Integrating Convolutional Neural Networks and Probabilistic Graphical Modeling for Epileptic Seizure Detection in Multichannel EEG

Jeff Craley[1(✉)], Emily Johnson[2], and Archana Venkataraman[1]

[1] Department of Electrical and Computer Engineering,
Johns Hopkins University, Baltimore, USA
jcraley2@jhu.edu
[2] Department of Neurology,
Johns Hopkins Medical Institute, Baltimore, USA

Abstract. Manual seizure detection in clinical electroencephalography (EEG) is time consuming and requires extensive training. In addition, the seizure origin and spreading pattern is valuable for therapeutic planning but cannot always be manually disambiguated. Prior work in automated seizure detection has focused on engineering new features that better capture the seizure activity. However, these methods ignore crucial information in the data and are not sensitive enough to track the seizure propagation. In this work we introduce a hybrid Probabilistic Graphical Model-Convolutional Neural Network (PGM-CNN) for seizure tracking in multichannel EEG. Our model leverages the power of deep learning for data driven analysis of the raw EEG time series while retaining clinically relevant information through the latent PGM prior. We validate our hybrid model on clinical EEG data from two hospitals with distinct patient populations. Our system achieves better detection performance than baseline methods, which exclusively use PGMs or neural networks.

1 Introduction

Epilepsy affects nearly 3.5 million people in the United States and is associated with a fivefold increase in mortality rate [1]. It has been estimated that 20–40% of epilepsy patients are medically refractory and do not respond to anti-epileptic drugs [2]. Subsequent treatments for these patients rely on clinicians being able to detect, and if appropriate, localize seizure activity in the brain. Due to the heterogeny of epilepsy disorders, scalp electroencephalography (EEG) recordings are critical for diagnosis and treatment planning. Typical in-patient evaluations for epilepsy involve continuous EEG recordings, sometimes for days. These recordings are manually inspected for seizure activity, a process which is time consuming, requires years of training, and is prone to human error.

Feature Engineering for EEG Analysis: Automated seizure detection has been an active field of research for the past three decades. Most algorithms

A. C. S. Chung et al. (Eds.): IPMI 2019, LNCS 11492, pp. 291–303, 2019.
https://doi.org/10.1007/978-3-030-20351-1_22

follow a two-stage machine learning pipeline consisting of (1) feature extraction from the EEG signal over short time windows, followed by (2) a binary classifier to identify seizure versus non-seizure intervals [3]. This end-to-end pipeline was exemplified by Shoeb et al. [4] where power features in different frequency bands were used in conjunction with a support vector machine classifier.

Prior work in the seizure detection community has focused largely on the feature extraction step. For example, Andrzejak et al. [5] noted that EEG during seizures exhibited a different degree of non-linearity than EEG recorded during baseline, inspiring the application of non-linear signal processing and chaos theory to EEG analysis. Similarly, Güler et al. [6] used Lyapunov exponents to discriminate between seizure and non-seizure EEG. While the above methods are promising, the generalization power is fundamentally limited by the chosen features. In addition, they perform classification independently for each time window and do not capture the seizure origin or manifestation.

Craley et al. [7] introduced a novel approach for seizure detection that used a Coupled Hidden Markov Model (CHMM) to track the propagation of a seizure across the scalp. However, this method relied on carefully selected features in order to learn a highly structured likelihood function. Despite leading to good performance, feature extraction focused specifically on a small number of spectral features. These features, combined with likelihood scoring using a restricted set of functions, likely missed relevant seizure information. In contrast, here we rely on data-driven strategies using deep architectures to directly learn more effective representations and analysis functions.

Data-Driven Representation Learning: Traditional representation learning techniques solve auxiliary problems in the pursuit of representations with desirable properties. While these properties are often useful, there are no guarantees that they are most appropriate for a given task. Alternatively, deep networks learn representations that capture facets of the data directly applicable to the task at hand [8]. This improved analytical power can come at the cost of interpretability, as these features may lack inuitive explanation.

This paper presents an integrated framework for epileptic seizure detection that blends the interpretability of Probabilistic Graphical Models (PGMs) with advancements in deep learning. Our PGM leverages the CHMM [7] for automated seizure tracking. We augment this PGM with a Convolutional Neural Network (CNN) likelihood model. Our PGM-CNN strategy can automatically learn the EEG features relevant for detection from limited amounts of training data. We demonstrate our PGM-CNN framework on multichannel EEG data acquired from two hospitals with distinct patient populations. Our PGM-CNN framework correctly identifies more of the annotated seizure activity in both datasets than comparable baseline methods. This performance suggests a new direction for automated seizure tracking in clinical EEG.

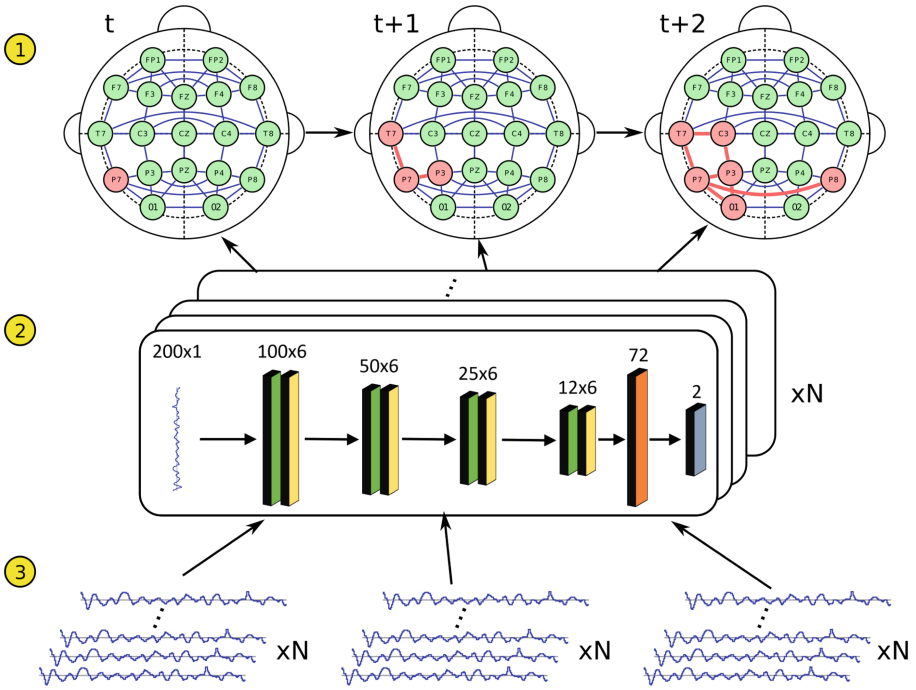


Fig. 1. Detail of the inference procedure. Time flows to the right while information flows upwards. In the third row, we depict the raw EEG signal. The signal from each channel is fed into a dedicated CNN for scoring in the second row. The first row depicts a hypothetical seizure spreading through the propagation network of the CHMM.

2 Integrating PGMs and Deep Learning

Figure 1 outlines our modeling strategy. Raw EEG signal from each channel in row three is fed directly into the CNNs in row two, where one CNN is trained for each channel. The CNNs score the signal for seizure activity and feed this information into the CHMM prior shown in row 1. The CHMM fuses these scores across the scalp and through time to perform posterior inference for seizure activity. Below, we formalize the mathematical relationships and inference procedure.

2.1 PGM Prior Based on the Coupled Hidden Markov Model

The PGM prior couples the hidden states of each EEG channel according to the previous states of the neighboring and contralateral channels [7]. For each channel i, we let the underlying seizure state at time t be represented by the variable $X_i^t \in \{0, 1, 2\}$ corresponding to pre-seizure baseline, seizure activity, and post-seizure baseline, respectively. The corresponding EEG data is represented by Y_i^t. We define the aunts of channel i, $au(i)$, as the neighbors in the graphs

shown in Fig. 1 and indicate the states of the ensemble of aunts of channel i at time t with $\mathbf{X}^t_{au(i)}$. The joint probability distribution can be written

$$P(\mathbf{X}, \mathbf{Y}) = \prod_{i=1}^{N} P(Y_i^0 \mid X_i^0) \prod_{t=1}^{T} P(Y_i^t \mid X_i^t) P(X_i^t \mid \mathbf{X}^{t-1}_{au(i)}, X_i^{t-1}), \tag{1}$$

where N indicates the number of channels and T is the length of the recording. Notice that we assume all channels are initially in baseline (i.e. $X_i^0 = 0, \forall i$) and have thus omitted the distribution over the initial state.

The transition probability $P(X_i^t \mid \mathbf{X}^{t-1}_{au(i)}, X_i^{t-1})$ for each chain depends only on the aunt states and the state of the chain in the previous timestep. These probabilities are encoded in a left-to-right time-inhomogenous transition matrix $\mathbf{A}_i^t$ where $P(X_i^t = k \mid X_i^{t-1} = j, X^{t-1}_{au(i)}) = A^t_{i,jk}$ as follows:

$$\mathbf{A}_i^t = \begin{bmatrix} 1 - g_i^t & g_i^t & 0 \\ 0 & 1 - h_i^t & h_i^t \\ 0 & 0 & 1 \end{bmatrix}. \tag{2}$$

Here g_i^t and h_i^t correspond to the probability that a channel enters or exits the seizure state, respectively. We model these probabilities as logistic regressions

$$\log\left(\frac{g_i^t}{1 - g_i^t}\right) = \rho_0 + \rho_1 \eta_i^t, \qquad \log\left(\frac{h_i^t}{1 - h_i^t}\right) = \phi_0 + \phi_1 \eta_i^t \tag{3}$$

such that ρ_0 corresponds to the base rate of seizure onset for each individual channel while ρ_1 corresponds to the influence of the aunt channels on the seizure onset. Likewise, the parameters ϕ_0 and ϕ_1 govern offset in an identical way.

2.2 Nonparameteric Likelihood via Convolutional Neural Networks

CNNs have become standard in computer vision due to their ability to learn spatially invariant features across multiple scales [8]. At a high level, the early layers learn simple features, such as edge detectors, while subsequent layers learn more and more complicated features. CNNs are also becoming popular for one-dimensional and time series data, where they provide a valuable alternative to the standard Recurrent Neural Network (RNN). While RNNs have been particularly effective in analyzing short sequences, CNNs with large receptive fields can be trained much faster than RNNs for long sequences. In addition, CNNs are restricted to learning highly structured functions composed of convolutions, which reduces their ability to overfit when training data is limited [8]. While CNNs are powerful tools for data analysis, they suffer from a lack of interpretability. However, from a clinical standpoint, we are primarily concerned with the seizure propagation patterns, as opposed to the underlying feature representation. Our hybrid approach captures the clinically relevant information by using a directly interpretable PGM prior while giving the CNN free rein over the

data likelihood to improve EEG signal analysis, resulting in gains in detection performance.

One important caveat to integrating a CNN data likelihood is that, by default, a CNN is trained for posterior inference. Namely, given the input data Y_i^t, the CNN will output a soft class assignment of seizure versus baseline, i.e. $P(X_i^t \mid Y_i^t)$. In contrast, the joint distribution in Eq. (1) relies on the data likelihood, $P(Y_i^t \mid X_i^t)$. We can obtain this factor by applying Bayes' rule:

$$P(Y_i^t \mid X_i^t) = \frac{P(X_i^t \mid Y_i^t)P(Y_i^t)}{P(X_i^t)} \propto \frac{P(X_i^t \mid Y_i^t)}{P(X_i^t)}. \tag{4}$$

Notice that we ignore the marginal probability $P(Y_i^t)$, as this term is the same regardless of the class label, and we only require data likelihoods up to a constant factor for posterior inference. Hence, we can rescale the CNN output by $P(X_i^t)$ to arrive at a surrogate likelihood term [9]. We approximate $P(X_i^t)$ by the proportion of seizure versus baseline in the dataset, i.e.

$$\hat{P}(X = 1) = \frac{\#\,\text{seizure windows}}{\#\,\text{windows}}, \qquad \hat{P}(X = 0) = 1 - \hat{P}(X = 1). \tag{5}$$

The rescaling of the discriminative posterior in Eq. (4) using the approximate prior over states in Eq. (5) will serve as the likelihood in our PGM-CNN model.

2.3 Fitting the PGM-CNN Model

We fit the PGN-CNN using a variational algorithm, similar to the one in [7]. We approximate the latent posterior as the product of independent HMM chains.

$$P(\mathbf{X} \mid \mathbf{Y}) \approx Q(\mathbf{X}) = \prod_{i=1}^{N} \frac{1}{Z_{Q_i}} Q_i(\mathbf{X}_i) = \prod_{i=1}^{N} \frac{1}{Z_{Q_i}} \prod_{t=1}^{T} T_i^t(X_i^t \mid X_i^{t-1}) E_i^t(X_i^t). \tag{6}$$

The factors $E_i^t(X_i^t)$ and $T_i^t(X_i^t \mid X_i^{t-1})$ encode the emission and transition terms of the approximating HMMs, respectively.

Posterior Inference: We infer the latent posterior distribution by iteratively running the forward-backward algorithm [10] over each of the individual chains, while holding the remaining chains constant. The forward-backward algorithm calculates the following posterior statistics under the distribution Q:

$$\tilde{\gamma}_i^t(j) \coloneqq E_{Q_i}\left[\mathbb{1}(X_i^t = j)\right] \qquad \tilde{\xi}_i^t(j,k) \coloneqq E_{Q_i}\left[\mathbb{1}(X_i^t = j, X_i^{t+1} = k)\right].$$

The variational form of $T_i^t(X_i^t \mid X_i^{t-1})$ closely resembles the original transition distribution but is now based on the statistics $\{\tilde{\gamma}_i^t(j), \tilde{\xi}_i^t(j,k)\}$. The emission parameters are the original likelihood multiplied by a correction factor $\alpha_i^t(\cdot)$.

$$E_i^t(0,2) = p(Y_i^t \mid X_i^t = 0, 2)\alpha_i^t(0), \qquad E_i^t(1) = p(Y_i^t \mid X_i^t = 1)\alpha_i^t(1)$$

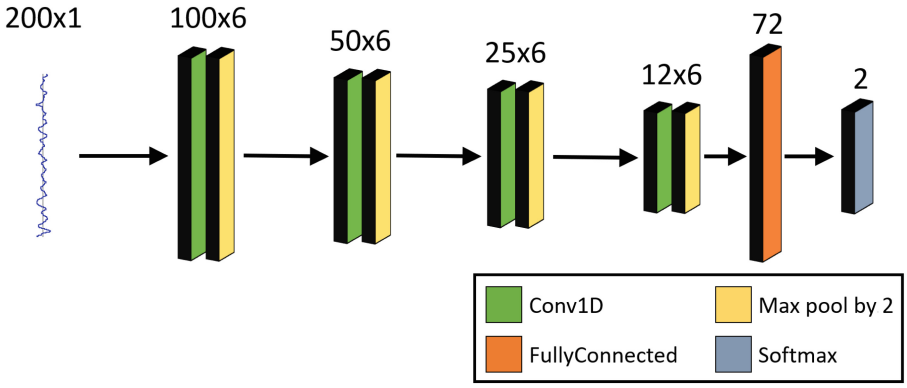


Fig. 2. Convolutional neural network architecture used in this work

$$\alpha_i^t(z) \approx \sum_{j \in au(i)} \Big[\tilde{\xi}_j^t(0,0) \left(-\rho_0 - \rho_1 \left(\nu_j^{t+1} + z\right)\right) - \tilde{\gamma}_j^t(0) \log\left(1 + e^{-\rho_0 - \rho_1\left(\nu_j^{t+1}+z\right)}\right) + \tilde{\xi}_j^t(1,1) \left(-\phi_0 - \phi_1 \left(\nu_j^{t+1} + z\right)\right) - \tilde{\gamma}_j^t(1) \log\left(1 + e^{-\phi_0 - \phi_1\left(\nu_j^{t+1}+z\right)}\right) \Big] \quad (7)$$

The factor shown in Eq. (7) encodes the influence of the aunts in the following timestep. In this term we define $\nu_i^t = \sum_{j \in au(i)} X_j^t$. We use Newton's method to learn the transition parameters $\{\rho_0, \rho_1, \phi_0, \phi_1\}$ based on the inferred $Q(\mathbf{X})$.

Neural Network Implementation: We implemented the CNN in PyTorch. The CNN consists of 4 convolution and pool layers as shown in Fig. 2. Each layer uses 6 channels with a kernel size of 5 samples and 2 sample zero padding to maintain a constant size. A LeakyReLU activation, where $\text{LeakyReLU}(x) = \max(0, x) + 0.01 \cdot \min(0, x)$, is applied at each layer. A max pooling operation with a kernel size of 2 and a stride of 2 is applied, halving the size of the representation at each layer. After the final convolution, the hidden units are concatenated and passed to a single linear layer for classification using a softmax activation.

During experimentation in the design of this network, we investigated similar architectures of varying depths, numbers of channels, and activation functions. Networks with saturating activations failed to train in some cases, perhaps due to the presence of artifact with extreme amplitudes. We found the LeakyReLU to be the most robust, likely due to the fact that it does not saturate.

The CNNs were trained discriminatively with a cross entropy loss function prior to posterior inference. We trained separate CNN classifiers for each EEG channel to capture behavior specific to different parts of the scalp. Stochastic gradient descent was performed using the Adam optimizer with a batch size of 32 samples and a learning rate of 0.5. We trained each CNN for 60 epochs, which was sufficient to achieve reliable performance without overfitting.

3 Evaluation

3.1 Baseline Comparisons

We compare our PGM-CNN detection performance to baseline methods ranging from simple classifiers on hand selected features to a fully CNN strategy. The features used in [7] (sum of spectral components and line-length) were used for all non-CNN baselines. Recordings were randomly assigned to 5-folds for cross validation. Training was performed on 4 folds while the remaining fold was used for testing. The baseline methods are summarized below.

CNN: We implement an end-to-end deep learning pipeline based on the CNN classification architecture described in Sect. 2.3. This comparison evaluates the predictive value of the CNN without the smoothing in the PGM prior.

CHMM: We implement the original CHMM model proposed in [7] which assumes a Gaussian Mixture Model (GMM) likelihood using the suggested parameter settings. This comparison will evaluate the performance gain in using a non-parametric likelihood with data driven learning from the raw EEG signal.

ANN: Similar to the CNN baseline, we evaluate the performance of the predefined features as inputs for an Artificial Neural Network (ANN) classifier. Due to the relatively small feature space we opted for a small ANN shown in Fig. 3 to avoid overfitting. Our networks are composed of two hidden layers with 10 units each. Each layer is fully connected with Rectified Linear Unit (ReLU) activations. The final output layer contains two nodes with a softmax activation applied. Thus the final layer represents the posterior probability of a hidden state given the associated feature vector.

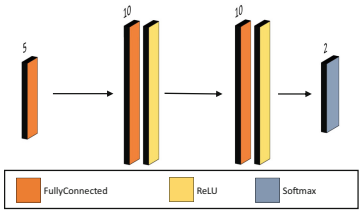
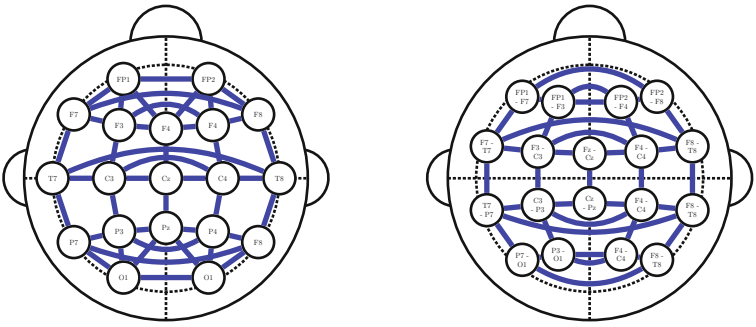


Fig. 3. Artificial neural network used for seizure detection in this work.

GMM: Finally, we implement a simple GMM classifier based on the precomputed EEG features. The inclusion of this baseline allows us to evaluate the relative performance of the CNN, ANN, and parametric GMM likelihoods directly without the inclusion of the latent seizure spreading prior.

3.2 Performance Metrics

Our performance metrics are based on the maximum a posterior (MAP) estimate of baseline versus seizure for each method and are presented as averages across test folds. Since the clinical seizure annotations tend to be overly generous and do not contain spatial information about onset, we aggregate the True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) across all windows, channels, and all seizure recordings. In general, the recordings contain muscle artifact directly following the seizure which confounds



a) Common Reference Coupling b) Longitudinal Montage Coupling

Fig. 4. Propagation paths for the (a) common reference and (b) longitudinal montage.

the offset for all methods. Therefore, we count any seizure classification occuring within the annotated seizure region as TP. However any contiguous classifications continuing past the annotated offset is not counted in our evaluation statistics.

Below we detail the summary statistics computed for each model. True Positive Rate (TPR), also known as recall, represents the total rate of correct classification. False Positive Rate (FPR) represents the rate of incorrect classification of baseline regions as seizure after excluding classifications beginning within the seizure region. We calculate a lower bound on the Area Under the Curve (AUC) using these two metrics. Precision (P) details the ratio of correct seizure classifications to the total number of seizure classifications. In addition to AUC, the F1 score offers a similar summary by computing the harmonic mean of P and TPR. Mathematically, these statistics are given by:

$$TPR = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N} (TP_i + FN_i)} \qquad FPR = \frac{\sum_{i=1}^{N} FP_i}{\sum_{i=1}^{N} (TN_i + FP_i)}$$

$$P = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N} (TP_i + FP_i)} \qquad F1 = 2\frac{P \cdot R}{P + TPR}$$

$$AUC = FPR \cdot TPR/2 + (1 - FPR)(1 + TPR)/2.$$

4 Experimental Results

Data and Preprocessing: Epileptic seizures are extremely heterogeneous. For example, generalized seizures manifest across the entire cortex at once, whereas focal seizures originate from a single area and may spread to other regions of the cortex. Given this heterogeneity, we evaluate our algorithm on two datasets. The first is taken from the Johns Hopkins Hospital (JHH) and contains 90 seizures from 15 adult patients with focal epilepsy. The second is a publicly available

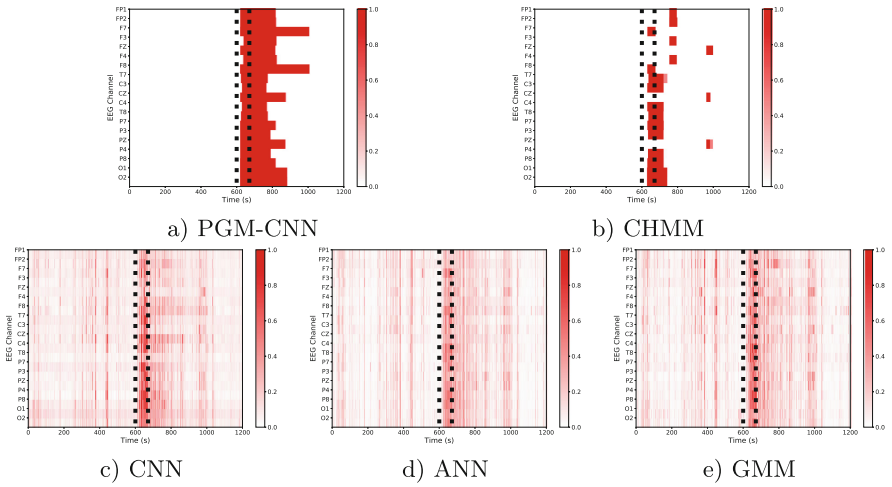


Fig. 5. Estimated posteriors for a single seizure from the JHH dataset. EEG channels are shown on the y-axis and time proceeds in the x-direction. The first row shows models with a CHMM prior. The second row shows channel-wise classifications.

Table 1. Results for the both datasets

	JHH dataset					CHB dataset				
Trial	TPR	FPR	AUC	P	F1	TPR	FPR	AUC	P	F1
PGM-CNN	**0.45**	0.010	**0.72**	0.79	**0.57**	**0.61**	0.013	**0.80**	0.74	**0.67**
CHMM	0.37	0.0083	0.68	**0.80**	0.50	0.571	**0.0067**	0.78	**0.83**	**0.67**
CNN	0.19	0.010	0.59	0.62	0.28	0.27	0.0071	0.63	0.70	0.39
DNN	0.11	**0.0070**	0.55	0.58	0.18	0.23	0.0071	0.61	0.66	0.34
GMM	0.18	0.015	0.58	0.52	0.27	0.26	0.010	0.62	0.61	0.37

dataset from Children's Hospital Boston (CHB) of unspecified epilepsy types [4]. We used 185 recordings from this dataset from 24 pediatric patients.

Besides the patient populations, another difference between the two datasets is the acquisition protocol. The JHH dataset contains the original recordings of the 10/20 international system in common reference. In contrast, the CHB dataset uses the longitudinal montage, which forms difference channels by subtracting the signals in neighboring electrodes. We specify a propagation network appropriate for this montage as shown in Fig. 4b. This coupling preserves neighboring and contralateral relationships on the scalp from the original prior.

Our recordings contain one seizure and up to ten minutes of pre- and post-seizure baseline. For the CHB data, these segments were clipped from the original release. EEG channels were low and highpass filtered at 50 and 1.6 Hz, respectively. A notch filter at 60 Hz was applied to remove any remaining power supply artifact. As in [7], 4 spectral features and one line-length feature were extracted

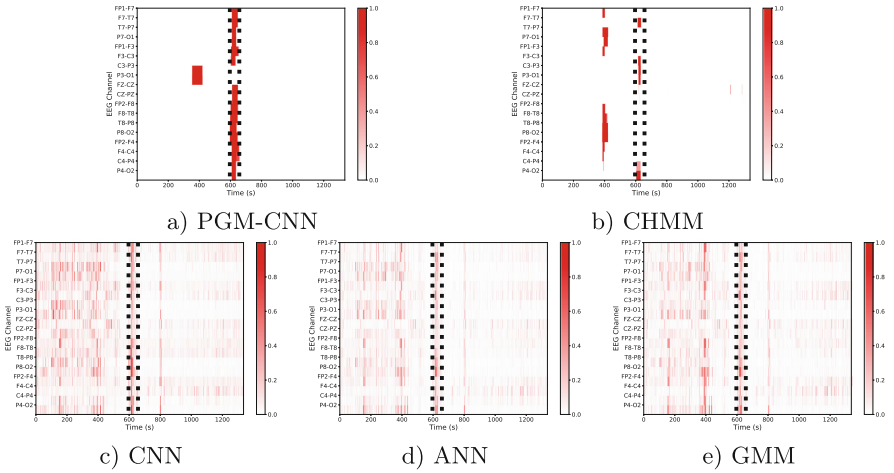


Fig. 6. Example posteriors from the CHB dataset. CHMM and likelihood models are shown in the first and second rows, respectively.

from 1 s windows with a 250 ms overlap. The CNN model was trained directly on the raw EEG signal from the 1 s windows.

Detection Performance. Table 1 reports the seizure detection performance averaged across the testing folds for both the JHH and CHB datasets. We have reported True Positive Rate (TPR), False Positive Rate (FPR), Area Under the Curve (AUC), Precision (P), and F1, as described in Sect. 3.2.

Our PGM-CNN dramatically outperforms all of the baseline methods on the JHH dataset. The only drawback is a slightly higher FPR, since our CNN shows more sensitivity to seizure activity, and classifies slightly more baseline as seizure. Despite the numerical increase in FPR, the increased sensitivity is valuable in the clinic, particularly when augmenting the expert manual inspections. Moreover, these spurious detections are compensated by more accurate true detections, which are reflected in the AUC, precision, and F1 measures. We emphasize that our evaluation metrics are much more conservative than in prior studies, which is why the TPR seems uniformly low. Instead of measuring singular detections within the annotated seizure period, we aggregate over channels and windows. This allows us to evaluate not only correct detections of seizures but *how much seizure activity* our algorithms are capable of discerning.

Interestingly, the same detection trends are seen in the CHB data, despite our PGM spreading prior being designed for focal and not generalized seizures. The PGM-CNN achieves the best TPR and AUC as well as a comparable F1 score. In short our flexible data likelihood based on the CNN allows us to learn complex data representations that better separate seizure from baseline. This leads to better detection rates, which is valuable for clinical planning.

Finally, we note that the channel-wise baselines are uniformly bad. Detecting seizure activity is, in general, a relatively difficult problem. Both seizure and

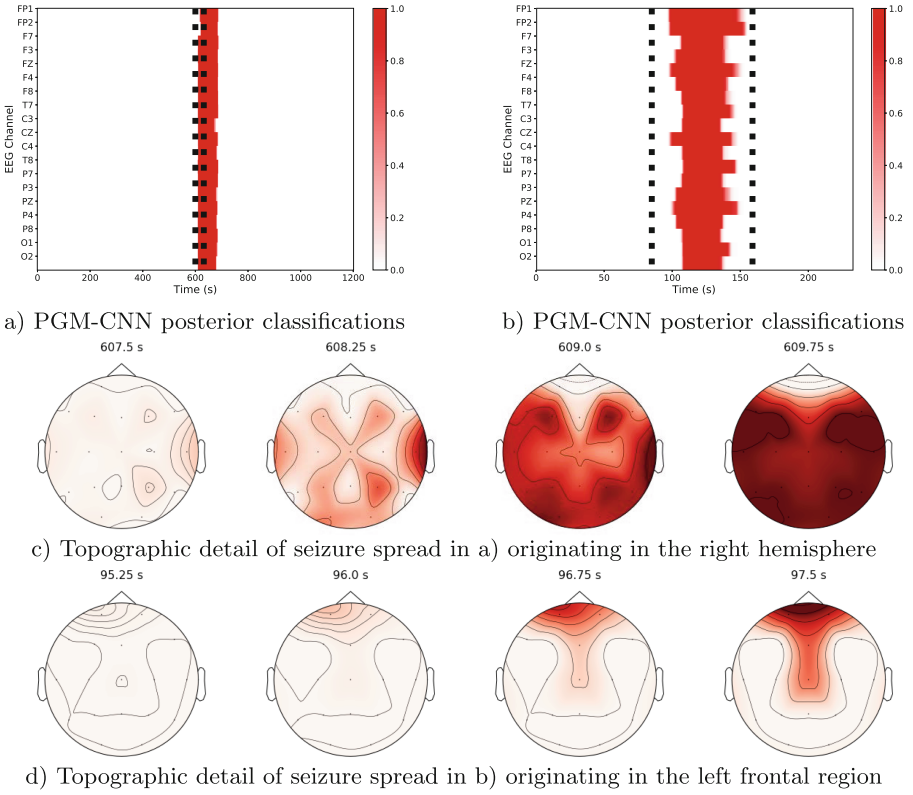


Fig. 7. Example seizure tracking from the JHH dataset. (a, b) Posteriors for all channels. (c, d) Topographic detail showing posterior onsets in clinically annotated regions.

baseline contain high amplitude muscle artifact, which confound the detection over short time windows. In addition, the data distributions are highly overlapping, with seizure activity often resembling normal behavior. The effect of the prior in the PGM-CNN and CHMM for data fusion across channels is apparent.

Figures 5 and 6 show the classification posteriors of each model for an example seizure in the JHH and CHB datasets, respectively. EEG channels are presented on the y-axis while the x-axis shows time. The dashed black lines correspond to annotations for seizure onset and offset. Red indicates the posterior probabilities of the seizure state. The PGM-CNN correctly classifies more of the annotated seizure in Fig. 5 than any of the other models. Each model incorrectly activates during the period immediately following the seizure, responding to the presence of artifact. However, the CNN likelihood model places more confidence in the seizure region, allowing for more correct classification. In contrast, the ANN and GMM identify strong seizure-like activity in the artifact following the actual seizure, causing an incorrect classification by the CHMM [7]. In Fig. 6 the PGM-CNN correctly classifies more of the annotated seizure than the CHMM but

makes a false positive, while the CHMM classifies only a small portion of the seizure and responds strongly to the artifact prior to the seizure.

Seizure Localization: Surgical resection is the standard-of-care for medically refractory focal epilepsy. The latent propagation prior of our PGM-CNN has the potential to aid in seizure localization. Figure 7 shows two classifications from the PGM-CNN. Clinical annotations for the seizure in (a) and (c) suggest an origin in the right temporal lobe and spreading left. Likewise, the annotations in (b) and (d) suggest a left frontal lobe onset. The localization information provided by our model agrees with the annotated foci. Remarkably, this spreading behavior is *learned in a completely unsupervised manner* based on the clinical hypotheses embedded in the PGM prior. This result highlights the promise of integrating model-based and data-driven approaches for medical imaging applications.

5 Conclusion

We have presented the first generative model-deep learning hybrid for epileptic seizure detection. Our framework captures the spatio-temporal spread of a seizure through a structured PGM prior, while allowing for a complex likelihood function that is implicitly learned via a CNN. This data driven approach learns representations directly from the raw EEG signal, improving upon feature extraction techniques. At the same time the PGM preserves clinical interpretability and acts as a local smoothing process for the CNN outputs based on limited training examples. We evaluate our method on clinical data from two hospitals with distinct patient populations. In both cases, our PGM-CNN achieved higher true positive detection and AUC than any of the baseline methods.

Future work will explore alternate deep learning architectures with larger receptive fields and evaluate the effectiveness of training multichannel CNNs for fusing information across the scalp. In addition, modeling improvements such as restrictions on allowed onsets and enforcement of concurrent offsets across channels would likely reduce false positives and is an ongoing direction of work.

Acknowledgments. This work was supported by a JHMI Synergy Award (Venkataraman/Johnson) and NSF CAREER 1845430 (Venkataraman).

References

1. Zach, M., et al.: National and state estimates of the numbers of adults and children with active epilepsy - United States, 2015. CDC MMWR **66**, 821–825 (2017)
2. French, J.A.: Refractory epilepsy: clinical overview. Epilepsia **48**, 3–7 (2007)
3. Osorio, I., Zaveri, H.P., Frei, M.G., Arthurs, S.: Epilepsy: the Intersection of Neurosciences, Biology, Mathematics, Engineering, and Physics. CRC Press, Boca Raton (2016)
4. Shoeb, A.H., Guttag, J.V.: Application of machine learning to epileptic seizure detection. In: International Conference on Machine Learning, pp. 975–982 (2010)

5. Andrzejak, R.G., et al.: Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. Phys. Rev. E **64**(6), 061907 (2001)
6. Güler, N.F., et al.: Recurrent neural networks employing Lyapunov exponents for EEG signals classification. Expert Syst. Appl. **29**(3), 506–514 (2005)
7. Craley, J., Johnson, E., Venkataraman, A.: A novel method for epileptic seizure detection using coupled hidden markov models. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11072, pp. 482–489. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00931-1_55
8. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge (2016). http://www.deeplearningbook.org
9. Hinton, G., et al.: Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Process. Mag. **29**(6), 82–97 (2012)
10. Murphy, K.P.: Machine Learning: A Probabilistic Perspective. MIT Press, Cambridge (2012)