

# A Spatio-Temporal Model of Seizure Propagation in Focal Epilepsy

Jeff Craley, *Student Member, IEEE*, Emily Johnson, and Archana Venkataraman, *Member, IEEE*,

**Abstract**—We propose a novel Coupled Hidden Markov Model (CHMM) to detect and localize epileptic seizures in clinical multichannel scalp electroencephalography (EEG) recordings. Our model captures the spatio-temporal spread of a seizure by assigning a sequence of latent states (i.e. baseline or seizure) to each EEG channel. The state evolution is coupled between neighboring and contralateral channels to mimic clinically observed spreading patterns. Since the latent state space is exponential, a structured variational algorithm is developed for approximate inference. The model is evaluated on simulated and clinical EEG from two different hospitals. One dataset contains seizure recordings of adult focal epilepsy patients at the Johns Hopkins Hospital; the other contains publicly available non-specified seizure recordings from pediatric patients at Boston Children’s Hospital. Our CHMM model outperforms standard machine learning techniques in the focal dataset and achieves comparable performance to the best baseline method in the pediatric dataset. We also demonstrate the ability to track seizures, which is valuable information to localize focal onset zones.

**Index Terms**—Seizure detection, focal epilepsy, coupled hidden Markov models, variational inference, electroencephalography

## I. INTRODUCTION

EPILEPSY is a heterogeneous neurological disorder characterized by recurrent and unprovoked seizures [1]. Epilepsy affects between 1–3% of the world’s population, making it one of the most prevalent neurological disorders. While epilepsy can often be controlled with medication, it is estimated that 20–40% of patients are medically refractory and do not respond to anti-epileptic drugs [2]. Alternative therapies for these patients rely on our ability to detect and localize epileptic seizures in the the brain. Epileptic seizures can be broadly characterized as either focal or generalized. Generalized seizures manifest simultaneously across the cortex. Conversely, focal seizures originate in a specific onset zone, but may subsequently spread to neighboring regions of the brain until potentially the entire cortex is involved [3]. In medically refractory focal epilepsy, resection of the onset zone may be the only treatment available to completely eliminate seizures.

Manuscript submitted January 23, 2019; revised August 30, 2019. This work was supported by the JHH synergy award (PIs: Johnson/Venkataraman) and the National Science Foundation CAREER award 1845430 (PI: Venkataraman).

J. Craley and A. Venkataraman are with the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, 21218 USA email: (jcraley2@jhu.edu, archana.venkataraman@jhu.edu)

E. Johnson is with the School of Medicine, Johns Hopkins University, Baltimore, MD, 21205 USA email: (ejohns92@jhmi.edu)

Copyright (c) 2019 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

Scalp EEG can provide information about whether the seizures are focal or generalized. In the case of focal seizures, it can be used to coarsely localize the onset zone. Scalp EEG is a natural complement to other noninvasive imaging modalities, such as PET [4] and MRI [5], which can be used to refine the localization. These modalities have the advantage of a higher spatial resolution, but they are more costly to acquire. After multimodal localization is performed using EEG and other noninvasive imaging, the extent of the onset zone may be identified using more invasive techniques, such as electrocorticography (ECoG), just prior to surgical resection. EEG studies play a critical role in this process, affording clinicians a noninvasive and cost efficient means of establishing early information necessary for treatment planning.

Scalp EEG recordings are typically acquired over the course of several days after any medication is withdrawn. Continuous EEG monitoring is performed over several days until an adequate number of seizures are recorded. Visual inspection of the EEG recordings remains the standard procedure for seizure detection. This process is time intensive, requires extensive training, and has the potential to miss important events.

### A. Prior Work on Epileptic Seizure Detection

Automated seizure detection in clinical EEG recordings has been under investigation since the early 1980s [6]. Interested readers can find a survey of a variety of techniques in [7] and [8]. Broadly, most seizure detection algorithms follow a standard machine learning pipeline. First the EEG signals are divided into windowed epochs, from which a set of features are extracted. Next, a classifier is trained on these features to declare each epoch as seizure or baseline [9]. This approach is exemplified by the work of [10], where the power in different spectral bands are used in conjunction with a Support Vector Machine (SVM) on two-second epochs of EEG data. Seizure detectors are trained in one of two manners: patient-specific or patient-agnostic, the latter of which can be used for a broad patient cohort. Intuitively, patient-specific seizure detectors tend to outperform patient-agnostic ones. However, they rely on having many good quality seizure recordings per patient, which is not always possible in a clinical setting. In contrast, patient-agnostic detectors are easier to train but tend to suffer in heterogeneous cohorts.

Much of the research in seizure detection has focused on identifying features that capture the specific morphologies of *ictal*, or seizure related, EEG. However, this task is complicated by the presence of artifacts in the EEG signal, which are often of greater magnitude than the true neural measurements.

In addition, many ictal morphologies closely resemble normal EEG patterns. For example, the rhythmic sinusoidal waves that can indicate the onset of a seizure share similarities with the posterior dominant rhythm present in normal EEG [11].

Given the above challenges, prior work in EEG feature extraction has studied a diverse range of signal properties. These strategies fall broadly into time domain and time-frequency domain approaches. Time domain methods extract signal statistics, such as power, skewness, kurtosis, and maximum amplitude, from the raw EEG signal as further described in [12]. Techniques from non-linear signal processing and chaos theory are an important sub-class of time domain features and have also received significant attention in the seizure detection literature. For example, the work of Andrzejak et al. identified differences in the non-linear dynamics of EEG signals in ictal versus non-ictal periods [13]. Other examples include the application of Lyapunov exponents [14] and approximate entropy [15], which measure the complexity in the evolution of non-linear dynamical systems.

Time-frequency decompositions have become increasingly popular in the seizure detection literature. From a biological standpoint, brain waves are believed to fall within functionally specialized frequency bands: theta (1–4 Hz), delta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma ( $\geq 30$  Hz) bands. By construction, time-frequency decompositions can isolate activity in these different frequency bands that may be relevant to the seizure onset. For example, the work of [16] used the relative power in these brain wave bands across EEG channels to perform patient-specific seizure detection. In contrast, [10] took a more data-driven approach by using evenly spaced band-pass filters to compute relative power across the frequency spectrum. The power in each sub-band is then fed into an SVM for seizure vs. non-seizure classification. Similarly, in [17] the Fourier transform was applied to calculate the power in the 2.5–12 Hz range, achieving accurate seizure detection in a patient specific-context.

Recently, wavelet based methods, such as the Discrete Wavelet Transform (DWT) and wavelet packet decompositions, have gained in popularity due to the dyadic band-pass nature of the sub-band decomposition. Depending on the sampling rate, these sub-bands may naturally align with the clinically relevant brain wave bands. For example, the works of [18] and [19] extract simple features, such as amplitude, energy, and coefficient of variation from each wavelet sub-band for use in seizure classification.

To better leverage their advantages, features from the time domain and time-frequency domain are often combined to produce better detection performance. For example, the work of [20] compares the individual performance of 65 time-domain and frequency-domain features; it identifies line-length and relative energy in the 12.5–25 Hz as the most robust features. Likewise, fractal dimension and energy in each band of the wavelet packet decomposition were combined for patient-specific classification in [21]. A similar approach was used in [12] where time domain features are combined with Fourier domain features and by [22], who added wavelet features to the mix. Finally, the works of [23], [24], [25] extract nonlinear signal characteristics following the wavelet decomposition.

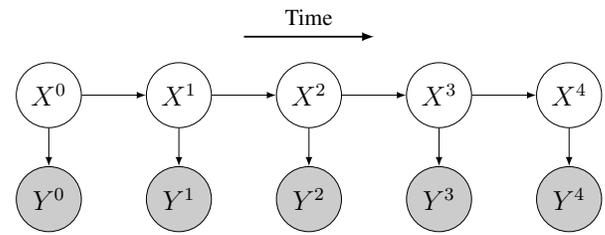


Fig. 1. Graphical model depicting an HMM. Observed variables,  $Y^t$ , are shown shaded. Hidden variables,  $X^t$ , are unshaded.

Almost as broad as the signal processing techniques applied to the feature selection process are the classifiers used to detect seizure versus non-seizure activity. The SVM has enjoyed wide popularity in the seizure detection literature, especially for patient-specific classification [16], [26]. Random forest (RF) classifiers have also enjoyed success [12], [27]. Other classifiers employed in the literature include k-nearest neighbors [27], relevance vector machines, and adaptive thresholding [17]. Recent investigation has applied deep learning to seizure detection. In [27] multi-layer perceptrons were compared to an array of other classifiers, achieving comparable performance.

### B. HMM Background

A complementary approach to frame-wise feature extraction is to model the temporal evolution of the EEG signal. The Hidden Markov Model (HMM) is a popular sequence model used often in time series applications, where an evolving latent state governs the emission of observed variables. HMMs are popular in speech, natural language processing, bioinformatics, and genomics applications [28], [29].

A graphical model depicting the HMM is shown in Fig. 1. The model consists of observed nodes  $Y^t$  and hidden nodes  $X^t$  for times  $t = 0, \dots, T$ . Nodes  $X^t$  form a Markov chain of discrete states. At every timestep, an observed emission  $Y^t$  is generated from an emission likelihood  $P(Y^t | X^t)$ , i.e.  $Y^t$  is conditionally independent of all other variables in the model given the latent state of the Markov chain at  $t$ . Given a distribution over the initial state  $X^0$ , the joint probability distribution of ensemble variables  $\mathbf{X}$  and  $\mathbf{Y}$  factorizes as:

$$P(\mathbf{X}, \mathbf{Y}) = P(Y^0 | X^0)P(X^0) \prod_{t=1}^T P(Y^t | X^t)P(X^t | X^{t-1}).$$

The conditional distribution  $P(X^t | X^{t-1})$  governing transitions between latent states is often expressed by a stochastic transition matrix  $A$ , where  $P(X^t = j | X^{t-1} = i) = a_{i,j}$ .

The forward-backward algorithm [28], [29] is used for exact marginal inference in the HMM. At a high level, this algorithm uses dynamic programming to propagate information back and forth along the HMM chain using forward and backward message passing. More details about this procedure can be

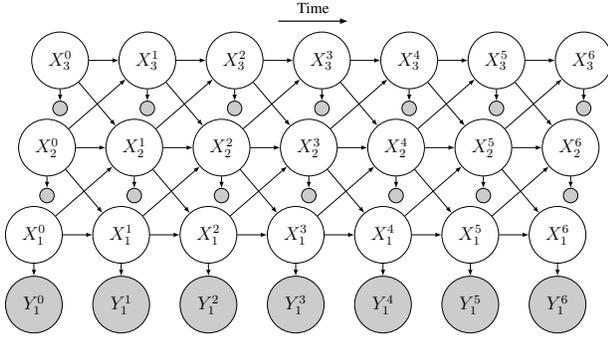


Fig. 2. Graphical model depicting a CHMM. Observed variables,  $Y_i^t$ , are shown shaded. Hidden variables,  $X_i^t$ , are unshaded.

found in [28], [29]. The forward messages  $\alpha_i^t(k)$  and backward messages  $\beta_i^t(k)$  are computed via the following recursions:

$$\begin{aligned} \alpha^t(i) &:= P(Y^0, Y^1, \dots, Y^t, X^t = i) \\ &= \sum_j P(Y^t | X^t = i) a_{j,i} \alpha^{t-1}(j) \\ \beta^t(i) &:= P(Y^{t+1}, Y^{t+2}, \dots, Y^T | X^t = i) \\ &= \sum_j \beta^{t+1}(j) P(Y^{t+1} | X^{t+1}) a_{i,j}. \end{aligned}$$

The data likelihood can be easily calculated by noting that  $P(Y^0, Y^1, \dots, Y^t) = \sum_j \alpha^t(j)$ . The singleton and pairwise marginals,  $\gamma^t(i)$  and  $\xi^t(i, j)$  respectively, are obtained by normalizing the following expressions:

$$\begin{aligned} \gamma^t(i) &= P(X^t = i | \mathbf{Y}) \propto \alpha^t(i) \beta^t(i) \\ \xi^t(i, j) &= P(X^t = i, X^{t+1} = j | \mathbf{Y}) \\ &\propto \alpha^t(k) a_{i,j} P(Y^{t+1} | X^{t+1} = j) \beta^{t+1}(j). \end{aligned}$$

### C. Bayesian Models for EEG Analysis

Extensions to the HMM have been applied to several problems in EEG analysis. For example, the work of [30] develops an Autoregressive HMM (AR-HMM) to model the changing correlation structure in raw EEG data as an unsupervised way to discover different dynamical regimes in the EEG signal. However, no labels were used in training the model, so expert labeling was required to apply the learned states to related problems. For example, to apply the outcome of this model to seizure prediction in canine EEG, the pre-seizure states had to be manually identified from the AR-HMM output.

The Coupled Hidden Markov Model (CHMM) extends the original HMM formulation to include multiple latent chains. Within the context of EEG, each latent chain corresponds to a single EEG electrode. Coupling is defined such that each latent chain may be affected by and may influence the states of other chains. CHMMs made their debut in modeling audio-visual relationships [31]. In another domain, the work of [32] proposed a CHMM to model the spread of infectious disease by defining a coupling structure based on the physical proximity of individual people.

Small two channel CHMMs have been used in behavioral EEG experiments [33]. In addition, the work of [34] developed

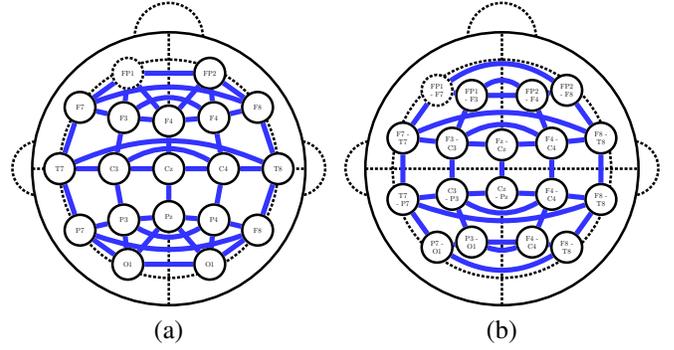


Fig. 3. Electrode placement in the 10/20 international system [35] with seizure propagation pathways shown in blue. The edges in the graph indicate conditional independences in between nodes in consecutive timesteps of our model. (a) Graph defined on the common average montage. (b) Graph defined for the longitudinal montage.

a distance coupled HMM to model EEG signals in both alcoholics and their healthy peers. Classification was performed between groups by assigning a test sequence to the model class under which its likelihood was maximum. While capable of discerning patients from controls, these models were designed to classify entire EEG sequences and not to label pathological activity within a single recording. In addition, these works simplify the analysis to just two EEG channels, for which exact inference using the forward-backward algorithm is tractable. However, the methodology outlined in [34] does not generalize to more dense recordings.

### D. Our Contribution

In this work we present a high-dimensional multichannel CHMM for seizure detection from scalp EEG. Our CHMM fuses information from the individual EEG channels via a spatio-temporal model of seizure spreading. The CHMM model is evaluated on two clinical datasets to demonstrate its superior performance over standard approaches. A preliminary version of this work was introduced in [36]. This version provides a more complete description of inference and learning with further algorithmic refinements which improve performance. Additional real-world and simulated experiments are provided as well.

The latent chains in our CHMM correspond to the standard electrode placement locations and represent the key areas of interest on the scalp. Interactions between EEG channels are coupled such that as channels enter the seizure state, their neighboring and contralateral electrodes are more likely to also enter a seizure state. Due to the high dimensional state space, exact inference in this model is intractable. Thus a structured mean field variational inference algorithm is developed to compute the latent posterior distributions of seizure activity and infer the seizure spreading characteristics.

We evaluate the performance of our model on simulated data and on two clinical EEG datasets from inpatient monitoring. The first testbed consists of 15 focal epilepsy patients from the Johns Hopkins Hospital. The second is drawn from the publicly available Children’s Hospital of Boston (CHB) dataset [10], which contains pediatric patients with both focal and general epilepsy. Spectral power and line length are utilized

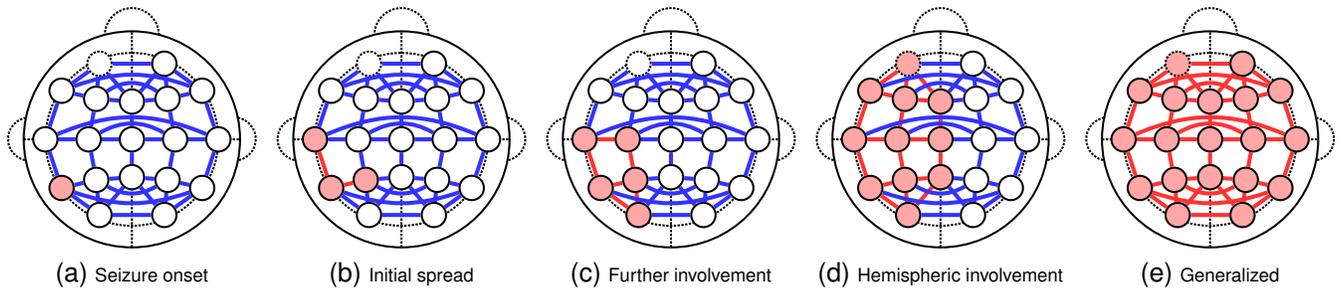


Fig. 4. Hypothetical spreading of a focal seizure. (a) A seizure originates in a single channel. (b) The seizure propagates to neighboring EEG channels. (c) Further spreading progresses to involve more EEG channels. (d) The left hemisphere is involved. (e) The seizure becomes generalized to the entire scalp.

as observed features in our CHMM. These features are simple yet robust variants of popular features in the seizure detection literature, which we have observed to perform well.

The generative framework is discussed in Section II. Section III contains an overview of the variational inference and Expectation-Maximization (EM) procedure for fitting the model to data. Sections IV and V present our results using simulated and real-world EEG data, respectively. In Section VI the implications of our results are discussed and future directions for our model are suggested. Section VII reviews the findings of our experimentation and concludes the paper.

## II. GENERATIVE MODEL OF SEIZURE PROPAGATION

This section details the generative process governing our CHMM based seizure detection algorithm. Fig. 2 illustrates the graphical model for a generic three chain CHMM. Here, observed emissions  $Y_i^t$  for chain  $i$  at time  $t$  remain conditionally independent given latent states  $X_i^t$ . The transition prior factorizes such that  $P(\mathbf{X}^t | \mathbf{X}^{t-1}) = \prod_{i=1}^N P(X_i^t | \mathbf{X}^{t-1})$  where  $N$  denotes the number of chains. Accompanying these latent states are observed emission variables  $\mathbf{Y}_i$  which are conditionally independent given the latent variables.

Unlike past work, we design a transition prior  $P(\mathbf{X}^t | \mathbf{X}^{t-1})$  capable of tracking the spread of seizures in focal epilepsy. In this prior, each electrode channel is represented by a chain of latent states  $\mathbf{X}_i$  where  $i$  indexes the EEG electrode. These states represent the current seizure vs. baseline label of the signal at each electrode. The observations  $\mathbf{Y}_i$  represent the features calculated from channel  $i$  of the EEG data.

Mathematically, our transition prior includes a contribution from all other chains via the above factorization. The observed emissions  $Y_i^t$  for chain  $i$  at time  $t$  remain conditionally independent given the latent states  $X_i^t$ . In the rest of this paper,  $X_i^t$  and  $Y_i^t$  will denote latent and observed random variables, respectively, for a single chain  $i$  at timestep  $t$ . Variable ensembles are bolded, with the corresponding subscript or superscript dropped. Thus  $\mathbf{X}_i = \{X_i^0, X_i^1, \dots, X_i^T\}$  refers to the ensemble of latent states from chain  $i$ ,  $\mathbf{X}^t = \{X_1^t, X_2^t, \dots, X_N^t\}$  refers to the latent states for all chains at time  $t$ , and  $\mathbf{X}$  refers to all latent variables. Similarly, the subscript  $-i$ , as in  $\mathbf{X}_{-i}$ , will indicate a collection of random variables taken over all chains excluding those of chain  $i$ . For convenience, Table I defines symbols for random variables and non-random parameters used in our model.

### A. Transition Prior

At the heart of our model is a graph which encodes the clinically informed spreading of a focal seizure. Connections between the latent chains of the CHMM, as illustrated in Fig. 2, are constructed according to this propagation graph. This graph is defined in the sensor space, using the common average and bipolar montage from the 10/20 international system [35]. The bipolar montage is popular with neurologists for tracking phase changes in the raw EEG signal. For both the common reference signals and the longitudinal bipolar montage, we define a network  $\mathcal{S}$  of seizure propagation by connecting neighboring and contralateral EEG channels. These graphs are shown in Fig. 3. Neighboring connections capture local seizure spreading between adjacent EEG channels. Contralateral connections account for seizure activity that appears to manifest simultaneously on each hemisphere. An example of the seizure spreading our model encodes is shown in Fig. 4.

For notational convenience, let  $au(i)$  be the set of aunt indices of a node  $i$  as the neighbors in the graph  $\mathcal{S}$ , i.e.  $au(i) = \{j | j \in ne_{\mathcal{S}}(i)\}$ . Our transition prior now simplifies to

$$P(X_i^t | \mathbf{X}^{t-1}) = P\left(X_i^t | X_i^{t-1}, \mathbf{X}_{au(i)}^{t-1}\right). \quad (1)$$

As seen in (1), transitions in channel  $i$  depend only on the previous state of the chain and the previous states of chains  $au(i)$ . This notation is used to avoid potential ambiguities between the terms “neighbors” and “parents” in the directed and undirected graphical modeling literature. Furthermore, it allows us to distinguish between the previous state in chain  $i$ ,  $X_i^{t-1}$ , and the previous states of the neighboring electrodes,  $\mathbf{X}_{au(i)}^{t-1} := \{X_j^{t-1}\}_{j \in au(i)}$ , more compactly.

A three state left-to-right time inhomogenous transition matrix is used to encode the probability of transitions between latent states. States 0 and 2 represent pre- and post-seizure baseline while state 1 represents a seizure as shown below.

$$A_i^t = \begin{bmatrix} 1 - g_i^t & g_i^t & 0 \\ 0 & 1 - h_i^t & h_i^t \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

$$\begin{aligned} \log\left(\frac{g_i^t}{1 - g_i^t}\right) &= \rho_0 + \rho_1 \eta_i^t \\ \log\left(\frac{h_i^t}{1 - h_i^t}\right) &= \phi_0 + \phi_1 \eta_i^t \end{aligned} \quad (3)$$

TABLE I  
RANDOM VARIABLES (TOP) AND NON-RANDOM PARAMETERS (BOTTOM) IN OUR GRAPHICAL MODEL SHOWN IN FIGS. 2 AND 3

$X_i^t$	Latent state in chain $i$ at time $t$
$Y_i^t$	Observed variables for chain $i$ at time $t$
$\eta_i^t$	Sum of the aunt nodes for channel $i$ at time $t$
$A_i^t$	Transition matrix for chain $i$ at time $t$
$g_i^t$	Probability of seizure onset in chain $i$ at time $t$
$h_i^t$	Probability of seizure offset in chain $i$ at time $t$
$\mathcal{S}$	Connectivity graph between electrode channels
$\rho_0$	Seizure onset parameter
$\rho_1$	Seizure onset spread parameter
$\phi_0$	Seizure offset parameter
$\phi_1$	Seizure offset spread parameter
$\pi_{im}^k$	Emission mixture weight for mixture $m$ for chain $i$ for state $X_i^t = k$
$\mu_{im}^k$	Emission mean for mixture $m$ for chain $i$ for state $X_i^t = k$
$\Sigma_{im}^k$	Emission covariance for mixture $m$ for chain $i$ for state $X_i^t = k$

State 2 is the final state and, once entered, the chain remains there for the duration of the recording. The transition probabilities in (2) are computed via the logistic functions in (3) based on the the neighboring state assignments. Here, let  $\eta_i^t$  be defined as the number of aunt nodes in the seizure state in the previous timestep, i.e.  $\eta_i^t := \sum_{j \in au(i)} \mathbf{1}(X_j^{t-1} = 1)$ . The parameters  $\rho_0$  and  $\rho_1$  represent the base parameter and aunt influence, respectively, for transitions from pre-seizure to seizure. Namely, at any timestep, there is a small base probability that a channel that has not transitioned into a seizure state may enter one. We expect to learn a positive value for  $\rho_1$  indicating more aunts in a seizure state will encourage a transition into seizure. Similarly,  $\phi_0$  and  $\phi_1$  represent corresponding parameters for the transition out of the seizure state into the post-seizure baseline.

### B. Emission Likelihood

Emission likelihoods  $P(Y_i^t | X_i^t)$  are modeled using Gaussian Mixture Models (GMMs). Let  $M$  be the number of mixtures. The parameter  $\pi_{ij}^k$  is the weight of mixture  $m$  in chain  $i$  when  $X_i^t = k$ . Likewise,  $\mu_{im}^k$  and  $\Sigma_{im}^k$  are the mean and covariance, respectively, for mixture  $m$  and chain  $i$  when  $X_i^t = k$ , i.e.,

$$P(Y_i^t | X_i^t = k) = \sum_{m=1}^M \pi_{im}^k \mathcal{N}(Y_i^t; \mu_{im}^k, \Sigma_{im}^k). \quad (4)$$

As seen, the likelihood of the emission variable  $Y_i^t$  is the weighted sum of Gaussian densities with weights  $\pi_{im}^k$ . For simplicity, we tie the parameters for the pre- and post-seizure baseline states, i.e.  $\pi_{im}^0 = \pi_{im}^2$ ,  $\mu_{im}^0 = \mu_{im}^2$ , and  $\Sigma_{im}^0 = \Sigma_{im}^2$ .

## III. INFERENCE AND LEARNING

The joint distribution of our CHMM can be written as

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

Notice that our transition prior allows all possible latent configurations, which amounts to  $3^{19}$  states under the 19 channel 10/20 system. Due to the high dimensionality of this latent space, exact inference is intractable. However, the structure of our model lends itself well to approximation by variational inference. We develop a structured mean field algorithm that approximates the latent posterior probability  $P(\mathbf{X} | \mathbf{Y})$  using separate independent HMM chains for each channel.

A variational EM algorithm [37] is used to fit our model to the observed data. This algorithm alternates between an Expectation (E) step that computes current posterior beliefs of the latent seizure states given fixed values of the likelihood and transition parameters. The Maximization (M) step updates the model parameters according to these beliefs. The E- and M-steps are iterated until convergence, to obtain both the model parameters and the marginal posterior beliefs. The following subsections outline the E-step, M-step, initialization, and training of the model.

### A. E-step: Variational Inference

Structured mean field variational inference is performed by defining an analytically tractable family of approximating distributions  $\mathcal{Q}$  and minimizing an upper bound on the data negative log-likelihood, known as the variational free energy:

$$\mathcal{F}\mathcal{E} := -E_Q[\log P(\mathbf{X}, \mathbf{Y})] + E_Q[\log Q(\mathbf{X})] \geq -\log P(\mathbf{Y}). \quad (5)$$

The bound in (5) is derived via Jensen's inequality. Notice that the distribution  $Q$  that minimizes the free energy also minimizes the KL divergence between the approximating distribution and the true posterior distribution, i.e.,  $D(Q(\mathbf{X}) \| P(\mathbf{X} | \mathbf{Y}))$ . Said another way, this variational inference process finds the closest distribution  $Q \in \mathcal{Q}$  to the posterior  $P(\mathbf{X} | \mathbf{Y})$  in an information theoretic sense.

Let the family of approximating distributions  $\mathcal{Q}$  for the CHMM be the product of  $N$  independent HMM chains across

the dynamic latent states  $\mathbf{X}_i$  for each of the  $N$  EEG channels:

$$\begin{aligned} Q(\mathbf{X}) &= \prod_{i=1}^N \frac{1}{Z_{Q_i}} Q_i(\mathbf{X}_i) \\ &= \prod_{i=1}^N \frac{1}{Z_{Q_i}} \prod_{t=1}^T \psi_i^t(X_i^t, X_i^{t-1}) \omega_i^t(X_i^t). \end{aligned} \quad (6)$$

The distribution of each chain in (6) is defined by singleton factors  $\omega_i^t(X_i^t)$  and pairwise factors  $\psi_i^t(X_i^t, X_i^{t-1})$ . Substituting this approximating distribution into (5) allows us to decompose  $\mathcal{F}\mathcal{E}$  into terms dependent on chain  $i$  and those dependent on all the other chains, denoted  $-i$ .

$$\begin{aligned} \mathcal{F}\mathcal{E} &= -E_{Q_i} [E_{Q_{-i}} [\log p(\mathbf{X}_i, \mathbf{Y}_i | \mathbf{X}_{-i}, \mathbf{Y}_{-i})]] \\ &\quad + E_{Q_i} [\log Q_i(\mathbf{X}_i)] \\ &\quad - E_{Q_{-i}} [\log p(\mathbf{X}_{-i}, \mathbf{Y}_{-i})] + E_{Q_{-i}} [\log Q_{-i}(\mathbf{X}_{-i})] \\ &= -E_{Q_i} [E_{Q_{-i}} [\log p(\mathbf{X}_i, \mathbf{Y}_i | \mathbf{X}_{au(i)})]] \\ &\quad + E_{Q_i} [\log Q_i(\mathbf{X}_i)] + \text{constant} \end{aligned} \quad (7)$$

In this substitution, we have used the factorization  $p(\mathbf{X}, \mathbf{Y}) = p(\mathbf{X}_{-i}, \mathbf{Y}_{-i}) p(\mathbf{X}_i, \mathbf{Y}_i | \mathbf{X}_{au(i)})$  to isolate terms pertaining to chain  $i$  as in the last two lines of (7). Notice that this factorization leads to a natural coordinate descent algorithm. Namely, by holding chains  $-i$  constant and minimizing the free energy with respect to chain  $i$ , the upper bound on the negative log-likelihood can be iteratively refined. Since  $\mathcal{F}\mathcal{E}$  is bounded from below (i.e., it cannot diverge to  $-\infty$ ), this coordinate descent procedure is guaranteed to converge to a local optima of the free-energy objective.

$$\begin{aligned} \arg \min_{Q_i} \mathcal{F}\mathcal{E} &= \arg \min_{Q_i} -E_{Q_i} [E_{Q_{-i}} [\log p(\mathbf{X}_i, \mathbf{Y}_i | \mathbf{X}_{au(i)})]] \\ &\quad + E_{Q_i} [\log Q_i(\mathbf{X}_i)] + \text{constant} \\ &= \arg \min_{Q_i} -E_{Q_i} [E_{Q_{-i}} [\log p(\mathbf{X}_i, \mathbf{Y}_i | \mathbf{X}_{au(i)})]] \\ &\quad + E_{Q_i} [\log Q_i(\mathbf{X}_i)] \\ &= \arg \min_{Q_i} D(Q(\mathbf{X}) \| p(\mathbf{X}_i | \mathbf{X}_{au(i)}, \mathbf{Y})) \end{aligned} \quad (8)$$

Hence, we perform inference, i.e. optimize  $Q_i(\mathbf{X}_i)$ , over the individual chains in a coordinate descent procedure until  $\mathcal{F}\mathcal{E}$  converges. From the last line in (8), note that at optimality, the approximating distribution  $Q_i$  is related to the expected value of the aunt chains as follows:

$$Q_i \propto \exp \left\{ E_{Q_{au(i)}} [\log p(\mathbf{X}_i, \mathbf{Y}_i | \mathbf{X}_{au(i)})] \right\}. \quad (9)$$

Effectively, the approximating distribution  $Q_i$  incorporates information from neighboring chains via the  $p(\mathbf{X}_i, \mathbf{Y}_i | \mathbf{X}_{au(i)})$  terms in (9). Notice that the exponent of (9) factors into  $p(\mathbf{X}_i, \mathbf{Y}_i | \mathbf{X}_{au(i)}) = p(\mathbf{X}_i | \mathbf{X}_{au(i)}) p(\mathbf{Y}_i | \mathbf{X}_i)$ . These two factors can be matched to the pairwise and singleton terms, respectively, of the approximating distribution  $Q_i(\mathbf{X}_i)$  in (6). In addition, we approximate the contribution of future  $X_j^{t+1}$  for chains  $j \in au(i)$  via a linearized approximation that we incorporate into the singleton terms of (6). This approximation is further described in Section III-A2.

The expectations in (7) are computed by iteratively applying the forward-backward algorithm to a single chain, while

fixing the approximate posterior probabilities of the remaining chains. Let the singleton and pairwise marginals in the approximating distribution for chain  $i$  at time  $t$  be defined as  $\tilde{\gamma}_i^t(j) := E_{Q_i} [\mathbf{1}(X_i^t = j)]$  and  $\tilde{\xi}_i^t(j, k) := E_{Q_i} [\mathbf{1}(X_i^t = j, X_i^{t+1} = k)]$  where  $\mathbf{1}(\cdot)$  is the indicator function.

1) *Pairwise factors:* The pairwise factors of the approximating distribution  $\psi_i^t(X_i^t, X_i^{t-1})$  mimic the transition parameters of the original distribution, i.e.  $\mathcal{F}\mathcal{E}$  is minimized via a left-to-right time inhomogenous structure

$$\tilde{A}_i^t = \begin{bmatrix} 1 - \tilde{g}_i^t & \tilde{g}_i^t & 0 \\ 0 & 1 - \tilde{h}_i^t & \tilde{h}_i^t \\ 0 & 0 & 1 \end{bmatrix}$$

where  $\psi_i^t(X_i^t = k, X_i^{t-1} = j) = (\tilde{A}_i^t)_{jk}$ , where  $(\cdot)_{jk}$  corresponds to the entry in row  $j$  and column  $k$  of the matrix argument. From (9) the pairwise term becomes

$$\psi_i^t(X_i^t, X_i^{t-1}) \propto \exp \left\{ E_{Q_{au(i)}} \left[ p(X_i^t | X_i^{t-1}, \mathbf{X}_{au(i)}^{t-1}) \right] \right\} \quad (10)$$

Substituting in the parameters of our distribution, this relationship implies that the onset probability satisfies  $\tilde{g}_i^t \propto \exp\{E_{Q_{au(i)}} \log g_i^t\}$  with  $(1 - \tilde{g}_i^t) \propto \exp\{E_{Q_{au(i)}} \log(1 - g_i^t)\}$ . A similar relationship is true for the variational offset parameter  $\tilde{h}_i^t$ . Dividing these terms and taking the logarithm, the variational transition terms are given by the expected value of the logits in the original transition prior.

$$\begin{aligned} \log \left( \frac{\tilde{g}_i^t}{1 - \tilde{g}_i^t} \right) &= \rho_0 + \rho_1 E_{Q_{au(i)}} [\eta_i^t] \\ \log \left( \frac{\tilde{h}_i^t}{1 - \tilde{h}_i^t} \right) &= \phi_0 + \phi_1 E_{Q_{au(i)}} [\eta_i^t] \\ E_{Q_{au(i)}} [\eta_i^t] &= \sum_{j \in au(i)} \tilde{\gamma}_j^{t-1} \end{aligned}$$

These equations bare a strong resemblance to the original transition terms presented in (2) and (3) and incorporate cross-channel information via the  $E_{Q_{au(i)}} [\eta_i^t]$  terms.

2) *Singleton Factors:* The singleton factors in our approximating distribution  $\omega_i^t(X_i^t)$  mimic the emission likelihood. However, these terms also absorb information from the aunt chains in the subsequent timestep. This information is captured by the multiplicative factor  $\exp\{\alpha_i^t(z)\}$ . We use a linearized approximation of this term, shown in (13), to easily fold it into the singleton factors. Namely the expectation in these equations is easily computed as the sum of the aunt's marginals in the previous timestep  $\nu_i^{t+1} = \sum_{j \in au(i)} \tilde{\gamma}_j^t$ :

$$\omega_i^t(X_i^t = 0, 2) = P(Y_i^t | X_i^t = 0, 2) \exp \{ \alpha_i^t(0) \} \quad (11)$$

$$\omega_i^t(X_i^t = 1) = P(Y_i^t | X_i^t = 1) \exp \{ \alpha_i^t(1) \} \quad (12)$$

$$\begin{aligned} \alpha_i^t(z) &\approx \sum_{j \in au(i)} \left[ \tilde{\xi}_j^t(0, 0) (-\rho_0 - \rho_1 (\nu_j^{t+1} + z)) \right. \\ &\quad - \tilde{\gamma}_j^t(0) \log \left( 1 + e^{-\rho_0 - \rho_1 (\nu_j^{t+1} + z)} \right) \\ &\quad + \tilde{\xi}_j^t(1, 1) (-\phi_0 - \phi_1 (\nu_j^{t+1} + z)) \\ &\quad \left. - \tilde{\gamma}_j^t(1) \log \left( 1 + e^{-\phi_0 - \phi_1 (\nu_j^{t+1} + z)} \right) \right]. \end{aligned} \quad (13)$$

For simplicity, the  $\alpha_i^t(\cdot)$  terms were omitted in the preliminary conference version of this work. However, their inclusion in this paper improves both the approximation quality and the detection performance on real-world EEG data.

### B. M-Step: Update Model Parameters

In the M-step, the parameters of the generating distributions are updated using the marginals computed in the E-step.

1) *Emission Parameters*: Updating each GMM emission likelihood requires a nested EM update. The inner E-step in (14) computes the expected latent state and mixture combination for every emission.

$$\tilde{\tau}_{i,j}^t(k) := E_{Q_i} [X_i^t = k, Y_i^t \text{ drawn from mixture } j] \quad (14)$$

The inner M-step in (15)–(20) updates the parameters of the GMM via standard mean and variance updates [28].

$$\mu_{ij}^{0,2} = \frac{\sum_{k \in \{0,2\}} \sum_{t=0}^T \tilde{\tau}_{i,j}^t(k) Y_i^t}{\sum_{k \in \{0,2\}} \sum_{t=0}^T \tilde{\tau}_{i,j}^t(k)} \quad (15)$$

$$\mu_{ij}^1 = \frac{\sum_{t=0}^T \tilde{\tau}_{i,j}^t(k) Y_i^t}{\sum_{t=0}^T \tilde{\tau}_{i,j}^t(k)} \quad (16)$$

$$\Sigma_{ij}^{0,2} = \frac{\sum_{k \in \{0,2\}} \sum_{t=0}^T \tilde{\tau}_{i,j}^t(k) (Y_i^t - \mu_{ij}^k)^2}{\sum_{k \in \{0,2\}} \sum_{t=0}^T \tilde{\tau}_{i,j}^t(k)} \quad (17)$$

$$\Sigma_{ij}^1 = \frac{\sum_{t=0}^T \tilde{\tau}_{i,j}^t(k) (Y_i^t - \mu_{ij}^k)^2}{\sum_{t=0}^T \tilde{\tau}_{i,j}^t(k)} \quad (18)$$

$$\pi_{ij}^0 = \pi_{ij}^2 = \frac{\sum_{t=0}^T \tilde{\tau}_{i,j}^t(0) + \tilde{\tau}_{i,j}^t(2)}{\sum_{j'} \sum_{t=0}^T \tilde{\tau}_{i,j'}^t(0) + \tilde{\tau}_{i,j'}^t(2)} \quad (19)$$

$$\pi_{ij}^1 = \frac{\sum_{t=0}^T \tilde{\tau}_{i,j}^t(1)}{\sum_{j'} \sum_{t=0}^T \tilde{\tau}_{i,j'}^t(1)} \quad (20)$$

The nested E-step and nested M-step are repeated until the algorithm converges. This iterative procedure is initialized using the previous settings for the emission likelihoods.

2) *Transition Parameters*: The transition parameters form a logistic regression onto the expected transition posteriors  $\tilde{\xi}_i^t(j, k)$ . Here we provide the update equations for the onset parameters  $\rho_0$  and  $\rho_1$ . Equations for offset parameters  $\phi_0$  and  $\phi_1$  are almost identical and are omitted for space. Newton's method is used to minimize the  $\mathcal{F}\mathcal{E}$ . Let  $\nabla_{\rho} \mathcal{F}\mathcal{E}_k$  and  $\nabla_{\rho}^2 \mathcal{F}\mathcal{E}_k$  be the gradient and Hessian of the free energy with respect to the vector of onset parameters  $\rho = (\rho_0, \rho_1)$ . A single iteration of the Newton's method algorithm is

$$p_k = -(\nabla_{\rho}^2 \mathcal{F}\mathcal{E}_k)^{-1} \nabla_{\rho} \mathcal{F}\mathcal{E}_k \quad \rho_{k+1} = \rho_k + \alpha_k p_k$$

where the subscript  $k$  indicates the iteration number and  $\alpha_k$  is the step size. Newton's method is prone to oscillation in logistic regression in some cases. Therefore we employ backtracking to ensure our updates remain within a stable region around the minimum. Specifically, we require our step size to fulfill the second strong Wolfe condition  $|\nabla f(\rho_k + \alpha_k p_k)^T p_k| \leq |\nabla f(\rho_k)^T p_k|$  [38]. This ensures that each step approaches a stationary point. Defining the logistic sigmoid

function as  $\sigma(x) := \frac{1}{1+e^{-x}}$ , the first and second derivatives making up the gradient and Hessian are shown below.

$$\begin{aligned} \frac{\partial}{\partial \rho_0} E_Q [\log P(\mathbf{X})] &= \sum_{t=1}^T \sum_{i=1}^N \left( \tilde{\xi}_i^t(0, 1) \right. \\ &\quad \left. - \tilde{\gamma}_i^t(0) E_{Q_{au(i)}} \left[ \sigma(\rho_0 + \rho_1 \eta_i^t) \right] \right) - 2\lambda \rho_0 \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \rho_1} E_Q [\log P(\mathbf{X})] &= \sum_{t=1}^T \sum_{i=1}^N \tilde{\xi}_i^t(0, 1) \left( E_{Q_{au(i)}} [\eta_i^t] \right. \\ &\quad \left. - \tilde{\gamma}_i^t(0) E_{Q_{au(i)}} \left[ \eta_i^t \sigma(\rho_0 + \rho_1 \eta_i^t) \right] \right) - 2\lambda \rho_1 \end{aligned}$$

$$\begin{aligned} \frac{\partial^2}{\partial \rho_0^2} E_Q [\log P(\mathbf{X})] &= \sum_{t=1}^T \sum_{i=1}^N -\tilde{\gamma}_i^t(0) \\ &\quad \cdot E_{Q_{au(i)}} \left[ \sigma(\rho_0 + \rho_1 \eta_i^t) (1 - \sigma(\rho_0 + \rho_1 \eta_i^t)) \right] - 2\lambda \end{aligned}$$

$$\begin{aligned} \frac{\partial^2}{\partial \rho_1^2} E_Q [\log P(\mathbf{X})] &= \sum_{t=1}^T \sum_{i=1}^N -\tilde{\gamma}_i^t(0) \\ &\quad \cdot E_{Q_{au(i)}} \left[ (\eta_i^t)^2 \sigma(\rho_0 + \rho_1 \eta_i^t) \cdot (1 - \sigma(\rho_0 + \rho_1 \eta_i^t)) \right] - 2\lambda \end{aligned}$$

$$\begin{aligned} \frac{\partial^2}{\partial \rho_0 \partial \rho_1} E_Q [\log P(\mathbf{X})] &= \sum_{t=1}^T \sum_{i=1}^N -\tilde{\gamma}_i^t(0) \\ &\quad \cdot E_{Q_{au(i)}} \left[ \eta_i^t \sigma(\rho_0 + \rho_1 \eta_i^t) (1 - \sigma(\rho_0 + \rho_1 \eta_i^t)) \right] \end{aligned}$$

Here  $\ell_2$  norm regularization with weight  $\lambda$  is used to stabilize the learning and deal with identifiability issues.

### C. CHMM Initialization and Semi-Supervised Training

Our model is trained on multichannel EEG snippets in which an expert has annotated the approximate start and end of a single seizure. We emphasize that we do not use localization information about where the seizure originates and how it spreads. Rather, our model automatically learns this information from the data. Pre-seizure is fixed as state  $X_i^t = 0$  and post-seizure is fixed as  $X_i^t = 2$  throughout the course of training. Inference is performed over the seizure interval with a required transition into the seizure state. This semi-supervised strategy overcomes both the lack of exact onset and offset labels and the lack of spreading labels.

We initialize the emission distribution based on the seizure interval annotations. A GMM for the seizure state is trained on all data from the seizure interval while non-seizure state GMM is trained on data from the rest of the recording. Transition prior parameters are initialized to  $\rho_0 = -7$ ,  $\rho_1 = 2$ ,  $\phi_0 = -3$ , and  $\phi_1 = 0$ . These settings correspond to one expected seizure every 13 minutes lasting an expected length of 15 seconds. An aunt in a seizure state raises the probability of seizure onset in a given channel by a multiple of roughly 7, with no change in offset probability due to aunts in a seizure state. *We emphasize that these settings are just for initialization. The model updates these parameters through the variational EM algorithm.* In fact, we observe convergence to a stable set of parameter values regardless of initialization.

### D. Comparison to Machine Learning Baselines

We compared the performance of our model to classifiers from the machine learning literature, performing classification using a Deep Neural Network (DNN) and a Random Forests classifier (RF) [28]. The neural network includes 2 hidden layers with a discriminative output layer as in [39]. In contrast, RFs are an ensemble of simple decision trees that perform classification using majority vote over the decisions of the ensemble. By combining relatively simple classifiers, random forests can create complicated decision functions while remaining robust to overfitting.

In addition, a GMM based Likelihood Ratio Test (LRT) was used to perform posterior inference. The GMM-LRT is analogous to our original model with no prior over the hidden states. One GMM is trained for all non-seizure intervals to model  $P(Y | X = 0)$  and another is trained on the seizure intervals to model  $P(Y | X = 1)$ .  $\delta := P(X = 1)$  is fixed to the proportion of seizure in the dataset. The posterior probability of a test frame belonging to the seizure class is

$$P(X = 1 | Y) = \frac{\delta P(Y | X = 1)}{\delta P(Y | X = 1) + (1 - \delta) P(Y | X = 0)}.$$

For each classifier, DNN, RF, or GMM, two approaches to seizure detection are evaluated. In the first approach, features from each EEG channel are concatenated to form a single stacked feature vector used for detection. The second approach trains classifiers on each channel independently to evaluate the performance from a channel-wise perspective. These comparisons allow us to quantify the gain from fusing information across channels. When presenting our baseline results, the prefixes S and I are used to represent stacked feature vectors and independent channel-wise classification, respectively, e.g. SGMM for stacked feature Gaussian mixture model or IRF for independent channel random forest.

## IV. EVALUATION ON SYNTHETIC DATA

Synthetic data is generated by simulating different seizure propagation patterns. The CHMM and baseline algorithms are then used to infer the underlying spatio-temporal dynamics. The latent seizure states are sampled from a modified version of the transition prior outlined in Section II-A governed by onset parameters  $\{\rho_0, \rho_1\}$ , and offset parameters  $\{\phi_0, \phi_1\}$ . Each seizure recording begins in a non-seizure state. Prior to a seizure occurring, the probability of a channel entering the seizure state depends on its neighbors via  $\sigma(\rho_0)$ . After onset, the probability a channel enters the seizure state is given by  $\sigma(\rho_0 + \rho_1 \eta_i^t)$ . Thus,  $\rho_1$  controls the speed of seizure spreading, where higher values cause faster spreading. Departing from our prior, we enforce that all channels must enter the seizure state before offset is allowed. Once all channels enter the seizure state, the probability a channel returns to the normal state is given by  $\sigma(\phi_0 + \phi_1 \eta_i^t)$ .

Parameters are fixed  $\rho_0 = -9.0$ ,  $\phi_0 = -3.0$ , and  $\phi_1 = 0.0$  to simulate seizures occurring after roughly 425 timesteps. These parameters control the likelihood of seizure onset, the base rate of offset, and the between channel influence during offset. As these parameters are less clinically relevant

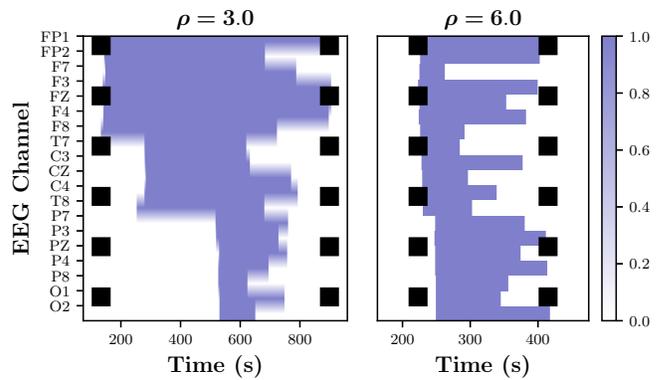


Fig. 5. Simulated underlying seizure stats for (left) slow and (right) fast propagation. Seizure onset and offset are shown with dashed black vertical lines. Seizure is shown in blue while non-seizure is shown in white.

than the speed of seizure spreading, we do not vary them between tests. Instead, we vary the neighborhood influence  $\rho_1$ , which indirectly controls the spreading rate. In our experiments,  $\rho_1 \in \{3.0, 4.0, 5.0, 6.0\}$  to explore a range of seizure spreading speeds, where higher values of  $\rho_1$  correspond to quicker spreading. While recording length in the real-world datasets varies, this variation is not clinically meaningful. Thus the length of the simulations is fixed to 1600 samples, corresponding to 20 minute recordings. Fig. 5 shows two simulated seizures for  $\rho_1 = 3.0$  (slow) and  $\rho_1 = 6.0$  (fast).

Emissions are sampled from a univariate normal distributions with mean 0 for non-seizure and mean 1 for seizure. The intra-class variance parameter is swept in  $[0.1, 1]$  to evaluate the model performance under different degrees of separability between the seizure and non-seizure classes. If the variance is small, the two classes remain easily separable. However, as the variance increases, the data distributions have a higher degree of overlap, making classification more difficult. For each setting, 10 sets of simulated training and testing data  $X$  are generated, each containing 100 seizures. Classifiers are trained using the training sets and test performance is reported for each classifier using the average across all folds.

Figs. 6 and 7 show the results of the simulated experiment. For both the IDNN and SDNN we use a network with two hidden layers of 10 nodes. The Area Under the Curve (AUC) for each test is shown on the y-axis while the intra-class variance parameter is shown on the x-axis. Fig. 6 shows the CHMM models and independent baselines for the full range of intra-class variances. Because the channels are evaluated individually, baselines in this figure achieve extremely similar performance for all values of  $\rho_1$ . At higher noise levels we observe that performance of the CHMM model degrades faster under slower spreading seizures. Intuitively, this makes sense as seizures originating close to simultaneously in each channel should result in easier cross channel information fusion, leading to increased performance. Fig. 7 shows the performance for the CHMM and stacked baselines. In general we observe that the performance of the stacked baselines increases as  $\rho_1$  increases. This makes intuitive sense, as faster spreading seizure activity will be present in more channels

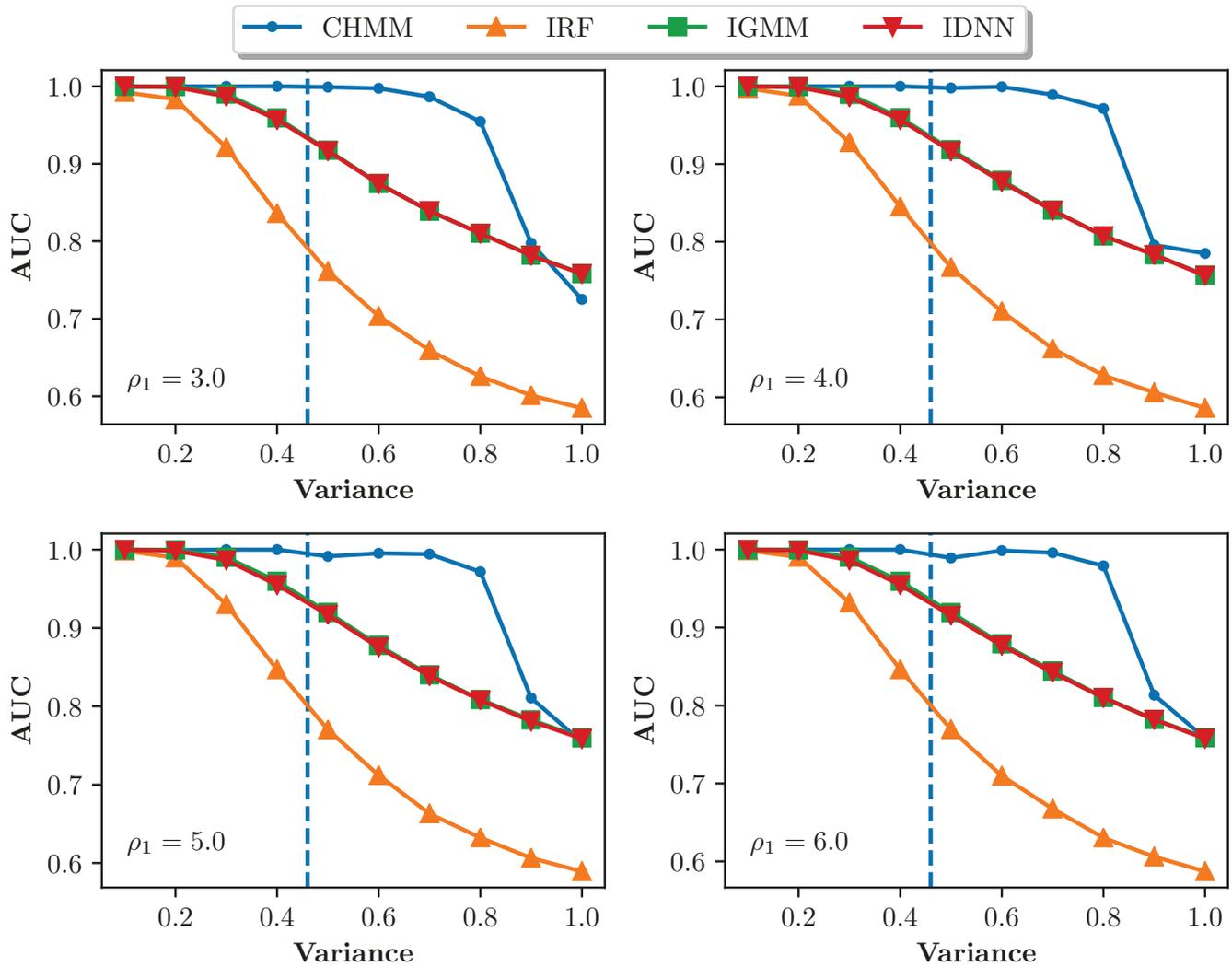


Fig. 6. AUC results for the CHMM and channel-independent baseline methods across a range intra-class variance values. Class separability estimated from the real-world EEG data is shown by the vertical dashed blue line. Values of  $\rho$  correspond to rate of seizure spread.

concurrently and is thus easier to classify.

Finally, we have estimated a lower bound on the class separability of our real-world EEG data by computing the Hellinger distance between the seizure and non-seizure classes in the focal JHH dataset. Specifically, a multivariate normal distribution with full covariance is fit to the temporal features extracted for each class (see Section V for details on the data preprocessing). The Hellinger distance between these two distributions for each EEG channel is then computed. These distances were then averaged across all EEG channels and recordings. Based on the average Hellinger distance, we computed the real-world data to have an approximate emission variance of 0.46. This value is marked with a dashed vertical line in Figs. 6–7.

## V. EVALUATION ON CLINICAL DATA

We evaluate our model on clinical EEG data recorded in two different hospitals. Details of the datasets, preprocessing, and feature extraction are given below. In our experiments on real

data, the two hidden layers of the IDNN and SDNN contain 10 and 50 units, respectively.

**JHH Dataset:** This data is acquired in the Epilepsy Monitoring Unit (EMU) of the Johns Hopkins Hospital. The recordings were screened and annotated by epilepsy fellows in the Department of Neurology. Due to the liberal annotation procedures followed at this hospital, many of our annotated seizure intervals are overly generous and contain periods of baseline before and after the seizure event. In total our dataset includes 90 seizures from 15 patients. Each of these seizure recordings contains up to 10 minutes of baseline EEG before and after the seizure. Recordings were sampled at 200 Hz in 10/20 reference space using the common average montage. In early experimentation, we evaluated our methods after applying the bipolar montage but found no sizable change in performance. Though not used during training, a subset of recordings contain clinical annotations of the likely seizure onset. When possible, we validate the seizure spread information generated by our model using these annotations.

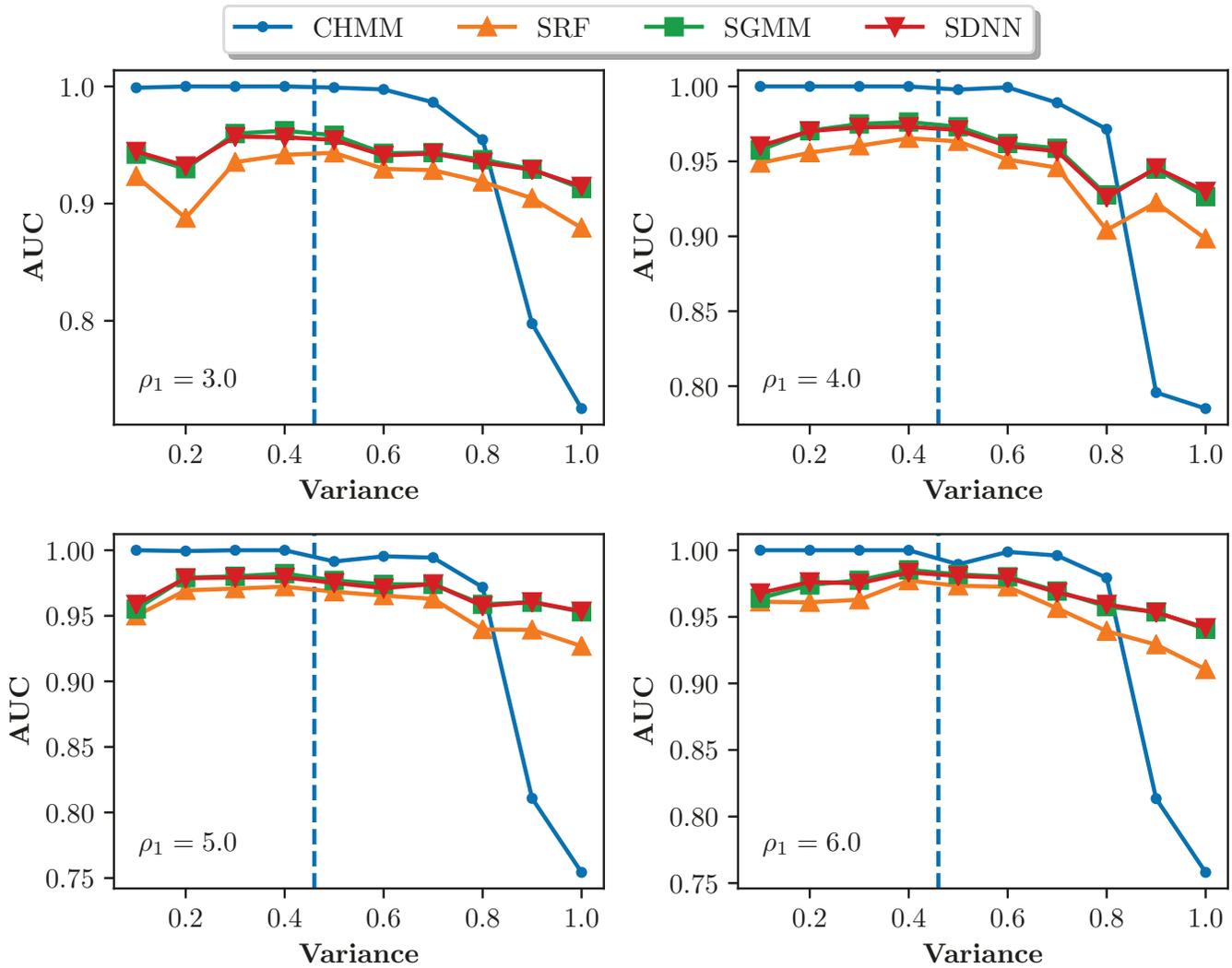


Fig. 7. AUC results for the CHMM and stacked-channel baseline methods across a range intra-class variance values. Class separability estimated from the real-world EEG data is shown by the vertical dashed blue line. Values of  $\rho$  correspond to rate of seizure spread.

**CHB Dataset:** We also tested our algorithm on a publicly available dataset recorded at Children’s Hospital in Boston (CHB) [10],[40]. The dataset contains scalp EEG recordings from pediatric patients and one adult ranging from age 1.5 to age 22. Seizure regions from the CHB dataset were trimmed with a random amount of pre- and post-seizure baseline not exceeding 10 minutes in each case. In total, 185 seizures from 24 patients were used. This dataset contains both focal and generalized seizures with more accurately annotated seizure intervals. The data has been released in the bipolar montage; hence the network depicted in Fig. 3b for was used inference.

#### A. Preprocessing and Feature Extraction

Each EEG recording is minimally preprocessed using a high-pass filter at 1.6 Hz and a low-pass filter at 50 Hz to remove DC trends and high frequency noise. In addition, a second order notch filter at 60 Hz with  $Q = 20$  was used to remove any remaining interference from the power supply.

We extracted features based on 1 second Tukey windows with shape parameter 0.25 and a 750 ms advance. First, short time Fourier transform coefficients for each window are computed. The coefficient magnitudes were summed according to the standard EEG frequency bands: theta (1–4 Hz), delta (4–8 Hz), alpha (8–13 Hz), and beta (13–30 Hz). A logarithm was applied to the summed features. These features track activity in each brain wave band, which has been noted to change during seizures and closely resembles feature extraction techniques in [16], [26]. Log line length features [41], computed as  $\log L = \log \left( \sum_{t=1}^T |s[t] - s[t-1]| \right)$  where  $s$  is a time series, were also included. Line length captures the smoothness of a signal. The features for each channel were normalized to mean 0, standard deviation 1 for each recording. This combination of features echoes those cited as optimal in [20]. Our prior experimentation with different feature extraction methods verified that spectral power and line length outperformed more sophisticated EEG features in the literature such as wavelet and entropy measures. In addition,

similar features have been employed for use with implanted EEG sensors [42].

### B. Evaluation

Five-fold cross validation was used to evaluate the methods. Each recording was randomly assigned to a fold irrespective of patient. This approach stands in contrast to prior work, which trains patient specific classifiers. Rather, we evaluate our model in a general setting where we may not have prior data from any given patient. Training was performed on four folds while testing was performed on the remaining fold. Reported metrics are averaged across all test folds.

We quantitatively evaluate each model based on the amount of detected seizure activity within the annotated region. Here, each channel detection within the seizure interval is counted as True Positive (TP). Each channel detection outside the interval counted as a False Positive (FP). Let  $t = 0, \dots, T$  index the one-second time windows within a single recording and let  $t_s$  and  $t_e$  denote the starting and ending time of the annotated seizure interval. Mathematically let the TP, FP, True Negatives (TN), and False Negatives (FN) for channel  $i$  be defined as:

$$TP_i = \sum_{t=t_e}^{t_s} \mathbf{1}(X_i^t = 1) \quad FP_i = \sum_{t=1}^T \mathbf{1}(X_i^t = 1) - TP_i$$

$$FN_i = \sum_{t=t_e}^{t_s} \mathbf{1}(X_i^t \neq 1) \quad TN_i = \sum_{t=1}^T \mathbf{1}(X_i^t \neq 1) - FN_i .$$

Detections are aggregated across channels to yield True Positive Rate (TPR) and True Negative Rate (TNR):

$$TPR = \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{t_e - t_s}$$

$$TNR = \frac{1}{N} \sum_{i=1}^N \frac{TN_i}{T - (t_e - t_s)} .$$

As we lack onset annotations for each individual channel we calculate these statistics based on the single clinician provided onset annotation. This strategy is based on the assumption that given liberal onset annotations, any positive classification within the annotation are likely to be correct.

Our reported performance is averaged over (potentially generous) seizure regions. Notice that our evaluation criterion is more stringent than the metrics reported in prior work as we report *percentages of correctly classified activity* rather than a single correct detections within the seizure interval. Hence, lower overall TPR than is presented in other seizure detection papers is expected. Precision (P) and Recall (R) averaged across channels are reported along with the AUC and F1 score to evaluate overall performance of each detector:

$$P = \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{TP_i + FP_i}$$

$$R = \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{TP_i + FN_i} .$$

### C. Experimental Results

Fig. 8 depicts the output of our model for a single recording from the JHH dataset. Fig. 8a shows the posterior beliefs of our model in blue. EEG channels are arranged along the y-axis of the image while time progresses horizontally. The dashed black lines indicate the annotated onset and offset of the seizure. Once again, these annotations serve as a rough guide, rather than a precise demarcation of onset. Fig. 8b-d shows baseline classifications for the same recording shown in Fig. 8a. Fig. 9 shows the same posterior distributions superimposed on the raw EEG signal while Fig. 10 shows them topographically on the scalp as the seizure progresses.

These results illustrate the ability of our CHMM to correctly label seizure intervals. Clinical annotations for the seizure in Fig. 8 note rhythmic theta activity in the left frontal area at the onset of the seizure, which correlates with the earliest CHMM detection in Fig. 8a. Likewise, the superimposed posteriors in Fig. 9 and topographic detail in Fig. 10 both show the earliest response of our model occurring in the left frontal region and spreading through the rest of the channels.

Figs. 8b and 8c illustrate the behavior when using the concatenated spectral power and line length features to make a single framewise classification. Notice that both models place higher weight on the seizure interval but lack contiguity. In addition, the GMM model in Fig. 8b reacts strongly to activity after the seizure that is likely muscle artifact. Conversely, Fig. 8d shows the results when channels were classified independently using a GMM classifier. This strategy allows us to isolate seizure activity in different channels but does not impose spatial or temporal contiguity in classification.

Mean and standard deviations of evaluation metrics are shown for the JHH dataset in Table II. As seen, our method outperforms all machine learning baselines. The clinical annotations demarcating seizure intervals have a tendency to extend beyond onset and offset. This coarse labeling and our stringent evaluation explains the low TPR across all models relative to results in the literature. The effect of the spatio-temporal transition prior is clear by comparing the CHMM with the IGMM. The prior allows our model to correctly place more posterior confidence in seizure, resulting in a higher TPR, while ignoring non-seizure baseline behavior that resembles seizure activity, simultaneously improving TNR. Of the baselines evaluated, only the SDNN surpassed the CHMM in TPR. The CHMM model surpasses the baseline methods in the summary scores AUC and F1, standing as much as a standard deviation above the best performing baselines.

Similar to the JHH dataset, Fig. 11 depicts posterior beliefs for a representative patient from the CHB dataset. Again, black dashed lines indicate annotated seizure onset and offset. Fig. 11a shows posteriors from the CHMM. Since the dataset does not include annotations of seizure type or focus, images of posteriors superimposed on the raw EEG signal and topographic details have been omitted. Once again, posteriors from both GMM baselines and the SRF are included.

Notice that in Fig. 11a the CHMM correctly labels the region containing the seizure but its detection extends slightly beyond the annotated offset. This is behavior can be attributed

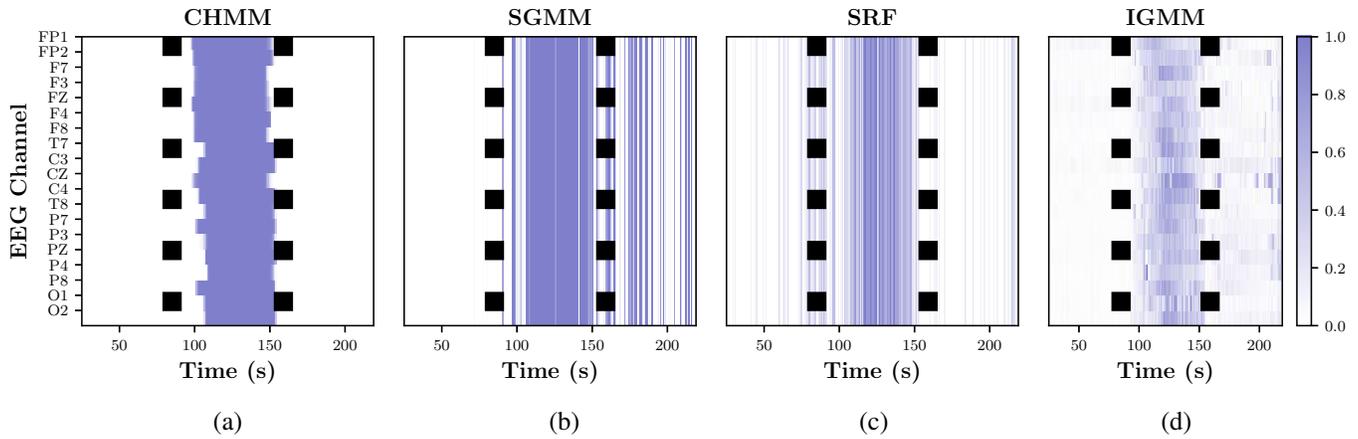


Fig. 8. CHMM and selected baseline classification posteriors for a representative JHH patient. EEG channels are arranged on the y-axis with time along the x-axis. Seizure onset and offset are indicated by the vertical dashed lines. (a) Classification results using our CHMM model. Stacked features are used in conjunction with GMM and RF classifiers in (b) and (c), respectively. (d) Classification performed on each channel with a GMM. Posterior beliefs are shown in blue where intensity depicts the strength of the belief.

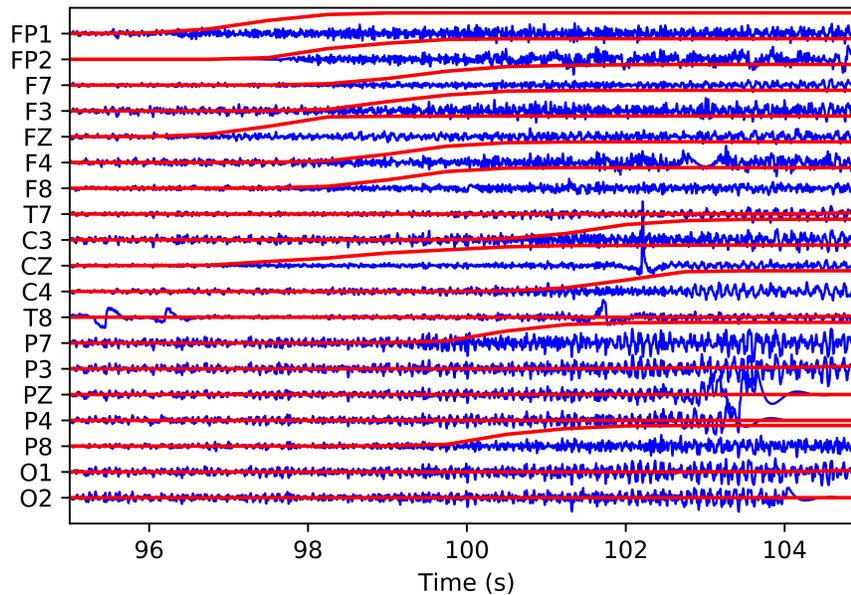


Fig. 9. CHMM posteriors superimposed on EEG for the recording in Fig. 8. Raw EEG signal is shown in blue while CHMM posteriors are shown in red. EEG channels are organized on the y-axis, while time progresses along the x-axis.

to the high degree of artifact present in the EEG signal post-seizure. While we lack clinical annotations regarding seizure types for this dataset, the presence of rhythmic activity occurring simultaneously in all channels at the annotated seizure onset indicates that this is likely a generalized seizure. Our CHMM inference readily captures this phenomenon by turning all channels on simultaneously. Hence even though our CHMM prior assumes a focal spreading pattern, our method is flexible enough to capture multiple seizure types. Once again, the baseline classifiers suffer from the same drawbacks. The models trained on stacked features place higher beliefs in seizure regions but allow for many spurious onsets and offsets while models trained on each channel

individually place lower posterior beliefs in seizure regions.

Quantitative results for the CHB dataset are shown in Table III. The best performance under each metric is bolded. We have underlined when our model achieved the second highest performance in any metric. Our model outperforms baselines in most metrics while remaining within a standard deviation of the best performing baselines. In general, the GMM and RF classifiers using stacked features are slightly biased towards positive classifications, resulting in higher TPR and recall, but lower TNR and precision.

In summary, our Bayesian model outperforms all baseline methods in the JHH dataset. The focal epileptic seizures present in this population are most accurately classified using

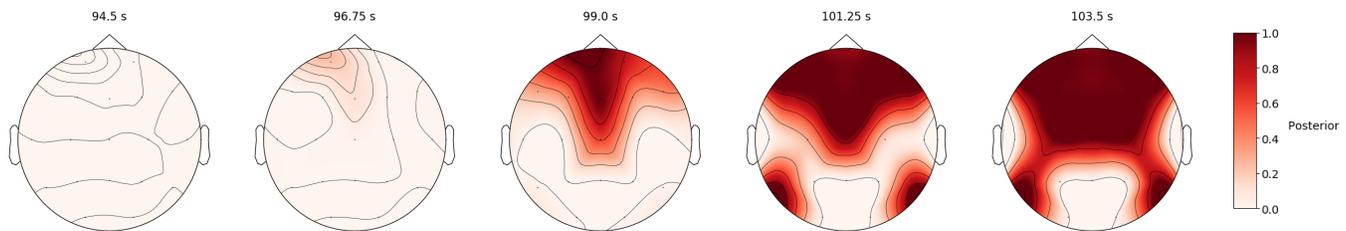


Fig. 10. Spread of the seizure depicted in Fig. 8, as computed by the CHMM. The CHMM classifies the earliest ictal activity occurring in the left frontal channels in agreement with clinical annotations.

TABLE II  
QUANTITATIVE RESULTS FOR THE JHH DATASET

Classifier	TPR	TNR	AUC	P	R	F1
CHMM	$37.40 \pm 7\%$	$98.29 \pm 0.54\%$	$0.84 \pm 0.05$	$0.66 \pm 0.11$	$0.44 \pm 0.09$	$0.49 \pm 0.09$
SDNN	$38.10 \pm 4\%$	$93.76 \pm 1.08\%$	$0.82 \pm 0.03$	$0.35 \pm 0.04$	$0.41 \pm 0.05$	$0.35 \pm 0.04$
SGMM	$33.42 \pm 5\%$	$95.19 \pm 1.45\%$	$0.72 \pm 0.04$	$0.44 \pm 0.08$	$0.37 \pm 0.06$	$0.36 \pm 0.00$
SRF	$29.35 \pm 3\%$	$92.53 \pm 1.20\%$	$0.79 \pm 0.03$	$0.29 \pm 0.05$	$0.34 \pm 0.04$	$0.28 \pm 0.06$
IDNN	$22.31 \pm 2\%$	$92.93 \pm 1.02\%$	$0.80 \pm 0.03$	$0.24 \pm 0.03$	$0.25 \pm 0.02$	$0.22 \pm 0.02$
IGMM	$26.20 \pm 3\%$	$92.79 \pm 1.36\%$	$0.79 \pm 0.03$	$0.27 \pm 0.04$	$0.30 \pm 0.03$	$0.25 \pm 0.04$
IRF	$24.11 \pm 3\%$	$92.51 \pm 1.04\%$	$0.74 \pm 0.03$	$0.24 \pm 0.03$	$0.28 \pm 0.03$	$0.23 \pm 0.03$

the CHMM with a transition prior designed for this task. Baseline methods fared poorly in part due to the heterogeneity of focal seizure presentations. In general, the CHB dataset contains better annotations, resulting in higher performance across the board. Furthermore, stacked feature vector based classification in the CHB dataset is better due to the presence of generalized seizures. Despite being tailored to capture focal spreading patterns, our CHMM maintains robust performance across both datasets. This cross hospital evaluation is the first of its kind and demonstrates our model’s ability to generalize to diverse seizure types and patient populations.

## VI. DISCUSSION

We have developed a novel CHMM framework that captures the spatio-temporal propagation of a seizure for robust seizure detection. Using a variational approximation, we are able to efficiently perform inference and learn the model parameters despite its high dimensional state space. The CHMM is compared to baseline classifiers based on both individual and concatenated EEG features trained across patients. The framework is evaluated on EEG data acquired at two different hospitals, which has not previously been reported in the seizure detection literature. Our CHMM model outperformed or performed comparably to the best machine learning baselines in both our JHH dataset of focal epilepsy and the publicly available CHB dataset of pediatric epilepsy recordings.

Performance of our model in the JHH dataset exceeded that of all the baselines in all but one statistic. This improvement demonstrates our models efficacy in patient-agnostic seizure detection in a heterogeneous focal epilepsy dataset. In contrast, for the CHB dataset, our CHMM performed within a standard deviation of the best baseline approaches. We believe these differences arise from two clinically-relevant factors. First, our modeling choices regarding the spread of focal seizures mirror that of the patient cohort, as every patient in the

JHH dataset has focal epilepsy. Thus our model more closely models the data than any of the baselines, leading to increased performance. In the CHB dataset, seizures appear to spread faster, indicating the presence of patients with generalized seizures. The simulations demonstrate that stacked feature baselines perform better in these conditions, thus explaining their better performance in this more homogeneous patient cohort. Second, the CHB dataset has been well-curated prior to its release, allowing better training of baseline models. In contrast, the JHH dataset has undergone minimal preprocessing to better reflect clinical conditions.

Furthermore our model provided onset localization information for several patients in the JHH dataset, which highlights its potential use in localizing seizure foci. This coarse localization in the EEG sensor space mirrors the early stages of clinical diagnosis, where EEG provides localization to a channel or lobe. As EEG is cheap to acquire, this coarse localization provides diagnostic information to guide expensive or invasive modalities, such as PET, MRI, or ECoG. Findings from EEG can be used in conjunction with these modalities for more comprehensive localization and treatment planning.

Interestingly, the CHB dataset includes generalized seizures which do not fit the assumptions of our spreading prior. Correspondingly, our simulated experiments show that the stacked baselines achieve better performance with faster spreading seizures. This behavior is mirrored in the CHB dataset where the CHMM performs on par with the baselines, particularly for the stacked feature representations. The advantage of our model stems from the ability of the CHMM transition prior to isolate the highest probability seizure interval. Despite the lack simultaneous onset in generalized seizures, this temporal data fusion still increases the CHMM performance relative to the individual channel baselines.

Future work will explore several modeling improvements to our CHMM. For example, allowing synchronous onset

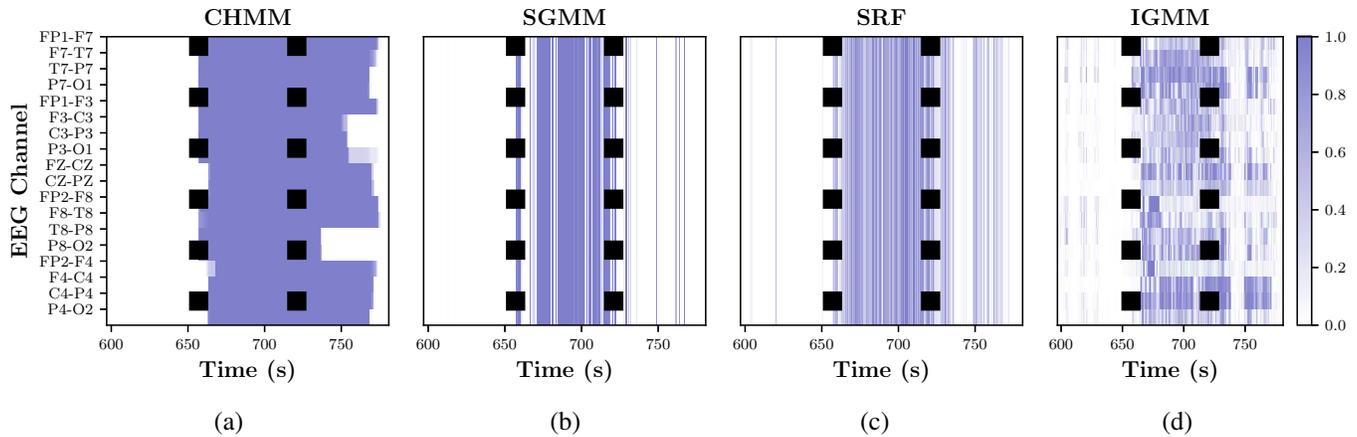


Fig. 11. CHMM and selected baseline classification posteriors for a representative CHB patient. EEG channels are arranged on the y-axis with time along the x-axis. Seizure onset and offset are indicated by the vertical dashed lines. (a) Classification results using our CHMM model. Stacked features are used in conjunction with GMM and RF classifiers in (b) and (c), respectively. (d) Classification performed on each channel with a GMM. Posterior beliefs are shown in blue where intensity depicts the strength of the belief.

TABLE III  
QUANTITATIVE RESULTS FOR THE CHB DATASET

Classifier	TPR	TNR	AUC	P	R	F1
CHMM	$57.43 \pm 5.64\%$	<b><math>98.67\% \pm 0.38</math></b>	$0.86 \pm 0.03$	<b><math>0.65 \pm 0.05</math></b>	$0.54 \pm 0.06$	<b><math>0.56 \pm 0.04</math></b>
SDNN	$52.31 \pm 8.44\%$	$97.79 \pm 0.41\%$	<b><math>0.91 \pm 0.01</math></b>	$0.51 \pm 0.05$	$0.48 \pm 0.09$	$0.47 \pm 0.07$
SGMM	<b><math>59.40 \pm 4.28\%</math></b>	$97.36 \pm 0.67\%$	$0.87 \pm 0.02$	$0.53 \pm 0.05$	<b><math>0.55 \pm 0.02</math></b>	$0.51 \pm 0.03$
SRF	$46.89 \pm 4.63\%$	$96.55 \pm 0.29\%$	$0.88 \pm 0.02$	$0.41 \pm 0.03$	$0.45 \pm 0.03$	$0.40 \pm 0.03$
IDNN	$27.73 \pm 2.09\%$	$95.62 \pm 0.22\%$	$0.83 \pm 0.02$	$0.27 \pm 0.03$	$0.27 \pm 0.02$	$0.24 \pm 0.02$
IGMM	$30.82 \pm 2.05\%$	$95.56 \pm 0.22\%$	$0.83 \pm 0.02$	$0.28 \pm 0.03$	$0.30 \pm 0.02$	$0.26 \pm 0.02$
IRF	$29.66 \pm 2.52\%$	$95.45 \pm 0.20\%$	$0.76 \pm 0.02$	$0.27 \pm 0.03$	$0.29 \pm 0.02$	$0.25 \pm 0.02$

in all channels could increase performance on generalized seizures. Likewise, evaluating concatenated features could increase model performance. This fusion could be performed by an auxiliary HMM chain that focuses on the concatenated feature vectors to initiate the spreading chains. In addition, we observe that seizures recordings typically progress from rhythmic behavior to intervals of muscle artifact. Including more seizure states to model this transition could aid in localization and improve performance in general. Finally, some elements of our model may be useful in seizure prediction tasks where the specific focus is generally known and the typical propagation pattern is of interest. Seizure prediction is typically performed using ECoG and some adaptation to this domain may be required. However, as scalp EEG becomes more available through wearable technology [43], techniques designed for scalp EEG may become more prevalent in seizure prediction.

Many features have been evaluated for the purpose of EEG-based seizure detection. In this work, we identified spectral bands and line length as robust and simple features. However, learning these representations from the data is an interesting direction for future work. For example, we may leverage the ability of deep learning to learn more discriminative representations for EEG analysis from data.

## VII. CONCLUSION

In this paper we developed a spatio-temporal propagation model for epileptic seizures based on a CHMM architecture. We demonstrated our model on a dataset of focal epilepsy recordings and on a publicly available dataset of pediatric EEG recordings. By specifically modeling the spread of focal seizures, our model outperforms baseline classifiers in the dataset containing focal epilepsy recordings. In the dataset comprised of pediatric seizure recordings, our model performs the best or comparably to our baselines. While commercial seizure detection algorithms exist, they have yet to supplant manual annotation. Accurate and reliable seizure detection remains a clinical necessity. Our experimentation here indicates that direct modeling of cross-channel interactions present in EEG signals can improve the performance of seizure detection algorithms. This modeling shows the ability to provide information capable of aiding the localization process for diagnosis and treatment planning of focal epilepsy.

## ACKNOWLEDGMENTS

This work was supported in part by the JHMI synergy Award (PIs: Johnson/Venkataraman) and the National Science Foundation CAREER award 1845430 (PI: Venkataraman).

## REFERENCES

- [1] J. W. Miller and H. P. Goodkin, *Epilepsy*. Wiley, 2014.

- [2] J. A. French, "Refractory epilepsy: clinical overview," *Epilepsia*, vol. 48, pp. 3–7, 2007.
- [3] H. O. Lüders, I. Najm, D. Nair, P. Widdess-Walsh, and W. Bingman, "The epileptogenic zone: general principles," *Epileptic disorders*, vol. 8, no. 2, pp. 1–9, 2006.
- [4] W. H. Theodore, "Presurgical focus localization in epilepsy: Pet and spect," in *Seminars in nuclear medicine*, vol. 47, no. 1. Elsevier, 2017, pp. 44–53.
- [5] J. S. Duncan, G. P. Winston, M. J. Koepp, and S. Ourselin, "Brain imaging in the assessment for epilepsy surgery," *The Lancet Neurology*, vol. 15, no. 4, pp. 420–433, 2016.
- [6] J. Gotman, "Automatic recognition of epileptic seizures in the eeg," *Electroencephalography and clinical Neurophysiology*, vol. 54, no. 5, pp. 530–540, 1982.
- [7] T. N. Alotaiby, S. A. Alshebeili, T. Alshawi, I. Ahmad, and F. E. A. El-Samie, "Eeg seizure detection and prediction algorithms: a survey," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, p. 183, 2014.
- [8] L. Orosco, "a survey of performance and techniques for automatic epilepsy detection," *Journal of Medical and Biological Engineering*, vol. 33, no. 6, pp. 526–537.
- [9] I. Osorio, H. P. Zaveri, M. G. Frei, and S. Arthurs, *Epilepsy: the intersection of neurosciences, biology, mathematics, engineering, and physics*. CRC press, 2016.
- [10] A. H. Shoeb and J. V. Guttag, "Application of machine learning to epileptic seizure detection," in *International Conference on Machine Learning*, 2010, pp. 975–982.
- [11] G. L. Krauss and R. S. Fisher, *The Johns Hopkins atlas of digital EEG: an interactive training guide*. Johns Hopkins University Press, 2006.
- [12] D. Wu, Z. Wang, L. Jiang, F. Dong, X. Wu, S. Wang, and Y. Ding, "Automatic epileptic seizures joint detection algorithm based on improved multi-domain feature of ceeg and spike feature of aeeg," *IEEE Access*, vol. 7, pp. 41 551–41 564, 2019.
- [13] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Physical Review E*, vol. 64, no. 6, p. 061907, 2001.
- [14] N. F. Güler *et al.*, "Recurrent neural networks employing lyapunov exponents for eeg signals classification," *Expert systems with applications*, vol. 29, no. 3, pp. 506–514, 2005.
- [15] U. R. Acharya *et al.*, "Automated diagnosis of epileptic eeg using entropies," *Biomedical Signal Processing and Control*, vol. 7, no. 4, pp. 401–408, 2012.
- [16] M. Bandarabadi, C. A. Teixeira, J. Rasekhi, and A. Dourado, "Epileptic seizure prediction using relative spectral power features," *Clinical Neurophysiology*, vol. 126, no. 2, pp. 237–248, 2015.
- [17] R. Hopfengärtner, B. S. Kasper, W. Graf, S. Gollwitzer, G. Kreiselmeier, H. Stefan, and H. Hamer, "Automatic seizure detection in long-term scalp eeg using an adaptive thresholding technique: a validation study for clinical routine," *Clinical Neurophysiology*, vol. 125, no. 7, pp. 1346–1352, 2014.
- [18] Y. Khan and J. Gotman, "Wavelet based automatic seizure detection in intracerebral electroencephalogram," *Clinical Neurophysiology*, vol. 114, no. 5, pp. 898–908, 2003.
- [19] M. Saab and J. Gotman, "A system to detect the onset of epileptic seizures in scalp eeg," *Clinical Neurophysiology*, vol. 116, no. 2, pp. 427–442, 2005.
- [20] L. Logesparan, A. J. Casson, and E. Rodriguez-Villegas, "Optimal features for online seizure detection," *Medical & Biological Engineering & Computing*, vol. 50, no. 7, pp. 659–669, Jul 2012. [Online]. Available: <https://doi.org/10.1007/s11517-012-0904-x>
- [21] L. S. Vidyaratne and K. M. Iftekharuddin, "Real-time epileptic seizure detection using eeg," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 2146–2156, 2017.
- [22] R. Meier, H. Dittrich, A. Schulze-Bonhage, and A. Aertsen, "Detecting epileptic seizures in long-term human eeg: a new approach to automatic online and real-time detection and classification of polymorphic seizure patterns," *Journal of Clinical Neurophysiology*, vol. 25, no. 3, pp. 119–131, 2008.
- [23] S. Ghosh-Dastidar, H. Adeli, and N. Dadmehr, "Mixed-band wavelet-chaos-neural network methodology for epilepsy and epileptic seizure detection," *IEEE transactions on biomedical engineering*, vol. 54, no. 9, pp. 1545–1551, 2007.
- [24] H. Adeli, S. Ghosh-Dastidar, and N. Dadmehr, "A wavelet-chaos methodology for analysis of eegs and eeg subbands to detect seizure and epilepsy," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 2, pp. 205–211, 2007.
- [25] H. Ocaik, "Automatic detection of epileptic seizures in eeg using discrete wavelet transform and approximate entropy," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2027–2036, 2009.
- [26] V. Sridevi, M. R. Reddy, K. Srinivasan, K. Radhakrishnan, C. Rathore, and D. S. Nayak, "Improved patient-independent system for detection of electrical onset of seizures," *Journal of Clinical Neurophysiology*, vol. 36, no. 1, p. 14, 2019.
- [27] E. Alickovic, J. Kevric, and A. Subasi, "Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction," *Biomedical signal processing and control*, vol. 39, pp. 94–102, 2018.
- [28] K. P. Murphy, "Machine learning: a probabilistic perspective," 2012.
- [29] F. Jelinek, *Statistical methods for speech recognition*. MIT press, 1997.
- [30] D. F. Wulsin, E. B. Fox, and B. Litt, "Modeling the complex dynamics and changing correlations of epileptic events," *Artificial intelligence*, vol. 216, pp. 55–75, 2014.
- [31] A. V. Nefian, L. Liang, X. Pi, L. Xiaoxiang, C. Mao, and K. Murphy, "A coupled hmm for audio-visual speech recognition," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, vol. 2. IEEE, 2002, pp. II–2013.
- [32] W. Dong, A. Pentland, and K. A. Heller, "Graph-coupled hmms for modeling the spread of infection," *arXiv preprint arXiv:1210.4864*, 2012.
- [33] I. Rezek and S. J. Roberts, "Estimation of coupled hidden markov models with application to biosignal interaction modelling," in *Neural Networks for Signal Processing X, 2000. Proceedings of the 2000 IEEE Signal Processing Society Workshop*, vol. 2. IEEE, 2000, pp. 804–813.
- [34] S. Zhong and J. Ghosh, "Hmms and coupled hmms for multi-channel eeg classification," in *Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on*, vol. 2. IEEE, 2002, pp. 1154–1159.
- [35] V. Jurcak *et al.*, "10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems," *Neuroimage*, vol. 34, no. 4, pp. 1600–1611, 2007.
- [36] J. Craley, E. Johnson, and A. Venkataraman, "A novel method for epileptic seizure detection using coupled hidden markov models," in *MICCAI*, 2018.
- [37] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, "An introduction to variational methods for graphical models," *Machine learning*, vol. 37, no. 2, pp. 183–233, 1999.
- [38] J. Nocedal and S. J. Wright, *Numerical Optimization*. Springer, 1999.
- [39] J. Craley, E. Johnson, and A. Venkataraman, "Integrating convolutional neural networks and probabilistic graphical modeling for epileptic seizure detection in multichannel eeg," in *Information Processing in Medical Imaging, A. C. S. Chung, J. C. Gee, P. A. Yushkevich, and S. Bao, Eds. Cham: Springer International Publishing, 2019, pp. 291–303*.
- [40] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [41] R. Esteller, J. Echaz, T. Tchong, B. Litt, and B. Pless, "Line length: an efficient feature for seizure onset detection," in *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, vol. 2. IEEE, 2001, pp. 1707–1710.
- [42] M. Shoaran, M. Farivar, and A. Emami, "Hardware-friendly seizure detection with a boosted ensemble of shallow decision trees," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2016, pp. 1826–1829.
- [43] W. Van Paesschen, "The future of seizure detection," *The Lancet Neurology*, vol. 17, no. 3, pp. 200–202, 2018.