Comparing SVM and Convolutional Networks for Epileptic Seizure Prediction from Intracranial EEG

Piotr W. Mirowski, *Member, IEEE*, Yann LeCun, Deepak Madhavan, and Ruben Kuzniecky.

Abstract— **Recent research suggests that electrophysiological changes develop minutes to hours before the actual clinical onset in focal epileptic seizures. Seizure prediction is a major field of neurological research, enabled by statistical analysis methods applied to features derived from intracranial Electroencephalographic (EEG) recordings of brain activity. However, no reliable seizure prediction method is ready for clinical applications. In this study, we use modern machine learning techniques to predict seizures from a number of features proposed in the literature. We concentrate on aggregated features that encode the relationship between pairs of EEG channels, such as cross-correlation, nonlinear interdependence, difference of Lyapunov exponents and wavelet analysis-based synchrony such as phase locking. We compare L1-regularized logistic regression, convolutional networks, and support vector machines. Results are reported on the standard Freiburg EEG dataset which contains data from 21 patients suffering from medically intractable focal epilepsy. For each patient, at least one method predicts 100% of the seizures on average 60 minutes before the onset, with no false alarm. Possible future applications include implantable devices capable of warning the patient of an upcoming seizure as well as implanted drug-delivery devices.**

I. Introduction

EPILEPSY is a chronic illness that affects approximately 1 to 2% of the world's population [1]. Among them, approximately 30-40% suffers from medication-refractory epilepsy, and may require surgical measures for either curative or palliative therapy. For those individuals, the prospect of experiencing unpredictable seizures during daily activity can be harrowing. A significant amount of research investigations have been recently pursued that focus on techniques that predict seizures prior to their onset. This is of marked value to patients with refractory epilepsy, and may allow them to take preparatory steps to protect themselves from injury or to attempt to take medication preventively. Additionally, seizure prediction mechanisms may play a critical value in implantable seizure prevention devices in the future (i.e. brain stimulators or intracranial drug delivery devices), signaling activation of these devices in the period prior to the impending seizure.

Manuscript received May 12, 2008. This work was supported in part by FACES (Finding A Cure for Epilepsy And Seizures), a non-profit organization of the NYU Medical Center.

P. W. Mirowski is with the Courant Institute of Mathematical Sciences, New York University, 719 Broadway, 12th Floor, New York, NY 10003, USA (e-mail: piotr.mirowski@computer.org).

Y. LeCun is with the Courant Institute of Mathematical Sciences, New York University, 719 Broadway, 12th Floor, New York, NY 10003, USA (e-mail: yann@cs.nyu.edu).

D. Madhavan is with the Department of Neurological Sciences, 982045 University of Nebraska Medical Center, Omaha, NE 68198, USA (e-mail: dmadhavan@unmc.edu).

R.. Kuzniecky is with the NYU Epilepsy Center, New York University School of Medicine, 403 East 34th Street, 4th Floor, New York, NY 10016, USA (e-mail: ruben.kuzniecky@med.nyu.edu).

A. Can Epileptic Seizures Be Predicted?

Recent multi-center clinical studies showed evidence of premonitory symptoms in 6.2% of 500 epileptic patients [2]. Another interview-based study found that 50% of 562 patients felt "auras" before seizures [3]. These clinical observations give an incentive to search for premonitory changes on Electroencephalographic (EEG) recordings from the brain. The epileptic brain consecutively transitions through different states of activity: from normal *interictal* (far from seizures) to *preictal* (minutes or sometimes hours before the seizure) then *ictal* (seizure) and *postictal*, before returning to the interictal state [4]. Despite the current lack of a complete neurological understanding of the preictal brain state, which is patient and condition specific, researchers increasingly hypothesize that brainwave synchronization patterns might differentiate interictal, preictal and ictal states [5].

The specific seizure prediction task thus becomes a classification problem where one aims at discriminating between interictal and preictal patterns of brain activity [4]. Ictal and postictal states are discarded from the classification because the task is not to detect undergoing seizures, but to warn the patient or clinician about future ones [4]. In this study, we arbitrarily define the preictal period as the 2 hours preceding a seizure.

B. Seizure Prediction Using Intracranial EEG

Each intracranial EEG electrode records local voltage potential over millions of brain cells. Multi-channel EEG can be viewed as an observation of a dynamical system generating electrophysiological waves.

Most current seizure prediction approaches can be summarized into (1) extracting *features* from EEG and (2) classifying them (and hence the patient's state) into preictal or interictal [4], [6].

The EEG is a temporal sequence of M-dimensional vectors $\mathbf{X}(t)$ in which each component $x_i(t)$ is a single electrode channel, at a fixed sampling rate. The data used in this paper has 6 channels sampled at 256Hz. We typically consider a window of N=1280 samples covering 5 seconds. The signal within this window is preprocessed to produce a vector of features $\mathbf{y}(t)$. A common hypothesis is that brainwave synchronization patterns are different in interictal and preictal phases [5]. For this reason, recent reviews of seizure prediction techniques [6]-[8] have advocated the use of *bivariate features* (which measure some relationship between two neighboring or distant EEG channels x_i and x_j), rather than *univariate features* (computed on each EEG channel separately).

C. Shortcomings of State-of-the-art Seizure Prediction

Despite over three decades of seizure prediction research, even the best methods suffer from a trade-off between *sensitivity* (being able to predict seizures) and *specificity* (avoiding false alarms). No method has achieved clinical applicability yet, with both a very high sensitivity and zero false alarms per hour [6]-[8]. We suggest two main limitation of existing seizure prediction algorithms: (1) unnecessary reduction of the number of features, and (2) the use of simplistic classification.

More specifically, the common approach is to average EEG-derived features (over time and/or over several EEG channels) and to ultimately perform binary classification of a single variable [6]. Binary classification consists in an a posteriori and in-sample tuning of a threshold (e.g. pre-ictal vs. interictal). In order to ensure the predictability, and in absence of testing data, this method requires sophisticated validation using the Seizure Time Surrogates method [9].

These typical shortcomings are illustrated in [10], where phase-locking synchrony is computed for all frequencies without band-pass filtering, and cross-correlation is computed for zero delay only. Bivariate features from several channels are collapsed to a single value. The final decision boundary is a simple line in a 2D space covered by the two bivariate features, unsurprisingly yielding very weak seizure prediction performance [10].

Machine Learning algorithms can alleviate these shortcomings, thanks to non-linear classification boundaries in a highly-dimensional features space, and by quantifying the efficiency of the learning process using in-sample learning (potentially with cross-validation) and out-of-sample testing. So far, only Genetic Algorithms [11], K-Means [12] and Quadratic Programming [13] have been applied, in a limited scope, to the seizure prediction problem, but only to select subsets of features and corresponding channels for further statistical classification, i.e. for data selection but not for the classification task itself.

As explained in sections II and III, our first contribution consists in aggregating bivariate features into temporally- and spatially-varying patterns. Our second contribution is to apply regularized machine learning methods (logistic regression, convolutional networks and support vector machines) to robustly classify these patterns of brain activity into interictal (away from seizures) and preictal (preceding seizures). We explain in section IV how we obtained 100% sensitivity and no false alarm seizure prediction performance on the reference Freiburg EEG database [14] containing EEG from 21 patients suffering from medically intractable focal epilepsy. Our results are considerably better than the 42% sensitivity and 3 false predictions per day reported on that dataset [15]-[19].

II. Bivariate Features From EEG

Bivariate features presented in this section and used in this study have the following common points:

(a) Bivariate features are computed on 5s windows (N=1280 samples at 256Hz) of any two EEG channels x_i and x_j.

(b) For M EEG channels, one computes features on $M \times (M-1)/2$ pairs of channels (e.g. 15 pairs for M=6).

(c) Features are aggregated for several consecutive time frames, e.g. 12 frames (1min) or 60 frames (5min).

This section details 4 different types of bivariate features.

A. Maximal Cross-Correlation [6]

The simplest linear measure of dependence between two signals is cross-correlation. Due to both the electrophysiological and cognitive nature of the brainwaves, delays between two spatially distant EEG signals are possible. To account for these delays, cross-correlation $C_{i,j}(\tau)$ between pairs (x_i, x_j) of EEG channels $x_i(t)$ and $x_j(t)$ is computed at delays τ ranging from -0.5s to 0.5s, and only the maximal value of $C_{i,j}(\tau)$ is retained [6], as in:

$$C_{ij} = \max_{\tau}\left\{\left|\frac{C_{i,j}(\tau)}{\sqrt{C_i(0)\cdot C_j(0)}}\right|\right\} \quad (2)$$

where:

$$C_{i,j}(\tau) = \begin{cases} \dfrac{1}{N-\tau}\displaystyle\sum_{t=1}^{N-\tau} x_i(t+\tau)x_j(\tau) & \tau \geq 0 \\ C_{j,i}(-\tau) & \tau < 0 \end{cases} \quad (3)$$

B. Nonlinear Interdependence [20]

Nonlinear Interdependence is another nonlinear bivariate statistics with good seizure prediction performance, and measures the distance, in state-space, between trajectories of two EEG channels [6], [20]. First, signals $x_i(t)$ and $x_j(t)$ from two EEG channels are time-delay embedded into respective vectors $\mathbf{x_i}(t)$ and $\mathbf{x_j}(t)$. We used embedding dimension d=10 and time-delay τ=6 samples or 23ms following suggestions in [6]. Then one computes an asymmetric statistic measuring the Euclidian distance, in reconstructed state-space, between trajectories described by $\mathbf{x_i}(t)$ and $\mathbf{x_j}(t)$. For each time point t, $\{t_1^i, t_2^i, \ldots, t_K^i\}$ are the time indices of the K=5 nearest neighbors of $\mathbf{x_i}(t)$, $\{t_1^j, t_2^j, \ldots, t_K^j\}$ are the time indices of the K nearest neighbors of $\mathbf{x_j}(t)$, and the nonlinear dependence of channel x_i on channel x_j averaged on N time points is:

$$S(x_i|x_j) = \frac{1}{N}\sum_{t=1}^{N}\frac{R(t,x_i)}{R(t,x_i|x_j)}, \text{ where:} \quad (4)$$

$$R(t,x_i)=\frac{1}{K}\sum_{k=1}^{K}\left\|\mathbf{x}_\mathbf{i}(t)-\mathbf{x}_\mathbf{i}\left(t_k^i\right)\right\|_2^2 \quad (5)$$

$$R(t,x_i|y_j)=\frac{1}{K}\sum_{k=1}^{K}\left\|\mathbf{x}_\mathbf{i}(t)-\mathbf{x}_\mathbf{i}\left(t_k^j\right)\right\|_2^2 \quad (6)$$

We use symmetric feature is $S_{ij}=(S(x_i|x_j)+S(x_j|x_i))/2$.

C. Difference of Short-Term Lyapunov Exponents [13]

A common measure of chaos is the estimation, from the observed time-delay embedded time-series $\mathbf{x_i}(t)$, of the largest Lyapunov exponent, i.e. the exponential rate of growth of an initial perturbation $\boldsymbol{\delta}\mathbf{x_i}(t)$. For EEG applications, short-time Lyapunov coefficients STL_{max} are typically [13] sampled and averaged on 5s windows, with time-delay τ=6 samples or 20ms, embedding dimension d=7 and evolution time Δt=12 samples or 47ms. Positive values of the largest Lyapunov exponent are an indication of a chaotic system, and this exponent increases with the unpredictability.

Whereas STL_{max} characterizes the chaotic behavior of a single EEG channel $x_i(t)$, the difference of STL_{max} values between two channels measures the convergence of chaotic behavior of the epileptic brain as it transits from the interictal to ictal state. A classification algorithm relying on differences of STL_{max} achieved very high sensitivity and specificity for the seizure detection problem.

A detailed implementation of the method is given in [13]. Assuming a perturbation evolution time Δt, the short-time Lyapunov exponent on channel x_i is estimated as:

$$STL_{\max}(x_i)=\frac{1}{N\Delta t}\sum_{t=1}^{N}\log_2\left|\frac{\delta\mathbf{x}_i(t+\Delta t)}{\delta\mathbf{x}_i(t)}\right| \quad (7)$$

The bivariate attribute is $DSTL_{ij}=|STL_{\max}(x_i)-STL_{\max}(x_j)|$.

D. Wavelet Analysis Based Measure of Synchrony [12]

An important family of bivariate features used in this study, which also yielded the best classification results, is phase synchrony. Frequency-specific and time-dependent phase $\varphi_{i,f}(t)$ and $\varphi_{j,f}(t)$ is extracted from the two respective EEG signals $x_i(t)$ and $x_j(t)$ using Wavelet Transform [21]. Statistics on the difference of phases between two channels are then made: phase-locking synchrony $SPLV$, entropy H of the phase difference and coherence Coh [21]. Phase-locking synchrony at frequency f is:

$$SPLV(f,x_i,x_j)=\left|\frac{1}{N}\sum_{k=1}^{N}\exp\left[i\varphi_{i,f}(t_k)-i\varphi_{j,f}(t_k)\right]\right| \quad (8)$$

In this study, wavelet-based synchrony is computed at several frequencies and averaged on 7 different frequency bands corresponding to so-called EEG rhythms: delta (below 4Hz), theta (4-7Hz), alpha (7-13Hz), low beta (13-15Hz), high beta (14-30Hz), low gamma (30-45Hz) and high gamma (65-100Hz). In other words, for a given time frame and given pair of channels, bivariate features consist in 7 synchrony values.

III. Classifying Bivariate EEG Feature Patterns

The first contribution of this study consists in using all the bivariate features for classification. Features are computed for every pair of EEG channels, for every window of 5s, (and, in the case of synchrony, for every frequency band). Features are not averaged over all the channels and/or over a 10min window [6]. To the contrary, they are aggregated to form highly dimensional patterns. In this study, patterns can be viewed as "movie" of 12 (1min) or 60 (5min) frames, each frame consisting of $M\times(M-1)/2$ pairs of EEG channels (possibly multiplied by 7 frequency bands), see Fig. 1. The dimensionality of the feature patterns ranges from 180 (e.g. cross-correlation on 1min windows, Fig. 1), to 6300 (e.g. wavelet phase-locking synchrony on 5min windows). Let us note $\mathbf{y}_t$ one pattern (sample of bivariate features, and z_t the associated label (-1 for preictal, 1 for interictal). $\mathbf{y}_t$ can either be one long vector or a matrix indexed by time and by channel pair and frequency band.

The second innovation relies on nonlinear classification methods such as convolutional networks (section III.B) or support vector machines with Gaussian kernels (section III.C), even if a simple logistic regression yields good results in several patients (section III.A).

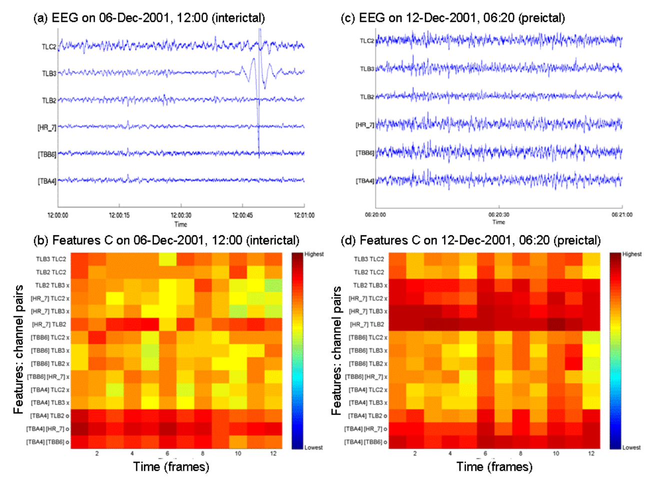

Fig. 1. Examples of EEG and bivariate features on 1min windows for Patient 12: (a) 1min of interictal EEG and (b) corresponding interictal cross-correlation features, (c) 1min of preictal EEG and (d) corresponding preictal cross-correlation. On the feature patterns, time frames are on the horizontal axis, and channel pairs on the vertical axis.

A. Logistic Regression

The decision function of a logistic regression classifier is

$$\bar{z}_t=\mathrm{sign}\left(\mathbf{w}^T\mathbf{y}_t+b\right), \quad (9)$$

where $\mathbf{w}$ is the weight parameter vector and b the bias. The optimal weight is obtained by minimizing loss function (10):

$$L(\mathbf{y}_t,z_t,\mathbf{w},b)=2\log\left(1+e^{-z_t\left(\mathbf{w}^T\mathbf{y}_t+b\right)}\right)+\lambda|\mathbf{w}| \quad (10)$$

Loss (10) comprises an L1 regularization term to select features. We minimize it using stochastic gradient descent [22]. A sensitivity analysis to the inputs can be performed by simply looking at individual weights w_i. We typically used values of 0.001 for lambda.

B. Convolutional Networks

Convolutional networks [23] are trainable, multi-layer, non-linear systems that are specifically designed to extract and classify high-dimensional patterns from images or multivariate time-series. They can be seen as multi-layer neural networks in which each layer is a bank of finite-impulse response filters followed by point-wise sigmoid squashing functions. All the layers are trained simultaneously using a version of the back-propagation learning algorithm. They can learn low-level features and high-level representations in an integrated manner. Their main advantage is that they can learn optimal time-invariant local feature detectors from input matrix $\mathbf{y_t}$ (which is indexed by time) and thus build representations that are robust to time shifts of specific feature motifs. This technique has already been applied to raw EEG data [24].

We use specific convolutional network architecture similar to [23]. For 12-frame patterns, the convolution kernels on the 1st, 3rd and 5th layers have sizes 1x5:*N*x3:1x1, with 1x2:1x2 subsampling on the 2nd and 4th layer. For 60-frame patterns, the corresponding convolution kernels have sizes 1x13:*N*x9:1x1, with 1x2:1x2 subsampling on the 2nd and 4th layer. In both cases, the second dimension is time and N is the number of pairs of channels (15) times the number of frequency bands (1 or 7). The thicknesses are 5:5 on the 1st and 3rd layer. There are 2 output nodes (one for preictal, one for interictal), with respective target output vectors [1,-1] and [-1,1], and the loss is the mean square distance to the prototypes. Having 2 outputs nodes enabled an asymmetric learning that penalized more false positives (false preictal alarms) than false negatives (missed preictal alarms), as illustrated in section IV.

Moreover, we evaluate the sensitivity of particular input features for the classification task, by back-propagating the gradients obtained for each testing sample onto the inputs, and then by summing the squares of these gradients on inputs. The input sensitivity analysis yields results comparable to logistic regression weights.

C. Support-Vector Machines

We use in this study support vector machines [25] with Gaussian kernels, implemented in the LibSVM library [26].

5-fold cross-validation helps to select the best regularization parameter C and the Gaussian parameter γ, using first a coarse grid search on $C \in \left[2^{1}, 2^{3}, \ldots, 2^{9}\right]$ and $\gamma \in \left[2^{-13}, 2^{-13}, 2^{-1}\right]$ followed by a refined grid search for $C \in C_0 \times \left[2^{-2}, 2^{-1.5}, \ldots, 2^{2}\right]$, and $\gamma \in \gamma_0 \times \left[2^{-2}, 2^{-1.5}, \ldots, 2^{-2}\right]$ around the best candidates C_0, γ_0.

IV. Results

A. 21 Patient EEG Database

To test our machine-learning based classification of bivariate features from EEG, we use the Freiburg EEG database [14] containing invasive EEG recordings of 21 patients suffering from medically intractable focal epilepsy. These EEG data had been acquired from intracranial grid-, strip-, and depth-electrodes at a 256 Hz sampling rate, and digitized to 16 bit by an analogue-to-digital converter. Then, three focal electrodes (1-3) were chosen from areas involved early in ictal activity, and three remaining electrodes (4-6) were selected as not involved during seizure spread.

We further apply Infinite Impulse Response (IIR) elliptical filters [27] to clean some artifacts: a 49-51Hz band-reject 12th-order filter to remove power line noise, a 120Hz cutoff low-pass 1st-order filter to avoid aliasing, and a 0.5Hz cutoff high-pass 5th-order filter to remove the dc component.

B. Training and Testing Data Selection

Each of the patients EEG recordings contains between 2 and 6 seizures and at least 50 min of pre-ictal data for most seizures, as well as approximately 24 hours of EEG-recordings without seizure activity and spanning the full wake-sleep cycle. Once bivariate features are computed for each patient, we select samples from the last 1 or 2 seizures (depending on their total number) and 33% of the interictal samples as testing data, and the rest as training data. In other words, we train the classifiers on the earlier seizures and on wake-sleep interictal data, and evaluate these classifiers on later seizures and on different wake-sleep interictal data.

All data samples are scaled on a per patient basis, to either zero mean and unit variance (logistic regression and convolutional networks) or between -1 and 1 for SVMs.

C. Classification Results on a the Freiburg Dataset

Because the ultimate goal is more the epileptic patient's quality of life rather than the classification task itself, seizure prediction performance is measured in terms of false positives (alarms) per hour and of sensitivity (number of seizures where at least one preictal sample is correctly classified).

Table 1 summarizes, for each patient, feature type and classifier, the false positive rate and time to first (and second) seizure, i.e. "how early a preictal alarm is sent", or how early is any preictal sample correctly classified. All results reported in Table 1 have less than 0.3 false positives per hour (as computed on total interictal time only) and each of the testing seizure has been predicted, i.e. there is at least one preictal alarm during the 2 hours preceding each seizure. In Table 1, crosses mark combinations of features and machines for which seizure prediction fails, i.e. with more than 0.3 false positives per hour or seizures not predicted. Results in bold indicate zero false alarm and 100% seizure prediction.

Although we tested 1min-long (12 frames) and 5min-long (60 frames) patterns, we reported in Table 1 only results obtained with 5min-long patterns.

Not all preictal samples are correctly classified: some are misclassified (false negatives) as interictal. We actually use the possibility of having some false negatives to achieve good seizure prediction results in patients 17, 19 and 21. During the stochastic gradient descent training of the convolutional network, we set a stronger penalty on false positives than on false negatives. These results are indicated in italic in Table 1.

The main conclusion is that **for each patient, at least one method predicts 100% of the seizures on average 60 minutes before the onset, with no false alarm**. The minimum prediction time is 3 minutes for patient 13, but has to be put on the account of very short preictal recording before the second seizure for that patient.

Convolutional networks achieve a zero-false-alarm seizure prediction on 20 patients out of 21, compared to 11 only using SVM (good results are obtained for patient 5 though, contrary to convolutional networks). Surprisingly, the linear classification boundary of logistic regression enables perfect seizure prediction on 14 patients.

It has to be noted that both for convolutional networks and logistic regression, the classification performance is 100% on the training dataset for reported results. The only exception are patients 17, 19 and 21, where we allow a larger penalty for false positives than for false negatives and obtain false negatives on the training dataset. In any case, we can claim that we obtain 100% sensitivity and no false positives on the full 88-seizure Freiburg dataset.

D. Feature Selection and Input Sensitivity

By enforcing sparse parameters through L1 regularization, and by observing the logistic regression weights or the convolutional networks input sensitivity, we see that only a subset of feature inputs is necessary, and that high frequency inputs are necessary for good classification (Fig. 2).

Cross-correlation features enable good seizure predictions with less than 0.3 false positives per hour on 12 patients out of 21; nonlinear interdependence on 16, wavelet phase-locking synchrony, phase difference entropy on 14, and wavelet coherence on 18 patients. The difference of short-term Lyapunov exponents is the weakest, as it works only on 2 patients. Unsurprisingly, 5min patterns achieve a better performance that 1min patterns (working respectively on 21 vs. 15 patients). Although 5min patterns have more dimensions than the number of training samples, using separate training, cross-validation and testing datasets enables generalization properties of the classifiers.

		pat 1		pat 2		pat 3			pat 4			pat 5		
		fpr	ts1	fpr	ts1	fpr	ts1	ts2	fpr	ts1	ts2	fpr	ts1	ts2
C	log	x	x	x	x	x	x	x	x	x	x	x	x	x
	lenet5	**0**	**68**	**0**	**40**	x	x	x	**0**	**54**	**61**	**0**	**25**	**52**
	svm	0.23	68	**0**	**40**	x	x	x	x	x	x	x	x	x
S	log	x	x	x	x	**0**	**48**	**3**	**0**	**54**	**61**	x	x	x
	lenet5	**0**	**68**	**0**	**40**	**0**	**48**	**3**	**0**	**54**	**61**	x	x	x
	svm	0.23	68	**0**	**40**	x	x	x	0.13	39	61	**0**	**45**	**52**
DSTL	svm	x	x	x	x	x	x	x	**0**	**39**	**51**	x	x	x
SPLV	log	**0**	**68**	**0**	**40**	**0**	**48**	**3**	**0**	**54**	**61**	x	x	x
	lenet5	**0**	**68**	**0**	**40**	**0**	**48**	**3**	**0**	**54**	**61**	x	x	x
	svm	0.12	68	**0**	**40**	**0**	**48**	**3**	**0**	**54**	**41**	x	x	x
H	log	x	x	**0**	**40**	**0**	**48**	**3**	**0**	**54**	**61**	x	x	x
	lenet5	**0**	**68**	**0**	**40**	**0**	**48**	**3**	**0**	**54**	**61**	x	x	x
	svm	0.23	68	**0**	**40**	**0**	**48**	**3**	**0**	**54**	**61**	x	x	x
Coh	log	**0**	**68**	**0**	**40**	**0**	**48**	**3**	**0**	**54**	**61**	x	x	x
	lenet5	**0**	**68**	**0**	**40**	**0**	**48**	**3**	**0**	**54**	**61**	**0**	**45**	**52**
	svm	0.12	68	**0**	**40**	**0**	**48**	**3**	**0**	**54**	**61**			

		pat 6		pat 7		pat 8		pat 9			pat 10			pat 11	
		fpr	ts1	fpr	ts1	fpr	ts1	fpr	ts1	ts2	fpr	ts1	ts2	fpr	ts1
C	log	x	x	**0**	**46**	x	x	x	x	x	**0**	**79**	**73**	x	x
	lenet5	x	x	**0**	**56**	x	x	x	x	x	x	x	x	x	x
	svm	0.12	66	**0**	**36**	x	x	x	x	x	0.12	79	73	x	x
S	log	x	x	**0**	**56**	x	x	x	x	x	x	x	x	x	x
	lenet5	x	x	**0**	**56**	x	x	**0**	**51**	**78**	x	x	x	**0**	**67**
	svm	0.12	16	**0**	**56**	**0**	**9**	0.13	51	43	0.12	79	73	0.25	67
DSTL	svm	x	x	x	x	x	x	x	x	x	0.24	9	3	x	x
SPLV	log	**0**	**66**	**0**	**56**	x	x	**0**	**51**	**78**	x	x	x	**0**	**57**
	lenet5	x	x	**0**	**56**	**0**	**39**	**0**	**51**	**78**	**0**	**79**	**73**	**0**	**67**
	svm	0.12	66	**0**	**56**	x	x	**0**	**51**	**78**	0.24	79	73	**0**	**27**
H	log	x	x	**0**	**56**	x	x	**0**	**51**	**78**	x	x	x	**0**	**67**
	lenet5	x	x	**0**	**56**	x	x	**0**	**51**	**78**	x	x	x	**0**	**67**
	svm	0.12	66	**0**	**56**	x	x	**0**	**51**	**78**	0.24	79	73	**0**	**27**
Coh	log	**0**	**66**	**0**	**56**	x	x	**0**	**51**	**78**	x	x	x	**0**	**37**
	lenet5	**0**	**71**	**0**	**56**	**0**	**44**	**0**	**51**	**78**	**0**	**79**	**73**	**0**	**67**
	svm	0.12	66	**0**	**56**	x	x	**0**	**51**	**78**	0.24	79	73	**0**	**32**

		pat 12		pat 13		pat 14		pat 15		pat 16		
		fpr	ts1	fpr	ts1	fpr	ts1	fpr	ts1	fpr	ts1	ts2
C	log	**0**	**25**	**0**	**2**	x	x	x	x	x	x	x
	lenet5	**0**	**25**	**0**	**7**	x	x	x	x	**0**	**65**	**25**
	svm	**0**	**25**	x	x	x	x	x	x	**0**	**60**	**20**
S	log	**0**	**25**	x	x	x	x	x	x	x	x	x
	lenet5	**0**	**25**	x	x	x	x	x	x	x	x	x
	svm			x	x	0.13	33	0.12	90	**0**	**55**	**55**
DSTL	svm			x	x	x	x	x	x	x	x	x
SPLV	log	**0**	**25**	x	x	x	x	x	x	x	x	x
	lenet5	**0**	**25**	x	x	x	x	**0**	**90**	x	x	x
	svm			x	x	0.26	33	**0**	**80**	x	x	x
H	log	**0**	**25**	x	x	**0**	**33**	**0**	**70**	x	x	x
	lenet5	**0**	**25**	x	x	**0**	**33**	**0**	**90**	x	x	x
	svm			x	x	0.13	33	**0**	**85**	x	x	x
Coh	log	**0**	**25**	x	x	x	x	**0**	**45**	**0**	**60**	**10**
	lenet5	**0**	**25**	x	x	x	x	**0**	**90**	x	x	x
	svm			x	x	0.26	28	**0**	**85**	**0**	**60**	**5**

		pat 17			pat 18			*pat 19*		pat 20			*pat 21*		
		fpr	ts1	ts2	fpr	ts1	ts2	fpr	ts1	fpr	ts1	ts2	fpr	ts1	ts2
C	log	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	lenet5	x	x	x	x	x	x	x	x	***0***	***91***	***96***	x	x	x
	svm	x	x	x	x	x	x	x	x	x	x	x	0.12	99	70
S	log	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	lenet5	x	x	x	x	x	x	***0***	***28***	**0**	**91**	**96**	x	x	x
	svm	x	x	x	x	x	x	x	x	x	x	x	x	x	x
DSTL	svm	x	x	x	x	x	x	x	x	x	x	x	x	x	x
SPLV	log	x	x	x	x	x	x	x	x	x	x	x	***0***	***99***	***75***
	lenet5	x	x	x	**0**	**20**	**70**	***0***	***28***	x	x	x	x	x	x
	svm	x	x	x	x	x	x	x	x	x	x	x	0.12	99	80
H	log	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	lenet5	***0***	***73***	***##***	x	x	x	x	x	x	x	x	x	x	x
	svm	x	x	x	x	x	x	x	x	x	x	x	0.12	14	75
Coh	log	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	lenet5	x	x	x	***0***	***25***	***90***	x	x	**0**	**99**	**20**	x	x	x
	svm	x	x	x	0.23	15	90	x	x	x	x	x	0.12	99	75

Table 1. False positive rates *fpr* and time of alarm before seizure *ts1* and *ts2* obtained on the testing dataset of 21 patients [14], using *log-reg* logistic regression-, *lenet5* convolutional networks- and *svm* support vector machines-based seizure prediction (classification of patterns of bivariate features into preictal and interictal). Features considered are maximal cross-correlation *C*, nonlinear interdependence *S*, difference of Lyapunov exponents *DSTL*, and three measures of wavelet analysis-based synchrony: phase-locking value *SPLV*, entropy *H* of phase difference and coherence *Coh*. Input patterns consisted of 5min (60 frames) of these features. For each patient, different classifiers and different features yielded different results. Crosses x mark combinations of features and machines for which seizure prediction failed, i.e. with more than 0.3 false positives per hour or seizures not predicted. Results in bold indicate zero false alarm and 100% seizure prediction. Results in italic indicate special training of the convolutional network.

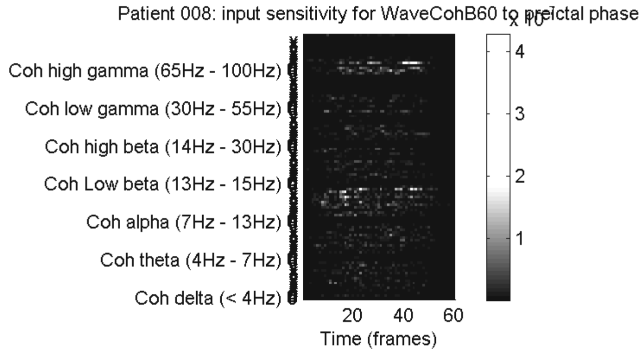

Fig. 2. Input sensitivity analysis for the seizure prediction task on patient 8 and on 60-frame patterns of wavelet-based coherence. Time frames are on the horizontal axis, and channel pairs on the vertical axis, sorted by 7 frequency bands. Thanks to L1 regularization, only some inputs play a significant role. We observe for this patient and set of features that the time dependence is much weaker than dependence on the frequency band and channel pair, and that high frequency coherence plays a significant role in the preictal/interictal classification task.

E. Implementation

The software for computing features from EEG is implemented in Matlab but can also be run under its free open-source counterpart, Octave. Support vector machine classification is performed using LibSVM [26] and its Matlab/Octave interface. Convolutional networks are implemented Lush [28], an open-source programming environment.

In order to translate seizure prediction research into engineering of human neuroprosthetic devices, we need to ensure a low computational complexity. Trained SVMs and convolutional networks can be run in real-time on test patterns, but the features detailed in this article are computationally expensive. We are currently investigating the use of neural networks to rapidly measure EEG synchronization.

Acknowledgment

P. W. Mirowski thanks Dr. Nandor Ludvig of the NYU Medical Center for useful advice and discussion about EEG.

References

[1] P. Kwan, M. J. Brodie, "Early Identification of Refractory Epilepsy", New England Journal of Medicine, n.342, 2000, pp.314-319.

[2] A. Schulze-Bonhage, C. Kurth, A. Carius, B. J. Steinhoff, T. Mayer, "Seizure anticipation by patients with focal and generalized epilepsy: a multicentre assessment of premonitory symptoms", *Epilepsy Research*, n.70, 2006, pp.83-88.

[3] P. Rajna, B. Clemens, E. Csibri, "Hungarian multicentre epidemiologic study of the warning and initial symptoms (prodrome, aura) of epileptic seizures", *Seizure*, n.6, 1997, pp.361–68.

[4] K. Lehnertz, B. Litt, "The first international collaborative workshop on seizure prediction: summary and data description", *Clinical Neurophysiology*, n.116, 2005, pp.493-505.

[5] M. Le Van Quyen, V. Navarro, J. Martinerie, M. Baulac, F. J. Varela, "Toward a Neurodynamical Understanding of Ictogenesis", *Epilepsia*, n.44, supp.12, 2003, pp.30-43.

[6] F. Mormann, T. Kreuz, C. Rieke, R. G. Andrzejak, A. Kraskov, P. David, C. E. Elger, K. Lehnertz, "On the predictability of epileptic seizures", *Clinical Neurophysiology*, n.116, 2005, pp.569-587.

[7] F. Mormann, C. E. Elger, K. Lehnertz, "Seizure anticipation: from algorithms to clinical practice", *Current Opinion in Neurology* n.19, 2006, pp.187-193.

[8] K. Lehnertz, F. Mormann, H. Osterhage, A. Müller, J. Prusseit, et al, *Journal of Clinical Neurophysiology*, vol.24, n.2, April 2007.

[9] R. G. Andrzejak, F. Mormann, T. Kreuz, C. Rieke, A. Kraskov, C. E. Elger, K. Lehnertz, "Testing the null hypothesis of the non-existence of the pre-seizure state", *Physical Review E*, 67, 2003.

[10] K. K. Jerger, S. L. Weinstein, T. Sauer, S. J. Schiff, "Multivariate linear discrimination of seizures", *Clinical Neurophysiology*, n.116, 2005, pp.545-551.

[11] M. D'Alessandro, R. Esteller, G. Vachtsevanos, A. Hinson, J. Echauz, et al, *IEEE Transactions in Biomedical Engineering*, vol.50, n.5, 2003, pp.603-615.

[12] M. Le Van Quyen, J. Soss, V. Navarro, R. Robertson, M. Chavez, M. Baulac, J. Martinerie, "Preictal state identification by synchronization changes in long-term intracranial recordings", *Clinical Neurophysiology*, n.116, 2005, pp.559-568.

[13] L. D. Iasemidis, D. S. Shiau, P. M. Pardalos, W. Chaovalitwongse, K. Narayanan, A. Prasad, K. Tsakalis, P. R. Carney, J. C. Sackellares, "Long-term prospective online real-time seizure prediction", *Clinical Neurophysiology*, n.116, 2005, pp.532-544.

[14] EEG Database at the Epilepsy Center of the University Hospital of Freiburg, Germany, available: https://epilepsy.uni-freiburg.de/freiburg-seizure-prediction-project/eeg-database/

[15] B. Schelter, M. Winterhalder, T. Maiwald, A. Brandt, A. Schad, et al, "Do False Predictions of Seizures Depend on the State of Vigilance? A Report from Two Seizure-Prediction Methods and Proposed Remedies", *Epilepsia*, n.47, 2006, pp.2058-2070.

[16] B. Schelter, M. Winterhalder, T. Maiwald, A. Brandt, A. Schad, et al, "Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction", *Chaos*, n.16: 013108, 2006.

[17] T. Maiwald, M. Winterhalder, R. Aschenbrenner-Scheibe, H. U. Voss, A. Schulze-Bonhage et al, "Comparison of three nonlinear seizure prediction methods by means of the seizure prediction characteristic", *Physica D*, n.194, 2004, pp.357-368.

[18] R. Aschenbrenner-Scheibe, T. Maiwald, M. Winterhalder, H. U. Voss, J. Timmer, et al, "How well can epileptic seizures be predicted? An evaluation of a nonlinear method", *Brain*, n.126, 2003, pp.2616-2626.

[19] M. Winterhalder, T. Maiwald, H. U. Voss, R. Aschenbrenner-Scheibe, J. Timmer et al, "The seizure prediction characteristic: a general framework to assess and compare seizure prediction methods", *Epilepsy Behavior*., vol.4 n.3, 2003, pp.318-325.

[20] J. Arnhold, P. Grassberger, K. Lehnertz, C. E. Elger, "A robust method for detecting interdependence: applications to intracranially recorded EEG", Physica D, vol.134, 1999, pp.419-430.

[21] M. Le Van Quyen, J. Foucher, J.-P. Lachaux, E. Rodriguez, A. Lutz, J. Martinerie, F. J. Varela, "Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony", *Journal of Neuroscience Methods*, n.11, 2001, pp.83-98.

[22] L. Bottou, "Stochastic gradient learning in neural networks", *Proceedings of Neuro-Nîmes 91*, EC2, Nîmes, France, 1991.

[23] Y. Le Cun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-Based Learning Applied to Document Recognition", Proceedings of the IEEE, vol.86, n.11, 1998, pp.2278-2324.

[24] P. W. Mirowski, D. Madhavan, Y. LeCun, "Time-Delay Neural Networks and Independent Component Analysis for EEG-Based Prediction of Epileptic Seizures Propagation", *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence*, July 22-26, 2007, AAAI Press 2007.

[25] C. Cortes, V. Vapnik, "Support-Vector Networks", *Machine Learning*, vol.20, n.3, 1995, pp.273-297.

[26] C.-C. Chang, C.-J. Lin, "LIBSVM: a library for support vector machines", 2001, available: http://www.csie.ntu.edu.tw/~cjlin/libsvm

[27] Delorme A., Makeig S., "EEGLab: an open-source toolbox for analysis of single-trial EEG dynamics including ICA", Journal of Neuroscience Methods, vol.134, n.1, 2004, pp.9-21.

[28] Bottou L., LeCun Y., 2002, "Lush : Lisp Universal Shell programming language", available: http://lush.sourceforge.net