac{H_{\max} - H(t, f_L, f_H)}{H_{\max}} \quad (9)$$

Each of the above PACs are computed for each of the 19 individual EEG channels described in Section III-A. Since this work is not informed of the most relevant channels to analyze, an average is taken of all channel PACs, giving one global PAC (GPAC) that varies in time. This method of averaging channels was also successfully demonstrated in previous work on seizure prediction [6].

To reduce the computational cost at training and testing time, the GPACs are binarized at thresholds P_{GPAC} varying from 0% to 100% of their maximum indices (e.g., at a threshold of 50%, any pixel values larger than 50% of the maximum value are set to 1 and the rest are set to 0). The value of P_{GPAC} was selected during cross validation while training the neural networks that made use of the GPACs as input features.

D. Machine Learning Processor: CNN-LSTM Architecture

This work uses a CNN-LSTM architecture (Fig. 1(b)) to process a temporal pattern of GPACs for seizure prediction. It takes as input 91x131 GPAC images where each pixel corresponds to a high and low frequency pair. The CNN consists of three

maximum pooled two-dimensional convolutional layers with 3 x 3 filters (16, 8, then 4 as shown in Fig. 1). The output of the final filter is flattened into an embedding layer which is then fed into a 256-dimensional LSTM network for temporal (sequence length 3) analysis. The GPAC images are distributed over time in the network where three sequential images (one word) are processed by the LSTM layer. The result is then fed into a fully connected layer of 100 units. The outputs of this layer are then used to compute the probability of an incoming spasm with a logistic activation function.

In previous work on seizure prediction in adults with epilepsy that used a random forest in a multistage state classifier [7], three windows were used to classify seizure states. To not diverge too greatly from this relatively successful model, the CNN-LSTM is constrained to analyze three temporal windows at a time. To increase the amount of training data for the CNN-LSTM, an overlap of at least 80% was used when computing GPACs in both UCLA and Montefiore subjects. This overlap was preserved in the testing phase of the CNN-LSTM.

In subsequent sections, each GPAC window is named as a “syllable”, each temporal sequence of three GPACs which are fed into the CNN-LSTM as a single data point is considered as a “word”, and each string of these words is considered as a “sentence”. The word length was chosen as three syllables, based on the number of GPACs that were analyzed in previous work on seizure prediction in IS [7]. The purpose of the sentence is to smooth out erratic predictions by the CNN-LSTM where a specific number of words must give a sufficient probability that is higher than a decision threshold T (determined at testing time). The sentence length was kept constant during the analysis. After a positive prediction is made, the sentences are analyzed immediately after the last positive one, meaning there is no overlap between sentences as the algorithm sweeps the segments of GPACs.

E. Machine Learning Processor: MVNN-BiLSTM Architecture

The CNN-LSTM architecture is further modified by adding filters that are more akin to the PDMs introduced by Marmarelis. As previously discussed, [22] and [24] proposed various methods to characterize nonlinear physiological systems in terms of their PDMs. In this work, their findings are integrated into the CNN-LSTM to create a more biomimetic network. In particular, the differentiating and integrating modes from [24] are added as a modular Volterra (MV) input layer preceding the CNN-LSTM.

Another biological inspiration is added by making the LSTM bidirectional. This is representative of gap junctions that are prevalent in glia, neurons, cardiac cells and other excitable cells [33], [34], [35].

The resulting MVNN-BiLSTM (Fig. 1(c)) is similar to the regular CNN-LSTM except for the additional PDM-like filter bank at the beginning which only focuses on differentiation and integration in image space as well as an additional hidden layer preceding the final probability output to increase the complexity of the network. The functions of the PDM-like filter bank include first-order differentiation (both horizontal and vertical in image

space), integration (horizontal and vertical), and second-order differentiation. The first-order derivatives are implemented with the Sobel operator, the integration filters with a cumulative sum function, and the second-order derivatives with the Laplacian operator. To allow the network to learn using these customized filters, additional weights and biases a_i and b_i , $i \in \{1, 2, 3, 4, 5\}$ are added to these operators as shown below:

$$\int_x \text{GPAC}(x, y) dx = a_1 \sum_{i \leq x} \text{GPAC}(i, y) + b_1 \quad (10)$$

$$\int_y \text{GPAC}(x, y) dy = a_2 \sum_{j \leq y} \text{GPAC}(x, j) + b_2 \quad (11)$$

$$\frac{d}{dx} (\text{GPAC}) = \text{GPAC} * a_3 \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} + b_3 \quad (12)$$

$$\frac{d}{dy} (\text{GPAC}) = \text{GPAC} * a_4 \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} + b_4 \quad (13)$$

$$\nabla^2 (\text{GPAC}) = \text{GPAC} * a_5 \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} + b_5 \quad (14)$$

F. Refractory Threshold Trigger (RTT)

The RTT produces an output signal when triggered followed by a refractory interval equal to the duration of a cluster of spasms representing a seizure episode to avoid repetitive triggers for each spasm within the cluster. It resembles an “hourglass ring device” [36] in that its states constitute a ring; it runs on a cycle of resting → excitation → refractoriness → resting → stop. It needs to be “triggered” to produce an output since the first and last state is a resting state. Its mathematical expression is based on the “Threshold Trigger” [37], [38]:

$$Y_{TT}(n) = 0.5 + 0.5 \text{sgn}(X_{TT} - T) \quad (15)$$

If $Y_{TT}(n) = 1$, then $Y_{TT}(n+m) = 0$, for $m = 1, \dots, t_r$, where $X_{TT}(\cdot)$ and $Y_{TT}(\cdot)$ are the input and output of the RTT, T is the decision threshold, and t_r is the refractory period.

G. Training and Testing of CNN-LSTM and MVNN-BiLSTM

We simplified the data separation by classifying EEG segments as either ictal (preceding a spasm onset within 1h) or not. It used both the interictal (over 1h away from a spasm onset)/control (non-IS) segments and ictal segments in training subjects. The interictal segments were defined to take place in babies with IS more than 1h before or after a spasm cluster, and the control segments were defined as the EEGs of babies without IS/seizures. Both interictal and control segments were collectively labelled as “non-ictal”. A seizure prediction horizon was defined as within 1h of the first spasm onset. This duration was chosen based on previous work in seizure prediction [39], and it was necessary to keep this horizon consistent for the

formulation of the seizure prediction problem. It should be noted that this work defines prediction as a classification of a segment as either non-ictal or ictal.

These segments were used to train the network to recognize whether the analyzed EEG had no spasms (the EEG being analyzed was a non-ictal segment) or an imminent spasm within the seizure horizon (ictal segment). Both architectures were trained on 10 IS and two control subjects (see Table IA), and they were tested on four IS and two control subjects (see Table IB). The Number of Windows columns show ranges because this value depends on the window size. A window size of 15s gives the lowest number of windows while a size of 5s gives the highest number. During testing, the sentence length was varied between 1 and 100, and the length that gave the best results was chosen for the final iteration of each network.

A receiver operating characteristic (ROC) curve was constructed by testing the CNN-LSTM and MVNN-BiLSTM over a range of values for T . By testing the network on ictal and non-ictal segments, we counted the number of true positives (TP), false negatives (FN), true negatives (TN), and false positives (FP). The values FP and TN corresponded to the number of non-ictal segments and ictal segments after clinical onset that were classified as positive and negative, respectively. The values TP and FN were the number of ictal segments before the clinical onset that were counted as positive and negative, respectively.

Overall accuracy, sensitivity, specificity, and $F1$ score of the best iterations of the two architectures corresponding to a T which maximized the sensitivity (the point with the least distance to the coordinate $(0, 1)$) were computed as follows:

$$\begin{aligned} \text{accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} \\ \text{sensitivity} &= \frac{TP}{TP + FN} \\ \text{specificity} &= \frac{TN}{TN + FP} \\ \text{F1 Score} &= \frac{2TP}{2TP + FP + FN} \end{aligned} \quad (16)$$

Since there was a relatively small number of segments, an exponential function was fitted to the ROC graph by tuning a parameter ξ_e called the specificity factor that represented how quickly the function can reach the optimal sensitivity as a function of specificity:

$$\text{sensitivity} = 1 - \exp\left(-\frac{1 - \text{specificity}}{1 - \xi_e}\right) \quad (17)$$

During a preliminary analysis of the networks, it became clear that the density of its outputs should be considered when making a final decision on whether a spasm would occur and reducing the false positive rate. Data that contained incoming spasms tended to contain higher densities of high probabilities. Therefore, another threshold (apart from the decision threshold T), the sentence length, was added to the classifiers as an additional parameter for optimization of the final versions. It is effectively the minimum number of probabilities in a sequence that must be

high (larger than T) for the algorithm to be sufficiently confident that a spasm will occur within the seizure horizon.

The detection latency was defined as the time between the end of the first detected sentence in an ictal segment and the onset of a spasm cluster. Both networks were tested on the ictal segments up to the clinical onsets of spasms, and all the detected windows for each subject were compiled and recorded.

H. Comparison With CHB-MIT Dataset

Part of the novelty of this work involves predicting infantile spasms in IS subjects, but another lies in the biological inspirations of the MVNN-BiLSTM. In order to compare this method to other seizure prediction methods found in the literature, this method is used to analyze 24 subjects from the standard CHB-MIT dataset [40]. Similar to other models presented by three deep learning networks [13], [15], [17] that primarily focused on the CHB-MIT dataset, the subjects are analyzed on a per patient basis. This means that the training and testing process is completed independently for each subject. The training and testing split is on a per segment basis where the train/test split is given by 70%/30% of the total amount of time.

I. Robustness Testing With Noise

To test the robustness of the CNN-LSTM and MVNN-BiLSTM, white and pink noise signals using Gaussian distributed noise with a power law spectrum [41] characterized by a signal to noise ratio of 0.1 were added to a subset of test subjects. The GPACs of simulated noise EEGs were computed and then added to the original GPACs of Subjects 7 and 14. Both networks (with the same parameters when trained on noiseless segments) were tested on IS and control subjects.

J. Post-Processing Stage

This stage includes an output stage that can produce (a) a warning signal, (b) a wireless transmission to a smart phone, and/or (c) a record in an electronic diary, e.g., in a home monitoring SPS system.

IV. RESULTS

A. Overview

During testing of the CNN-LSTM and MVNN-BiLSTM, it was determined that the CNN-LSTM performed overall best with 10s GPAC windows while the MVNN-BiLSTM performed best with 5s GPAC windows. Both networks were evaluated only on the test set.

Figs. 2 and 3 show the EEGs and GPACs that led to true predictions and false predictions made by the CNN-LSTM and the MVNN-BiLSTM in Subjects 10 and 14. According to Fig. 2, both networks were able to predict a spasm in Subject 14, although the timing of the MVNN-BiLSTM was slightly earlier. Fig. 3 shows that the MVNN-BiLSTM correctly detected no spasms in Subject 10 in an interictal segment whereas the CNN-LSTM made one incorrect prediction.

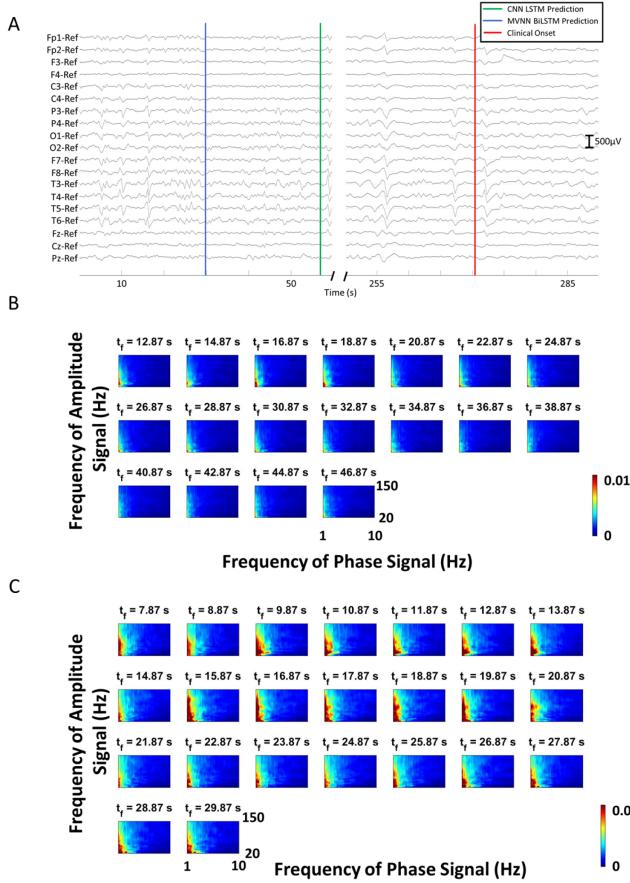


Fig. 2. (a) Ictal EEGs of subject 14. Both CNN-LSTM and MVNN-BiLSTM output a prediction prior to the clinical onset of the spasm, for a true positive. (b) GPAC images analyzed by CNN-LSTM which led to a true positive. (c) GPAC images analyzed by MVNN-BiLSTM which led to a true positive. The colour bar maps GPAC values to colours in the GPAC images.

Fig. 4 gives a holistic perspective of the performances of the CNN-LSTM and MVNN-BiLSTM. The CNN-LSTM correctly detected spasms in all 4/4 ictal segments, and it correctly detected no spasms in 2/5 non-ictal segments where it gave one false prediction in Subjects 10, 11, and 15 (Fig. 4(a)). The MVNN-BiLSTM method correctly detected spasms in all 4/4 ictal segments, and it correctly detected no spasms in all 5/5 non-ictal segments (Fig. 4(b)).

As shown in Fig. 5, ROC curves and detection latency box plots were generated for both the CNN-LSTM and MVNN-BiLSTM. The mean detection latency was calculated over all segments containing a spasm in the test set. Overall, the metrics (optimal sensitivity, optimal specificity, area under the ROC curve) of the CNN-LSTM are lower than those of the MVNN-BiLSTM.

B. Effect of Noise

White noise represents a more generic source of signal contamination, e.g., poor electrode placement, while pink noise represents more biological sources, such as EEG noise. White

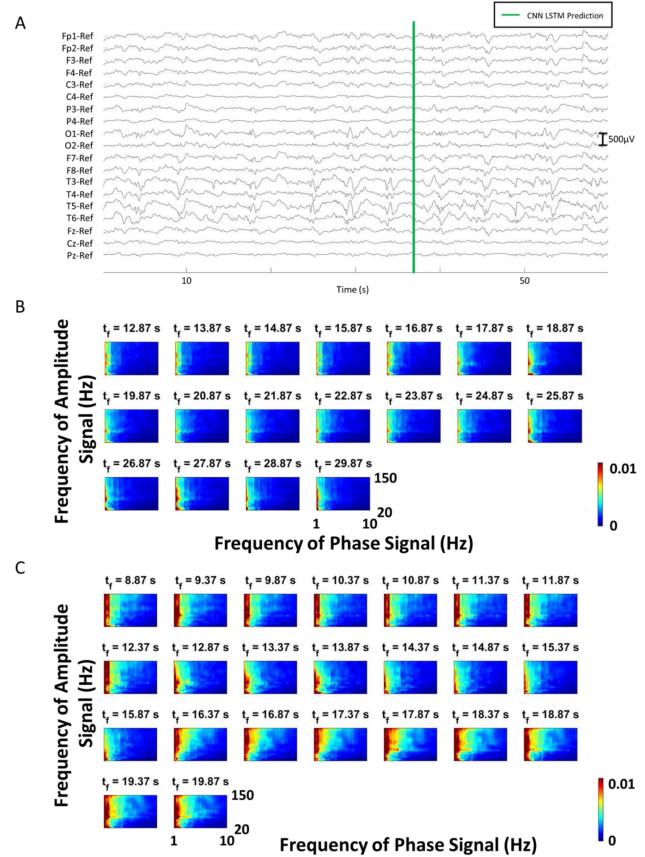


Fig. 3. (a) Interictal EEGs of subject 10. The CNN-LSTM outputs a prediction unrelated to a clinical spasm onset for a false positive, while the MVNN-BiLSTM does not give a true negative. (b) GPAC images analyzed by CNN-LSTM which led to a false positive. (c) GPAC images analyzed by MVNN-BiLSTM which led to a true negative. The colour bar maps GPAC values to colours in the GPAC images.

and pink noise were added to the GPACS (shown in Supplementary), and the resulting predictions are shown in Fig. 6, and their corresponding ROC curves are shown in Fig. 7.

Both the CNN-LSTM and the MVNN-BiLSTM only gave one false positive when tested on Subject 14 with white noise. The CNN-LSTM gave one false negative on Subject 14 with pink noise while the MVNN-BiLSTM gave none.

C. Comparison of SPS Methods

Table II shows a comparison between the performances of the CNN-LSTM and the MVNN-BiLSTM. The results are reported based on the number of ictal/interictal/control segments listed in Table IB.

The MVNN-BiLSTM achieved a better F1 score, area under the ROC curve, and false positive rate, and it also had a comparable mean detection latency with the CNN-LSTM, giving 0 false positives. Another LSTM-based method demonstrated similar performances, producing average sensitivities and specificities of 0.9963 and 0.9978, respectively [13]. A method which combined CNNs and LSTMs achieved sensitivities and specificities of 0.9188 and 0.8613, respectively [18]. However, since these

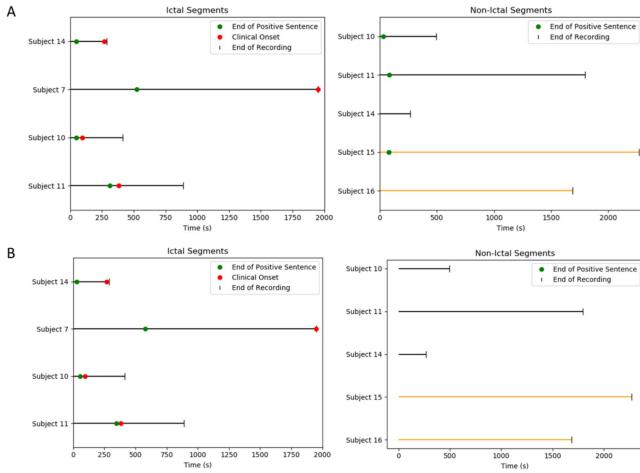


Fig. 4. Green circles mark the ends of positive GPAC sentences where the network predicted that a spasm would occur within a seizure horizon of 1h. Red circles denote the beginning of a spasm, and vertical lines mark the end of the given segment. The horizontal orange lines highlight segments from control subjects. (a) Overview of CNN-LSTM predictions in ictal segments (EEG segments within 1h of the first spasm onset) and non-ictal segments (interictal EEG segments before 1h of the first spasm onset or control EEG segments). (b) Same as part (a) but with the MVNN-BiLSTM.

TABLE II
COMPARISON ACROSS SPS METHODS

SPS Method	Mean Detection Latency (s)	F1 Score	Area under ROC Curve	False Positives	False Negatives
CNN-LSTM	-635.95 ± 740.14	0.727	0.7361	3	0
CNN-LSTM with Noise	-1064.02 ± 841.53	0.667	0.8281	1	1
MVNN-BiLSTM	-649.95 ± 741.88	1.000	0.9583	0	0
MVNN-BiLSTM with Noise	-1081.02 ± 841.53	0.889	0.9375	1	0

methods do not use databases containing IS subjects, they are not included in Table II.

D. Comparison With CHB-MIT Dataset

Table III lists comparisons of our MVNN Bi-LSTM with models presented by three deep learning works [13], [15], [18].

V. DISCUSSION

A. Comparison of SPS Methods

The MVNN-BiLSTM achieved a better performance than the CNN-LSTM in terms of F1 score and false positive rate, giving 0 false positives (Table II). In practice, the MVNN-BiLSTM

TABLE III
COMPARISON WITH CHB-MIT DATASET

Subject	Method	Sensitivity	Specificity	FPR (FP/h)
CHB01	[13]	100	100	0
	[17]	100	N/A	0
	[15]	98.7	99.7	N/A
CHB02	Our MVNN	100	100	0
	[13]	100	100	0
	[17]	100	N/A	0.06
CHB03	[15]	85.3	98.0	N/A
	Our MVNN	100	100	0
	[13]	100	100	0
CHB04	[17]	75	N/A	0
	[15]	99.3	96.3	N/A
	Our MVNN	100	100	0
CHB05	[13]	99.97	99.97	0
	[17]	N/A	N/A	N/A
	[15]	89.3	95.0	N/A
CHB06	Our MVNN	100	100	0
	[13]	99.65	100	0
	[17]	100	N/A	0
CHB07	[15]	98.0	99.0	N/A
	Our MVNN	100	100	0
	[13]	99.87	98.80	0.13
CHB08	[17]	N/A	N/A	N/A
	[15]	96.3	95.3	N/A
	Our MVNN	100	100	0
CHB09	[13]	99.94	99.75	0.06
	[17]	100	N/A	0.16
	[15]	97.3	98.0	N/A
CHB10	Our MVNN	100	100	0
	[13]	99.75	100	0
	[17]	N/A	N/A	N/A
CHB11	[15]	100	100	0
	Our MVNN	100	100	0
	[13]	100	100	0
CHB12	[17]	N/A	N/A	N/A
	[15]	91.0	99.7	N/A
	Our MVNN	100	100	0
CHB13	[13]	99.89	100	0
	[17]	N/A	N/A	N/A
	[15]	94.3	99.7	N/A
CHB14	Our MVNN	100	85.7	0.14
	[13]	98.72	99.36	0.08
	[17]	N/A	N/A	N/A
CHB15	[15]	95.0	93.3	N/A
	Our MVNN	100	100	0
	[13]	99.50	100	0
CHB16	[17]	N/A	N/A	N/A
	[15]	93.7	100	N/A
	Our MVNN	100	87.5	0.13
CHB17	[13]	99.68	99.68	0.05
	[17]	N/A	N/A	N/A
	[15]	92.0	93.7	N/A
CHB18	Our MVNN	100	100	0
	[13]	99.89	100	0
	[17]	100	N/A	0.22
CHB19	[15]	85.3	94.3	N/A
	Our MVNN	100	100	0
	[13]	99.83	99.49	0.08
CHB20	[17]	100	N/A	0.13
	[15]	93.7	100	N/A
	Our MVNN	100	100	0
CHB21	[13]	100	100	0
	[17]	100	N/A	0.5
	[15]	84.0	84.0	N/A
CHB22	Our MVNN	100	100	0
	[13]	100	100	0
	[17]	100	N/A	0.18
CHB23	[15]	90.0	95.0	N/A
	Our MVNN	100	100	0
	[13]	100	100	0
CHB24	[17]	100	N/A	N/A
	[15]	92.7	93.0	N/A
	Our MVNN	100	100	0

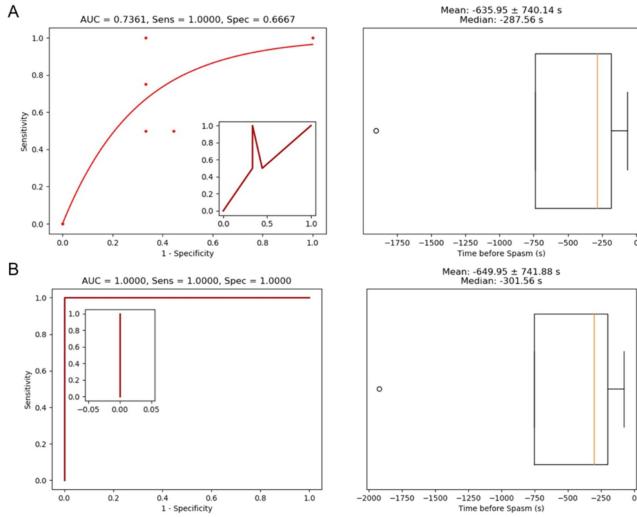


Fig. 5. ROC curve and detection latency box plots. Detection latency is the time between the spasm onset and the end of the last GPAC window giving a prediction. The raw ROC curves without fitting are shown as insets. (a) Results of CNN-LSTM using 10s GPAC windows binarized with $P_{GPAC} = 0.340$, $T = 0.005$, and a sentence length of 16. Overall, it achieves an F1 score of 0.727 and an AUC of 0.7361. It has a specificity factor of $\xi_e = 0.700$ and a goodness of fit of $R^2 = 0.695$. (b) Results of MVNN-BiLSTM using 5s GPAC windows binarized with $P_{GPAC} = 0.290$, $T = 0.450$, and a sentence length of 21. Overall, it achieves an F1 score of 1.00 and an AUC of 1.00. It has a specificity factor of $\xi_e = 0.922$ and a goodness of fit of $R^2 = -0.048$. The small R^2 value is due to most points having a specificity of 1.

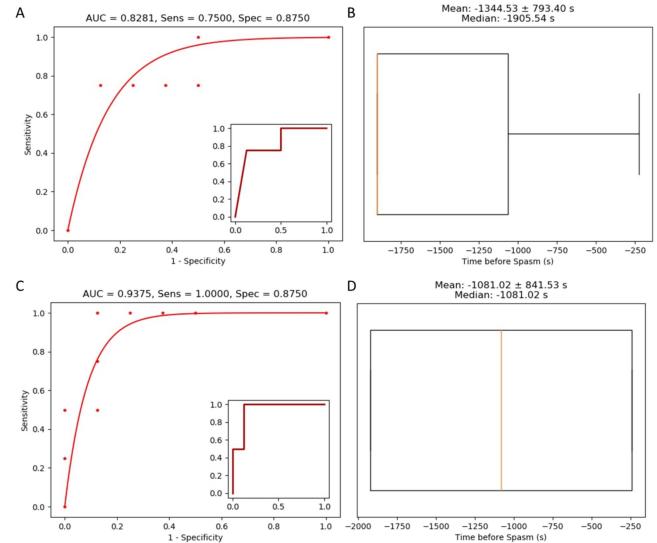


Fig. 7. ROC curve and detection latency box plots of performance on noisy data. Detection latency is the time between the spasm onset and the end of the last GPAC window giving a prediction. The raw ROC curves without fitting are shown as insets. (a) Results of a CNN-LSTM using 10s GPAC windows binarized with $P_{GPAC} = 0.340$, $T = 0.005$, and a sentence length of 16. Overall, it achieves an F1 score of 0.667. It has a specificity factor of $\xi_e = 0.843$ and a goodness of fit of $R^2 = 0.834$. (b) Results of an MVNN-BiLSTM using 5s GPAC windows binarized with $P_{GPAC} = 0.290$, $T = 0.065$, and a sentence length of 21. Overall, it achieves an F1 score of 0.889. It has a specificity factor of $\xi_e = 0.912$ and a goodness of fit of $R^2 = 0.640$.

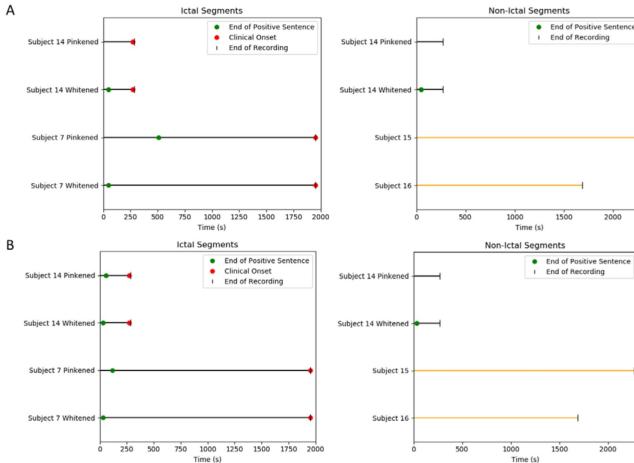


Fig. 6. Same labelling convention as Fig. 4 but applied to noisy segments. (a) Overview of CNN-LSTM predictions in ictal segments (EEG segments within 1h of the first spasm onset) and non-ictal segments (interictal EEG segments at least 1h away from a spasm cluster or control EEG segments). (b) Same as part (a) but with an MVNN-BiLSTM.

provides a more clinically significant result since false positives are typically the most difficult source of error for SPS methods.

Furthermore, to compare the MVNN-BiLSTM with other existing SPS methods, it was trained and tested on the subjects of the standard CHB-MIT dataset. It performed at least as well as other state-of-the-art models (Table III), suggesting that the MVNN-BiLSTM can generalize well to multiple types

of seizure events (not just IS). We highlight the two youngest subjects, CHB06 (aged 1.5 years) and CHB12 (aged 2 years). They provided 52.8h and 14.7h of EEG data, respectively. Their ages are relatively close to the ages of this study's primary age group (up to 2 years for the IS subjects). The MVNN-BiLSTM predicts all seizure events within the pre-ictal segments, and it gives no false positives during the interictal segments. Table III shows that the MVNN Bi-LSTM gave perfect results for these two subjects, suggesting that it may be a more general seizure prediction tool for infants.

B. Association With Neural Coding in the Brain

This work incorporates principles of neural coding into the problem of early seizure detection in two ways. The first involves extracting cross-rhythm interactions from EEG recordings, capturing frequency-based codes associated with neural phenomena. Transforming the raw EEG inputs into PAC images can allow the MVNN-BiLSTM to require a relatively small amount of training data to learn how to recognize the complex processes within the brain which precede a spasm. This is because the PAC images are richer in frequency-based information than the raw EEG signals, meaning it is easier for a machine learning algorithm such as the MVNN-BiLSTM to learn more relevant patterns that can predict spasms. Furthermore, the use of PAC features is biomimetic in that they are associated with the brain's processing and coding of information. For example, the Theta-Gamma code is used by the brain to code information about navigation in space [42].

Recent computational models have suggested a link between pathological PAC features and astrocytic activity [43]. In astrocytes, a large amount of activity occurs in a bi-directional manner, primarily driven by gap junctional connections between the cells [34]. Generally, gap junctions between excitable cells such as neurons and cardiac cells provide fast and bidirectional communication between them [35]. To incorporate these phenomena into our novel model, we utilized bidirectional LSTMs.

Another example of a brain-mimetic method used by this work is the incorporation of PDMs as the input layer to the network. Although the differentiating and integrating PDMs extracted from EEG signals in our previous work [24] were based on time, a filter bank of PDMs can be seen as equivalent for GPAC images with frequency axes. Hence, it is possible to combine existing image processing techniques such as Sobel operators and cumulative sum operations with our GPAC-based SPS to yield our best results on seizure prediction of IS to date.

Given the rich literature on the mechanisms of the brain, there are many possibilities to extend our algorithms to make them more brain-mimetic. One option involves adding cognitive rhythm generators to the architecture since these computational models have proven to be successful in capturing the neural rhythms of the brain while also using PDMs as inputs [25]. Since this network can understand the coding that occurs within the brain, it is highly feasible that it would provide a strong asset to the existing MVNN-BiLSTM.

C. Addition of Noise

While both the MVNN-BiLSTM and the CNN-LSTM performed comparably with the addition of white noise, the former was able to surpass the latter in performance when pink noise was added. This suggests that the MVNN-BiLSTM is more robust to noise, particularly pink noise, where other EEG processes may otherwise obfuscate spasm onset features.

D. Relevance of Epileptic Spasm Prediction

Infantile spasms make up a particular seizure type that needs identifying when suspicious behaviors are noticed in babies. Clinicians have difficulty even in just determining whether a seizure has occurred solely from the EEG traces. The significant clinical gap is that epileptic spasms can be easily missed in routine EEG studies. Even experienced pediatric electroencephalographers sometimes have a hard time picking up epileptic spasms on EEG. Without simultaneous video recordings, it would be even more challenging. The clinical utility of this algorithm would be to mark the occurrence of spasms at high accuracies just based on EEG recordings. The experts can then go to such segments and carefully verify the findings. This becomes more of a problem in post-treatment EEG or possible relapse cases, where hypsarrhythmia is not typically seen but epileptic spasms need to be found on EEG. We would like to emphasize that this prediction system has a good biological mechanistic basis (delta-HFO phase amplitude coupling which is well-described in ictogenesis in spasms), and it may possibly be used to evaluate treatment response or relapse risk in future studies. Thus, we clearly see the benefit of this approach in a way that machine

learning algorithms can complement human effort. Furthermore, the availability of algorithms that predict with high accuracy and specificity EEG-captured epileptic spasms will be an essential unbiased tool in clinical trials where the effects of treatments on spasms are being tested.

E. Limitations and Future Directions

The biological inspirations of the MVNN-BiLSTM had a few limitations. Specifically, the number of kernels used by the network is a hyperparameter, and this could have been modified to include derivatives that had a higher order than only two. This would effectively add more complexity to the model which could potentially allow it to learn the temporal sequence preceding a spasm more specifically, leading to an earlier prediction time. The fully connected layer was also biologically limited since it did not employ polynomial activations as performed in polynomial ANNs by Marmarelis [28]. In this study, a hyperbolic tangent activation function was used instead to avoid exploding gradients, but other methods that limit the weights of the network should also be explored in future work.

Another limitation is that spasms occurring too soon after a previous spasm event may be missed by the classifier. The MVNN-BiLSTM was trained using only the clinical onsets of the first spasm of a cluster. Future studies that train SPS methods based on clinical onsets of individual spasms may be helpful in refining them for individual ictal events. Furthermore, the detection latencies may depend on the total amount of data available for each subject since we observed that extending the amount of data to be analyzed in reverse time tended to increase the latency. Additional validation using prolonged EEG recordings would be helpful to report a more accurate false positive rate with respect to time instead of segments and more accurately determine the earliest detection latency with respect to a spasm cluster. It is not completely clear whether the reported MVNN-BiLSTM detections truly correspond to the ensuing spasm events or other features intrinsic to the ictal state. During training and testing, the EEG after the first cluster of spasms was not analyzed, meaning there could have been other spasms which might explain some of the recurrent alarms. In the future, testing on continuous EEGs including segments during and after spasm clusters should be carried out to make this analysis more thorough.

In this retrospective study, all subjects had hypsarrhythmia and seizures with electrodecremental responses, and the focus on this subset is a limitation of the study. Hypsarrhythmia is highly subjective and exhibits poor inter-rater reliability, especially in different institutions. In a multicenter study of 447 patients with IS, 82 percent had hypsarrhythmia [44]. This is consistent with an earlier report that approximately 80 percent of infants with infantile spasms have, at some time, a characteristic EEG pattern of hypsarrhythmia [45].

The number of subjects in this study was quite limited due to the rareness of IS. However, this also means that IS is a relatively misunderstood phenomenon, making the problem of prediction a much more difficult problem than other common types of epilepsy like temporal lobe epilepsy. However, the

MVNN-BiLSTM was able to generalize well to a test set that was completely separated from the training set and performed well on seizure prediction using a standard CHB-MIT data set.

The main benefit of this work when applied to IS subjects is that an epileptologist would be able to identify which EEG segments are at risk for containing a spasm. However, this still requires them to sift through the segments to manually identify when the spasm occurs. A future work for this study would be to have the model not only output the presence of an incoming spasm, but also to provide an estimate of the detection latency. The epileptologist would then have a better idea of where in the segment they should search for the spasm.

VI. CONCLUSION

Biomimetic, more specifically brain-mimetic, deep learning network methods provide a promising future application for neural networks in seizure prediction since they utilize some the extensive knowledge about coding and learning in the electrical networks of the brain. Not only was the MVNN-BiLSTM able to produce similar detection latencies as previous works, but it was even able to reduce the false positive count to 0, outperforming conventional architectures. This provides a significant contribution to the problem of identifying EEG segments containing epileptic seizures in adults and spasms in infants.

VI. CONFLICTS OF INTEREST

AS Galanopoulou is the Editor-in-Chief of Epilepsia Open and receives royalties from Elsevier, Medlink, and Morgan and Claypool for publications.

SL Moshé is serving as Associate Editor of Neurobiology of Disease. He is on the editorial board of Brain and Development, Pediatric Neurology, Annals of Neurology, MedLink and Physiological Research. He receives from Elsevier for his work as Associate Editor in Neurobiology of Disease; annual compensation from MedLink; and royalties from two books he co-edited.

ACKNOWLEDGMENT

BL Bardakjian acknowledges grant support by the Natural Sciences and Engineering Research Council of Canada (NSERC), and EpLink - the Epilepsy Research Program of the Ontario Brain Institute. The Ontario Brain Institute is an independent non-profit corporation, funded partially by the Ontario government. The opinions, results and conclusions are those of the authors and no endorsement by the Ontario Brain Institute is intended or should be inferred. Computations were performed on the Niagara supercomputer at the SciNet HPC Consortium. SciNet is funded by the Canada Foundation for Innovation; the Government of Ontario; Ontario Research Fund - Research Excellence; and the University of Toronto.

REFERENCES

- [1] I. Kuzovkin et al., "Activations of deep convolutional neural networks are aligned with Gamma band activity of human visual cortex," *Commun. Biol.*, vol. 1, no. 107, pp. 1–12, 2018, doi: [10.1038/s42003-018-0110-y](https://doi.org/10.1038/s42003-018-0110-y).
- [2] J. E. Lisman and O. Jensen, "The θ - γ neural code," *Neuron*, vol. 77, no. 6, pp. 1002–1016, 2013, doi: [10.1016/j.neuron.2013.03.007](https://doi.org/10.1016/j.neuron.2013.03.007).
- [3] W. Bedingham and W. G. Tatton, "Kinematic representation of imposed forearm movements by pericruciate neurons (areas 4 and 3a) in the awake cat," *J. Neurophysiol.*, vol. 53, no. 4, pp. 886–909, 1985, doi: [10.1152/jn.1985.53.4.886](https://doi.org/10.1152/jn.1985.53.4.886).
- [4] V. Grigorovsky et al., "Machine intelligence-based epileptic seizure forecasting," in *Neural Engineering*. Berlin, Germany: Springer, 2020, pp. 535–565.
- [5] M. Guirgis et al., "Defining regions of interest using cross-frequency coupling in extratemporal lobe epilepsy patients," *J. Neural Eng.*, vol. 12, no. 2, pp. 1–19, 2015, doi: [10.1088/1741-2560/12/2/026011](https://doi.org/10.1088/1741-2560/12/2/026011).
- [6] D. Jacobs et al., "Classification of pre-clinical seizure states using scalp EEG cross-frequency coupling features," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 11, pp. 2440–2449, Nov. 2018, doi: [10.1109/TBME.2018.2797919](https://doi.org/10.1109/TBME.2018.2797919).
- [7] D. Jacobs et al., "Classification of scalp EEG states prior to clinical seizure onset," *IEEE J. Transl. Eng. Health Med.*, vol. 7, 2019, Art. no. 2000203, doi: [10.1109/JTEHM.2019.2926257](https://doi.org/10.1109/JTEHM.2019.2926257).
- [8] H. Hashimoto et al., "Phase-amplitude coupling between infraslow and high-frequency activities well discriminates between the preictal and interictal states," *Sci. Rep.*, vol. 11, no. 1, pp. 1–13, Aug. 2021, doi: [10.1038/s41598-021-96479-1](https://doi.org/10.1038/s41598-021-96479-1).
- [9] Z. C. Lipton et al., "A critical review of recurrent neural networks for sequence learning," 2015, *arXiv:1506.00019v4*.
- [10] D. E. Rumelhart and J. L. McClelland, "Learning internal representations by error propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*. Cambridge, MA, USA: MIT Press, 1987, pp. 318–362.
- [11] Y. Duan, Y. L.V., and F.-Y. Wang, "Travel time prediction with LSTM neural network," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst.*, 2016, pp. 1053–1058, doi: [10.1109/ITSC.2016.7795686](https://doi.org/10.1109/ITSC.2016.7795686).
- [12] J. Chung, "Empirical evaluation of gated recurrent neural networks on sequence modeling," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1–9.
- [13] K. Tsioris et al., "A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals," *Comput. Biol. Med.*, vol. 99, pp. 24–37, 2018, doi: [10.1016/j.combiomed.2018.05.019](https://doi.org/10.1016/j.combiomed.2018.05.019).
- [14] J. W. Wheless et al., "Infantile spasms (West syndrome): Update and resources for pediatricians and providers to share with parents," *BMC Pediatrics*, vol. 12, no. 108, pp. 1–9, 2012, doi: [10.1186/1471-2431-12-108](https://doi.org/10.1186/1471-2431-12-108).
- [15] M. Zhou et al., "Epileptic seizure detection based on EEG signals and CNN," *Front. Neuroinform.*, vol. 12, Dec. 2018, Art. no. 95, doi: [10.3389/FNINF.2018.00095/BIBTEX](https://doi.org/10.3389/FNINF.2018.00095).
- [16] F. E. Ibrahim et al., "Deep-learning-based seizure detection and prediction from electroencephalography signals," *Int. J. Numer. Methods Biomed. Eng.*, vol. 38, no. 6, Jun. 2022, Art. no. e3573, doi: [10.1002/CNM.3573](https://doi.org/10.1002/CNM.3573).
- [17] M. Shahbazi and H. Aghajani, "A generalizable model for seizure prediction based on deep learning using CNN-LSTM architecture," in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, 2018, pp. 469–473, doi: [10.1109/GlobSIP.2018.8646505](https://doi.org/10.1109/GlobSIP.2018.8646505).
- [18] X. Wei et al., "Early prediction of epileptic seizures using a long-term recurrent convolutional network," *J. Neurosci. Methods*, vol. 327, no. 1, pp. 1–10, 2019, doi: [10.1016/j.jneumeth.2019.108395](https://doi.org/10.1016/j.jneumeth.2019.108395).
- [19] C. D. James et al., "A historical survey of algorithms and hardware architectures for neural-inspired and neuromorphic computing applications," *Biologically Inspired Cogn. Architectures*, vol. 19, pp. 49–64, Jan. 2017, doi: [10.1016/j.bica.2016.11.002](https://doi.org/10.1016/j.bica.2016.11.002).
- [20] A. M. Zador, "A critique of pure learning and what artificial neural networks can learn from animal brains," *Nature Commun.*, vol. 10, 2019, Art. no. 3770, doi: [10.1038/s41467-019-11786-6](https://doi.org/10.1038/s41467-019-11786-6).
- [21] V. Z. Marmarelis, "Modeling methodology for nonlinear physiological systems," *Ann. Biomed. Eng.*, vol. 25, no. 2, pp. 239–251, 1997, doi: [10.1007/bf02648038](https://doi.org/10.1007/bf02648038).
- [22] V. Z. Marmarelis, "Nonlinear dynamic modeling of physiological systems," in *Nonlinear Dynamic Modeling of Physiological Systems*. Hoboken, NJ, USA: Wiley, 2004, pp. 179–264.
- [23] G. D. Mitsis, M. G. Markakis, and V. Z. Marmarelis, "Nonlinear modeling of the dynamic effects of infused insulin on glucose: Comparison of compartmental with Volterra models," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 10, pp. 2347–2358, Oct. 2009, doi: [10.1109/TBME.2009.2024209](https://doi.org/10.1109/TBME.2009.2024209).
- [24] E. E. Kang et al., "Markers of pathological excitability derived from principal dynamic modes of hippocampal neurons," *J. Neural Eng.*, vol. 9, no. 5, pp. 1–11, 2012, doi: [10.1088/1741-2560/9/5/056004](https://doi.org/10.1088/1741-2560/9/5/056004).
- [25] O. C. Zalay and B. L. Bardakjian, "Theta phase precession and phase selectivity: A cognitive device description of neural coding," *J. Neural Eng.*, vol. 6, no. 3, pp. 1–18, 2009, doi: [10.1088/1741-2560/6/3/036002](https://doi.org/10.1088/1741-2560/6/3/036002).

- [26] O. C. Zalay and B. L. Bardakjian, "Synthesis of high-complexity rhythmic signals for closed-loop electrical neuromodulation," *Neural Netw.*, vol. 42, pp. 62–73, 2013, doi: [10.1016/j.neunet.2013.01.005](https://doi.org/10.1016/j.neunet.2013.01.005).
- [27] F. H. Farah et al., "Coupled oscillators model of hyperexcitable neuroglial networks," *Int. J. Neural Syst.*, vol. 29, no. 3, pp. 1–17, 2019, doi: [10.1142/S0129065718500417](https://doi.org/10.1142/S0129065718500417).
- [28] K. H. Chon et al., "Comparative nonlinear modeling of renal autoregulation in rats: Volterra approach versus artificial neural networks," *IEEE Trans. Neural Netw.*, vol. 9, no. 3, pp. 430–435, May 1998, doi: [10.1109/72.668884](https://doi.org/10.1109/72.668884).
- [29] G. R. Lee et al., "PyWavelets: A python package for wavelet analysis," *J. Open Source Softw.*, vol. 4, 2019, Art. no. 1237, doi: [10.21105/joss.01237](https://doi.org/10.21105/joss.01237).
- [30] A. B. L. Tort et al., "Measuring phase-amplitude coupling between neuronal oscillations of different frequencies," *J. Neurophysiol.*, vol. 104, no. 2, pp. 1195–1210, 2010, doi: [10.1152/jn.00106.2010](https://doi.org/10.1152/jn.00106.2010).
- [31] T. T. K. Munia and S. Aviyente, "Time-frequency based phase-amplitude coupling measure for neuronal oscillations," *Sci. Rep.*, vol. 9, no. 1, pp. 1–15, 2019, doi: [10.1038/s41598-019-48870-2](https://doi.org/10.1038/s41598-019-48870-2).
- [32] Z. Deng et al., "Massively parallel non-stationary EEG data processing on GPGPU platforms with Morlet continuous wavelet transform," *J. Internet Serv. Appl.*, vol. 3, no. 1, pp. 347–357, 2012, doi: [10.1007/s13174-012-0071-1](https://doi.org/10.1007/s13174-012-0071-1).
- [33] C. Steinhäuser et al., "Astrocyte dysfunction in temporal lobe epilepsy: K⁺ channels and gap junction coupling," *Glia*, vol. 60, no. 8, pp. 1192–1202, Aug. 2012, doi: [10.1002/GLIA.22313](https://doi.org/10.1002/GLIA.22313).
- [34] S. Mylvganam et al., "Roles of gap junctions, connexins, and panxins in epilepsy," *Front. Physiol.*, vol. 5, 2014, Art. no. 172, doi: [10.3389/FPHYS.2014.00172](https://doi.org/10.3389/FPHYS.2014.00172).
- [35] D. C. Spray and R. Dermietzel, *Gap Junctions in the Nervous System*, 1st ed. Berlin, Germany: Springer, 1996.
- [36] A. T. Winfree, *The Geometry of Biological Time*. Berlin, Germany: Springer, 1980, pp. 83–84.
- [37] V. Z. Marmarelis, "Signal transformation and coding in neural systems," *IEEE Trans. Biomed. Eng.*, vol. 36, no. 1, pp. 15–24, Jan. 1989, doi: [10.1109/10.16445](https://doi.org/10.1109/10.16445).
- [38] G. Mitsis et al., "Principal dynamic mode analysis of action potential firing in a spider mechanoreceptor," *Biol. Cybern.*, vol. 96, no. 1, pp. 113–127, Jan. 2007, doi: [10.1007/S00422-006-0108-2](https://doi.org/10.1007/S00422-006-0108-2).
- [39] S. Liu et al., "Exploring the time-frequency content of high frequency oscillations for automated identification of seizure onset zone in epilepsy," *J. Neural Eng.*, vol. 13, no. 2, pp. 1–15, 2016, doi: [10.1088/1741-2560/13/2/026026](https://doi.org/10.1088/1741-2560/13/2/026026).
- [40] A. H. 1981-Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," 2009. Accessed: Feb. 16, 2023. [Online]. Available: <https://dspace.mit.edu/handle/1721.1/54669>
- [41] J. Timmer and M. Konig, "On generating power law noise," *Astron. Astrophys.*, vol. 300, pp. 707–710, 1995.
- [42] J. E. Lisman and O. Jensen, "The θ - γ neural code," *Neuron*, vol. 77, no. 6, pp. 1002–1016, Mar. 2013, doi: [10.1016/J.NEURON.2013.03.007](https://doi.org/10.1016/J.NEURON.2013.03.007).
- [43] V. Grigorovsky, V. L. Breton, and B. L. Bardakjian, "Glial modulation of electrical rhythms in a neuroglial network model of epilepsy," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 7, pp. 2076–2087, Jul. 2021.
- [44] S. T. Demarest et al., "The impact of hypsarrhythmia on infantile spasms treatment response: Observational cohort study from the National Infantile Spasms Consortium," *Epilepsia*, vol. 58, no. 12, pp. 2098–2103, Dec. 2017, doi: [10.1111/EPI.13937](https://doi.org/10.1111/EPI.13937).
- [45] P. Jeavons and B. Bower, *Clinics in Developmental Medicine No. 15 Infantile Spasms: A Review of the Literature and a Study of 112 Cases*. London, U.K.: Heinemann, 1964.