A Novel Method for Epileptic Seizure Detection Using Coupled Hidden Markov Models

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

[1] Department of Electrical and Computer Engineering, Johns Hopkins University,
Baltimore, USA
jcraley@gmail.com
[2] Department of Neurology, Johns Hopkins Medical Institute, Baltimore, USA

Abstract. We propose a novel Coupled Hidden Markov Model to detect epileptic seizures in multichannel electroencephalography (EEG) data. Our model defines a network of seizure propagation paths to capture both the temporal and spatial evolution of epileptic activity. To address the intractability introduced by the coupled interactions, we derive a variational inference procedure to efficiently infer the seizure evolution from spectral patterns in the EEG data. We validate our model on EEG aquired under clinical conditions in the Epilepsy Monitoring Unit of the Johns Hopkins Hospital. Using 5-fold cross validation, we demonstrate that our model outperforms three baseline approaches which rely on a classical detection framework. Our model also demonstrates the potential to localize seizure onset zones in focal epilepsy.

1 Introduction

Epilepsy is a heterogenous neurological disorder characterized by recurring and unprovoked seizures [1]. It is estimated that 20–40% of epilepsy patients are medically refractory and do not respond to drug therapy. Alternative therapies for these patients crucially depend on being able to detect epileptic activity in the brain. The most common modality used for seizure detection is multichannel electroencephalography (EEG) acquired on the scalp. The clinical standard for seizure detection involves visual inspection of the EEG data, which is time consuming and requires extensive training. In this work, we develop an automated seizure detection procedure for clinically acquired multichannel EEG recordings.

There is a vast body of literature on epileptic seizure detection from a variety of viewpoints. The nonlinearity of EEG signals has inspired the application of techniques from chaos theory such as approximate entropy and Lyapunov exponents as in [2,3], respectively. Alternatively, wavelet and other time-frequency based features seek to capture the non-stationarity of the EEG signal as in [4]. These features are fed into standard classification algorithms to detect seizure activity. A fundamental limitation of the methods in [2,3] is that they are trained on a single channel of EEG data and fail to generalize in practice. Multichannel

strategies such as those in [4–7] rely heavily on prior seizure recordings to train patient specific detectors, which are often unavailable.

Unlike prior work, our approach explicitly models the spatial dynamics of a seizure through the brain over time. We build on existing work in Hidden Markov Models (HMMs) [7], adopting a Coupled HMM (CHMM) [8] to model interchannel dependencies. Specifically, the likelihood that an EEG channel will transition into a seizure state will increase if neighboring channels are in a seizure state. This coupling renders exact inference intractable. Therefore we develop a variational Expectation Maximization (EM) algorithm for our framework.

We evaluate our algorithm using 90 scalp EEG recordings from 15 epilepsy patients acquired in the Epilepsy Monitoring Unit (EMU) of the Johns Hopkins Hospital. These recordings contain up to 10 min of baseline activity before and after a seizure and have not been screened for artifacts. We compare our CHMM to classifiers evaluated on a framewise and channelwise basis. Our algorithm outperforms these baselines and demonstrates efficacy in classifying seizure intervals. Our algorithm provides localization information that could be useful for determining the seizure onset location in cases of focal epilepsy.

2   Generative Model of Seizure Propagation

We adopt a Bayesian framework for seizure detection. The latent variables $\mathbf{X}$ denote the seizure or non-seizure states. $\mathbf{Y}$ corresponds to observed data feature vectors computed from EEG channels as shown in Fig. 1(a). The random variable X_i^t denotes the latent state of EEG channel i at time t. We assume three possible states: pre-seizure baseline ($X_i^t = 0$), seizure propagation ($X_i^t = 1$), and post-seizure baseline ($X_i^t = 2$). The corresponding observed "emission" feature vectors Y_i^t are continuous statistics computed from time window t of the EEG channel i. For convenience, we also define the ensemble variables $\mathbf{X}^t \triangleq [X_1^t, \ldots, X_N^t]^T$ where N is the number of electrodes. Given the EEG observations, our goal is to infer the latent seizure state for each chain at all times.

2.1   Model Formulation and Inference

Figure 1(b) shows the coupling between electrodes of the 10/20 international system [9]. We define the aunts $au(\cdot)$ of a given node as the set of electrodes connected to it in Fig. 1(b). The joint distribution of $\mathbf{X}$ and $\mathbf{Y}$ factorizes into transition priors that depend on both a channel's own previous state and those of its aunts $P(X_i^t \mid \mathbf{X}_{au(i)\bigcup i}^{t-1})$, and emission likelihoods $P(Y_i^t \mid X_i^t)$ as in Eq. (1). For simplicity we assume that all recordings begin in a non-seizure state ($X_i^0 = 0 \forall i$).

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

Note that the observed emissions are conditionally independent given the latent states. Below, we detail the model formulation and inference algorithm.
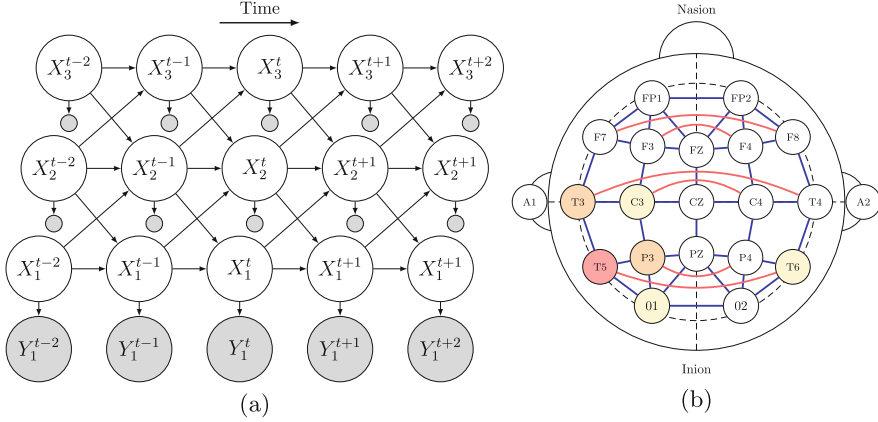
Fig. 1. (a) Graphical model depicting a three chain CHMM. Observed nodes are shaded gray, while latent nodes are shown in white. (b) EEG channels in the 10/20 international system [9], cross hemispheric (red) and neighboring (blue) channel connections. A seizure propagates from the red, to the orange, and finally yellow shaded electrodes.

Coupled State Transitions. The distribution over state vectors $\mathbf{X}^t$ forms a first order Markov chain. This distribution further factorizes into products of transition distributions of individual chains $P(\mathbf{X}^t \mid \mathbf{X}^{t-1}) = \prod_{i=1}^{N} P(X_i^t \mid \mathbf{X}_{au(i) \bigcup i}^{t-1})$. We encode these chainwise transition probabilities using time inhomogenous transition matrices as shown in Eq. (2). This structure ensures each channel begins in a non-seizure baseline state, transitions into an active seizure state, and transitions into a post-seizure state.

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

The transition matrix A_i^t is governed by neighboring and contralateral EEG channels to capture the main modes of seizure propagation, as shown in Fig. 1(b). Let η_i^t be the number of aunts in the seizure state in the previous timestep. We model the transition probabilities into and out of the seizure state via logistic regression functions of η_i^t as shown in Eq. (3). Parameters $\{\rho_0, \phi_0\}$ control the base onset and offset rates while $\{\rho_1, \phi_1\}$ control the effects of a channel's aunts.

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

Emission Likelihood. We use a Gaussian Mixture Model (GMM) to describe the emissions Y_i^t of each chain. Let C_i^t be the mixture from which Y_i^t was generated. Let π_{ij}^k be the prior probability of mixture component j when $X_i^t = k$ for $k = 0, 1, 2$. The joint distribution over Y_i^t and C_i^t can be expressed as follows

$$P(Y_i^t, C_i^t = j \mid X_i^t = k) = P(Y_i^t \mid C_i^t = j)P(C_i^t = j \mid X_i^t = k)$$
$$= \pi_{ij}^k \mathcal{N}\left(Y_i^t; \mu_{ij}, \Sigma_{ij}\right) \tag{4}$$

Effectively, the emission distributions for all observed variables share the same mean parameters μ_{ij} and covariance parameters Σ_{ij}, but use different mixture weights based on the latent seizure state k. We tie weights for both pre- and post-seizure states, i.e. $\pi_{ij}^0 = \pi_{ij}^2$ for all channels i and mixture components j. The data likelihood $P(Y_i^t \mid X_i^t)$ can be computed by marginalizing over j.

Approximate Inference Using Variational EM. Exact inference for the CHMM is intractable due to the coupled state transitions. Therefore we develop a structured variational algorithm [10], in which we approximate the posterior distribution over $\mathbf{X}$ as a set of N independent HMM chains:

$$Q(\mathbf{X}) = \prod_{i=1}^{N} \frac{1}{Z_{Q_i}} Q_i(\mathbf{X}_i) = \prod_{i=1}^{N} \frac{1}{Z_{Q_i}} \prod_{t=1}^{T} T_i^t(X_i^t \mid X_i^{t-1}) E_i^t(X_i^t). \tag{5}$$

As seen in Eq. (5), each approximating chain includes a normalizing constant Z_{Q_i}, a transition term $T_i^t(X_i^t \mid X_i^{t-1})$, and an emission term $E_i^t(X_i^t)$.

The transition distribution $T_i^t(X_i^t \mid X_i^{t-1})$ is encoded by a state transition matrix $\tilde{A}_i^t$ which mimics the structure of Eq. (2). Here $\tilde{g}_i^t$ and $\tilde{h}_i^t$ are variational transition parameters analogous to the original transition parameters g_i^t and h_i^t.

$$\tilde{A}_i^t = \begin{bmatrix} 1 - \tilde{g}_i^t & \tilde{g}_i^t & 0 \\ 0 & 1 - \tilde{h}_i^t & \tilde{h}_i^t \\ 0 & 0 & 1 \end{bmatrix} \tag{6}$$

In contrast to Eq. (6), the emission distribution $E_i^t(X_i^t)$ weighs the contribution of the observed data Y_i^t through variational parameters $\tilde{l}_{i0}^t$ and $\tilde{l}_{i1}^t$. Thus $E_i^t(X_i^t = 0, 2) = \tilde{l}_{i0}^t$ and $E_i^t(X_i^t = 1) = \tilde{l}_{i1}^t$.

We learn variational parameters for each chain by minimizing the free energy of the approximation. We perform this minimization by decoupling the free energy into expectations over a single channel and expectations over the remaining channels. The index "$-i$" in Eq. (7) denotes the set of channels excluding i.

$$\mathcal{FE} = -E_Q\left[\log p(\mathbf{X}, \mathbf{Y})\right] + E_Q\left[\log Q(\mathbf{X})\right]$$
$$= -E_{Q_i}\left[E_{Q_{-i}}\left[\log p(\mathbf{X}_i, \mathbf{Y}_i \mid \mathbf{X}_{-i}, \mathbf{Y}_{-i})\right]\right] + E_{Q_i}\left[\log Q_i(X_i)\right] \tag{7}$$
$$- E_{Q_{-i}}\left[\log p(\mathbf{X}_{-i}, \mathbf{Y}_{-i})\right] + E_{Q_{-i}}\left[\log Q_{-i}(X_{-i})\right]$$

Notice that the last line of Eq. (7) does not depend on the parameters of chain i, allowing a natural fixed point iteration over the parameters of a single chain while holding all other chains constant. This minimization fixes the variational parameters $\tilde{l}_i^t$ equal to the GMM likelihood of the observed data:

$$\tilde{l}_{i0}^t = p(Y_i^t \mid X_i^t = 0, 2), \qquad \tilde{l}_{i1}^t = p(Y_i^t \mid X_i^t = 1). \tag{8}$$

The updates for the variational transition parameters form logistic regressions based on the expected value of the original activations.

$$\log\left(\frac{\tilde{g}_i^t}{1-\tilde{g}_i^t}\right) = \rho_0 + \rho_1 E_{Q_{au(i)}}\left[\eta_i^t\right], \quad \log\left(\frac{\tilde{h}_i^t}{1-\tilde{h}_i^t}\right) = \phi_0 + \phi_1 E_{Q_{au(i)}}\left[\eta_i^t\right] \quad (9)$$

Once the variational parameters have been computed, the approximating distribution takes the form of an HMM where the $\tilde{l}$ parameters capture the likelihood of the data under each latent state. We can use the forward-backward algorithm [10] to compute the expected latent states $E_Q[X_i^t]$, the expected state transitions, and the expected number of aunts in the seizure state $E_Q[\eta_i^t]$.

Learning the Model Parameters. We use the expected values of the latent states and mixture components to update the transition parameters $\{\rho_i, \phi_i\}$ and the emission parameters $\{\mu_{ij}, \Sigma_{ij}, \pi_{ij}^k\}$. Let $\tau_i^t(j,k)$ be the expectation that channel i at time t is in state k with mixture j, we can update the emission parameters according to the soft counts of the occurrence of each mixture.

$$\mu_{ij} = \frac{\sum_{k=0}^2 \sum_{t=0}^T \tau_i^t(j,k) Y_i^t}{\sum_{k=0}^2 \sum_{t=0}^T \tau_i^t(j,k)}, \quad \Sigma_{ij} = \frac{\sum_{k=0}^2 \sum_{t=0}^T \tau_i^t(j,k)\left(Y_i^t - \mu_{ij}\right)^2}{\sum_{k=0}^2 \sum_{t=0}^T \tau_i^t(j,k)} \quad (10)$$

$$\pi_{ij}^0 = \pi_{ij}^2 = \frac{\sum_{t=0}^T \tau_i^t(j,0) + \tau_i^t(j,2)}{\sum_{j'} \sum_{t=0}^T \tau_i^t(j',0) + \tau_i^t(j',2)}, \quad \pi_{ij}^1 = \frac{\sum_{t=0}^T \tau_i^t(j,1)}{\sum_{j'} \sum_{t=0}^T \tau_i^t(j',1)} \quad (11)$$

The update for the transition parameters $\{\rho_i, \phi_i\}$ takes the form of a weighted logistic regression. We regress the expected η_i^t onto the expected transitions for each chain and use Newton's method to find the optimal transition parameters.

Implementation Details. We initialize our model by training the GMM emission distributions based on the expert annotations of seizure intervals. 3 emission mixtures resulted in a reasonable compromise between sensitivity and specificity. Transition parameters ρ_0, ρ_1, ϕ_0, and ϕ_1 were initialized to -7, 2, -3, and 0, respectively. This corresponds to expected seizures every 13 min lasting 15 s, channels turning on with a 7 fold increase per aunt node, and no cross channel influence for offset. Our EM proceedure is performed in an unsupervised fashion without further use of the labels. Model parameters are updated during the M-step of the EM algorithm. During inference, channels are updated sequentially until the scaled difference in $\mathcal{FE}$ converges to less than 10^{-4}.

2.2   Baseline Comparison

We compare our model to three alternative classification schemes. The first approach is to train a logistic regression function to distinguish between baseline and seizure intervals based on a linear combination of the EEG features. The second approach uses kernel support vector machines (SVMs) to learn a possibly nonlinear decision boundary in the EEG feature space that maximally separates the baseline and seizure conditions. Here we rely on a polynomial kernel. SVM classifiers have been used extensively for seizure detection [2,5,6]. Finally we

consider a GMM hypothesis testing scenario. This method trains GMMs for seizure and non-seizure states and classifies based on the ratio of the likelihoods under each GMM, roughly equating to our model with no transition prior.

3   Experimental Results

3.1   Data and Preprocessing

Our EEG data was recorded as part of routine clinical evaluation in the EMU of the Johns Hopkins Hospital. Our dataset consists of 90 seizures recordings from 15 patients with as much as 10 min of baseline before and after the seizure. Recordings were sampled at 200 Hz. We rely on expert clinical annotations denoting the seizure onset and offset to validate the performance of each method.

For preprocessing, each EEG channel was bandpass filtered through sequential application of fourth order Butterworth high and low pass filters at 1.6 Hz and 50 Hz respectively. This filtering mirrored clinical preprocessing practice for removing DC trends and high frequency components with no clinical relevance. In addition, a second order notch filter with $Q = 20$ was applied at 60 Hz to the EEG recordings to remove any remaining effect of the power supply.

We considered two emission features for analysis computed from channels in common reference: the sum of spectral coefficients in brain wave frequency bands and the log line length. Features were computed on windows of 1 s with 250 ms overlap. For spectral features, a short time Fourier transform was taken after the application of a Tukey window with shape parameter 0.25. The magnitudes of the STFT coefficients corresponding to frequencies in the theta (1–4 Hz), delta (4–8 Hz), alpha (8–13 Hz), and beta (13–30 Hz) bands were summed and the logarithm was taken, resulting in a length four feature vector. The log line length was computed as the logarithm of the sum of the absolute difference between successive samples i.e. given a signal s of length T, $\log L = \log\left(\sum_{i=0}^{T-1}|s(i+1) - s(i)|\right)$.

3.2   Seizure Detection Performance

We use a 5 fold cross validation strategy for evaluation. Four folds were used to train each model and detection was evaluated on the held-out fold. Each recording was randomly assigned to a fold independently of patient. For our model, the training phase was used to learn the emission and transition parameters. Table 1 summarizes the performance for each classifier based on the average accuracy of the testing fold. The sensitivity (TPR) and specificity (TNR) denote the prediction accuracy for seizure and non-seizure frames, respectively, computed across all channels. For the probabilistic classifiers (i.e. logistic regression, GMM, and CHMM), these rates are weighted by the posterior confidence of the classifier.

The transition prior allows our CHMM to place more confidence in contiguous regions exhibiting seizure-like activity. Figure 2(a) shows an example of our classifier correctly classifying the majority of the seizure across all channels. However, this confidence comes with a reduction in specificity, as the classifier tends
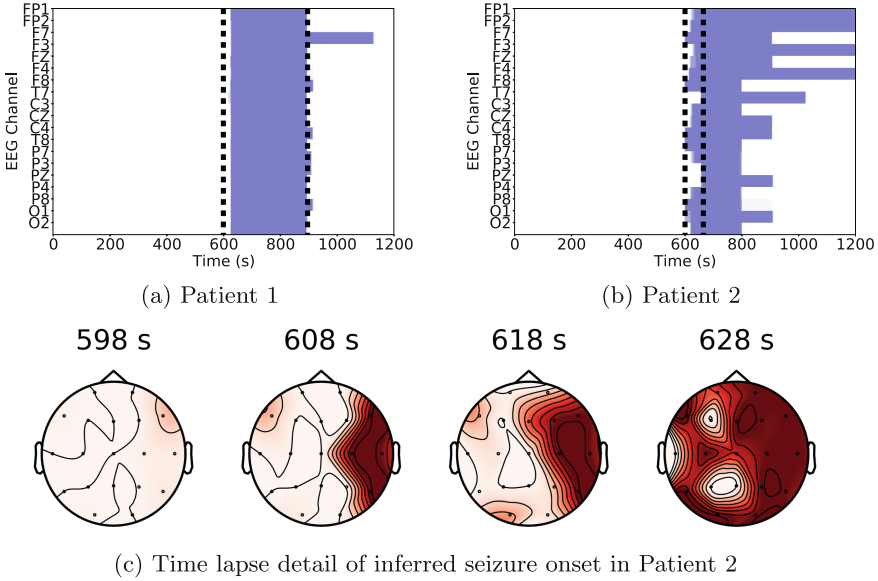
(a) Patient 1

(b) Patient 2

(c) Time lapse detail of inferred seizure onset in Patient 2

Fig. 2. Detection results for the CHMM. Top row: Estimated posterior probability of the latent seizure state for two epilepsy patients. White corresponds to pre- and post-seizure baseline while violet indicates seizure states. EEG channels corresponding to 10/20 system [9] channel locations are on the y axis. The expert identified seizure region is denoted by the dashed black lines. (a) shows the models ability to accurately classify seizures across the whole brain and (b) shows the outward spread of a right temporal lobe seizure. Bottom row: temporal evolution of the seizure depicted in (b).
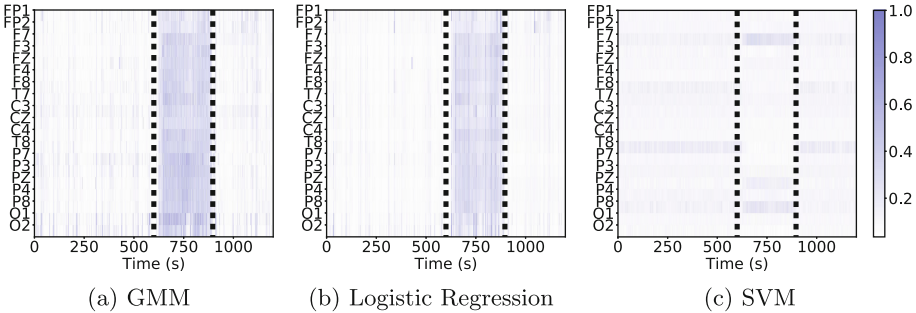

(a) GMM

(b) Logistic Regression

(c) SVM

Fig. 3. Detection results on Patient 1 for the three baseline methods. These algorithms place much lower posterior confidence in seizure intervals than the CHMM.

to associate post-seizure spectral artifacts with seizure as shown in Fig. 2(b). In future work we will investigate feature selection methods to combat this issue. Figure 2(c) shows the evolution of a focal right temporal seizure, which indicates our model's potential to localize epileptic activity on the scalp. This localization is highly relevant to clinical management of epilepsy.

Due to the heterogeneity of seizure presentations across patients, our baselines fail to perform well as shown in Fig. 3. The logistic regression and GMM correctly classify portions of seizure intervals but lack consistency. The GMM exhibits more confidence in classifying seizures than its probabilistic linear counterpart. The SVM performs poorly due to the inseperability of the EEG features in our noisy clinical dataset.

Table 1. Results for each method.

Model	TPR	TNR	AUC
GMM	21.91%	92.39%	0.784
Logistic regression	18.07%	**92.55%**	0.80
Kernel SVM	10.22%	90.27%	0.53
CHMM	**72.63%**	79.27%	**0.839**

4   Conclusion

We have presented a novel method for epileptic seizure detection based on a CHMM model. At a high level, we directly model seizure spreading by allowing the state of neighboring and symmetric EEG channels to influence the transition probabilities for a given channel. We have validated our approach on clinical EEG data from 15 unique patients. Our model outperforms three baseline approaches which perform classification on a framewise basis. By incorporating a transition prior that includes spatial and temporal contiguity to seizure regions we are able to better classify seizure intervals within EEG recordings.

References

1. Miller, J.W., Goodkin, H.P.: Epilepsy. Wiley, Hoboken (2014)
2. Acharya, U.R., et al.: Automated diagnosis of epileptic EEG using entropies. Biomed. Signal Process. Control. **7**(4), 401–408 (2012)
3. Güler, N.F., et al.: Recurrent neural networks employing Lyapunov exponents for EEG signals classification. Expert. Syst. Appl. **29**(3), 506–514 (2005)
4. Zandi, A.S., et al.: Automated real-time epileptic seizure detection in scalp EEG recordings using an algorithm based on wavelet packet transform. IEEE Trans. Biomed. Eng. **57**(7), 1639–1651 (2010)
5. Hunyadi, B., et al.: Incorporating structural information from the multichannel EEG improves patient-specific seizure detection. Clin. Neurophysiol. **123**(12), 2352–2361 (2012)
6. Shoeb, A.H., Guttag, J.V.: Application of machine learning to epileptic seizure detection. In: International Conference on Machine Learning, pp. 975–982 (2010)
7. Baldassano, S., et al.: A novel seizure detection algorithm informed by hidden Markov model event states. J. Neural Eng. **13**(3), 036011 (2016)
8. Brand, M., Oliver, N., Pentland, A.: Coupled hidden Markov models for complex action recognition. In: Proceedings of the 1997 IEEE Computer Society Conference on Computer vision and pattern recognition, pp. 994–999. IEEE (1997)
9. Jurcak, V., et al.: 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. Neuroimage **34**(4), 1600–1611 (2007)
10. Murphy, K.P.: Machine Learning: A Probabilistic Perspective. The MIT Press, Cambridge (2012)