Highlights

- We present a patient-specific seizure detection algorithm conveying structural information from the multichannel EEG using a novel machine learning technique applied to epileptic EEG data for the first time.
- The proposed system automatically incorporates crucial spatial information specific to the patient's seizure and facilitates improved detection performance.
- The proposed system reduces the number of required seizure examples, thus the training time needed to tune the system to individual patients

Incorporating structural information from the multichannel EEG improves patient-specific seizure detection

Borbala Hunyadi[a,b,*], Marco Signoretto[a,b], Wim Van Paesschen[c], Johan A.K. Suykens[a,b], Sabine Van Huffel[a,b], Maarten De Vos[a,b,d]

[a] *SCD - SISTA, Department of Electrical Engineering - ESAT, Katholieke Universiteit Leuven*
[b] *IBBT Future Health Department*
[c] *Department of Neurology, University Hospital Gasthuisberg, Leuven, Belgium*
[d] *Neuropsychology Lab, Department of Psychology, University of Oldenburg, Oldenburg, Germany*

Abstract

Objective: A novel patient-specific seizure detection algorithm is presented. As the spatial distribution of the ictal pattern is characteristic for a patient's seizures, this work aims at incorporating such information into the data representation and provides a learning algorithm exploiting it. *Methods:* The proposed training algorithm uses nuclear norm regularization to convey structural information of the channel-feature matrices extracted from the EEG. This method is compared to two existing approaches utilizing the same feature set, but integrating the multichannel information in a different manner. The performances of the detectors are demonstrated on a publicly available dataset containing 131 seizures recorded in 892 hours of scalp EEG from 22 pediatric patients. *Results:* The proposed algorithm performed significantly better compared to the reference approaches (p=0.0170 and p=0.0002). It reaches a median performance of 100% sensitivity, 0.11/h false detection rate and 7.8s detection delay, outperforming a method in the literature using the same dataset. *Conclusion:* The strength of our method lies within conveying structural information from the multichannel EEG. Such

*SCD - SISTA, Department of Electrical Engineering, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium. e-mail: borbala.hunyadi@esat.kuleuven.be

formulation automatically includes crucial spatial information and improves detection performance. *Significance:* Our solution facilitates accurate classification performance for small training sets, therefore, it potentially reduces the time needed to train the detector before starting monitoring.

Keywords: seizure detection, EEG, structural information, convex optimization, nuclear norm, LS-SVM

---

1. Introduction

Epilepsy, the second most common neurological disorder after stroke, occurs in over 0.5% of the world population (De Boer et al. (2008)). This disease is manifested through recurrent epileptic seizures, resulting from an abnormal, synchronous activity in the brain involving a large network of neurons. Depending on the brain regions involved in the seizure, the patient may have diverse clinical symptoms, including sensory dysfunction, loss of consciousness, motor automatisms, etc. Approximately 30% of epilepsy patients are non-responsive to anti-epileptic drugs (refractory epilepsy) (see Engel (1996)), hence their quality of life is seriously compromised and surgical resection of the epileptic focus has to be considered.

The electroencephalogram (EEG) measures the electrical activity of the brain and is a well-established technique in epilepsy diagnosis and monitoring. As EEG monitoring takes several days, the amount of information to be processed is huge. The benefits of the automatic analysis of EEG recordings are twofold. An automatic seizure detection technique supporting visual analysis would drastically decrease the workload of clinicians. Besides, due to its fine temporal resolution, EEG can provide accurate information about the onset of the seizure. As the seizure spreads quickly through the brain, the early detection of the seizure is essential. Preparing the patient for SPECT scan by injecting a contrast agent early in the course of the seizure helps revealing the focus of the epilepsy, and facilitates a successful surgical outcome.

The large inter-patient variability regarding seizure characteristics poses a great challenge for developers of a universal, automatic seizure detector. However, in case of unifocal epilepsy patients, patient-specific systems are more promising, due to the rather consistent intra-patient seizures regarding both temporal pattern and localization. We propose a seizure detector capable of exploiting this crucial structural information underlying the mul-

tichannel EEG data, therefore, providing better performance and requiring a lower amount of seizure data for training than other approaches.

Several existing seizure detection algorithms act on single channel data (e.g. Deburchgraeve et al. (2008), De Vos et al. (2011) Polychronaki et al. (2010), Zandi et al. (2010)). Methods attempting to integrate multichannel information include two-step systems, where, in the first step a decision is made for each channel by a separate classifier, and in the second step the outputs of these classifiers serve as the input of a combined, final decision procedure (e.g. Qu and Gotman (1997); Greene et al. (2008)) compared such a late integration method to an early integration method, where the features extracted from each channel are sorted and stacked into a long feature vector, which is then used to train a single classifier. The early integration method was shown to be superior in performance, by preserving the synchronously recorded nature of the EEG and exploiting the inherent inter-relationship of the channels. Shoeb and Guttag (2010) developed a patient-specific seizure detector, which relies on features describing the temporal evolution, the spectral and the spatial structure of the EEG. In order to capture spatial information, the features of each channel are concatenated to form one feature vector. As opposed to the former study, where the sorting operation was intended to remove spatial information, the goal of the stacking in this case is to drive the attention to the locations corresponding to the channels consistently showing seizure activity.

In the present paper a novel solution is investigated. The features extracted from the multichannel EEG are represented in the form of a matrix which serves as an input to a classifier. The matrix representation of the data helps preserving and exploiting the inherent spatial structure of the multichannel EEG data. Moreover, recent studies (He et al. (2006), Tao et al. (2007)) show that matrix and tensor representation of signals reduce the small sample-size problem, facilitating a precise classification performance even for low number of training points and outperform traditional vector representation.

The paper is organized as follows. In section 2 the EEG dataset used in this study is presented. Afterwards, the extracted features are described, together with three different strategies on integrating the feature information extracted from the individual channels: the early integration and late integration approaches, and the feature-channel matrix we propose. Further, the different machine learning solutions utilizing these data representations are explained. At last, the applied performance evaluation measures are enu-

merated. Section 3 presents the results of the study used to compare the performances of the different seizure detection approaches. Finally, section 4 and 5 is devoted to the discussion and concluding remarks.

2. Materials and Methods

Fig. 1 depicts the consecutive steps of the entire procedure involved in training and testing the seizure detectors. The following sections describe each step in detail.

2.1. Training and testing EEG data

The seizure detectors were evaluated using the CHB-MIT database, which includes the scalp EEG recordings of 23 pediatric patients, and is available online at PhysioNet (`http://physionet.org/physiobank/database/chbmit`, Goldberger et al. (2000)). All patients of the dataset were included in our study except for one, in case of which the entire electrode montage was altered during the recordings. A number of channels were changed in case of some other patients as well. In these cases 18 channels were considered, which remained unchanged throughout the whole duration of the recording. In the rest of the patients, all 23 channels of the electrode montage was used. Our dataset, therefore, consists of the scalp recordings of 22 patients. A second dataset was recorded 1.5 years after the first recording of patient 1. This recording was handled separately, resulting in 23 distinct datasets. A total of 131 seizures were recorded during 892 hours of monitoring. The data were sampled at 256Hz and a bipolar electrode montage was used. Detailed description of the dataset can be found online (see the link above).

In order to assess the capabilities of the seizure detectors trained with different amount of seizure information, several experiments were performed. Mutually non-overlapping training and testing datasets were created in each experiment for each patient in the following way. The seizures of each patient were ordered according to the times of their onset. In each experiment, seizures which occurred earlier in time were included in the training set, and the rest of the seizures were added to the test set. For instance, if four seizures were recorded from a patient, three experiments were performed. In the first experiment one seizure was included in the training, and three in the test; in the second experiment two seizures were included in the training and two in the test; finally, in the third experiment three seizures were included in the training and one in the test. Maximum five experiments were performed per

patient, even if the number of recorded seizures would have allowed larger training sets.

Non-seizure datapoints for the training set were selected from the EEG recorded over a twenty-four hour period, as this period covers the different brain functioning states connected to sleep and other activity levels affecting EEG (see Qu and Gotman (1997)). In order to limit the computational costs, but at the same time cover as many different EEG patterns as possible, one minute long segments recorded every fifteen minutes were included for the training.

The seizure detectors were tested on continuous EEG data recorded after the twenty-four hour period used for constructing the training set. All available seizure data, which were not included in the training were added to the test dataset, resulting in a total of 474 hours of test data, on average 21.56 hours per patient, ranging from 1.13 to 125.94 hours.

2.2. Preprocessing and feature extraction

Seizures often involve involuntary movements and muscle contractions, which contaminate EEG data seriously and hinder interpretation. Muscle artifacts were automatically removed applying canonical correlation analysis for blind source separation (BSS-CCA) (see De Clercq et al. (2006); De Vos et al. (2010)). Afterwards, each EEG channel was band-pass filtered between 1-30 Hz. The signals were segmented into 2s long non-overlapping epochs and a total of sixteen features were extracted from each channel of each epoch. The extracted features, listed in Table 1, are well established and widely used in the literature, not only for adult scalp EEG (Meier et al. (2008); Minasyan et al. (2010); Qu and Gotman (1997)), but also for intracranial recordings (Tito et al. (2009)) or neonatal seizure detection (Temko et al. (2011)). The definition and computation of the features are described in Appendix A. The feature set corresponding to one epoch (a total of 16 times 23 or 18 features) will be referred to as datapoint throughout this paper.

2.3. Integration of multichannel information

A distinct feature set is extracted from each EEG channel, however, the multichannel information has to be integrated in one global decision in each epoch. We compared three different integration approaches. The first two are described and studied in detail in Greene et al. (2008), the third one is a novel approach we propose and evaluate in this paper. The different

integration approaches result in different data representation, and require different learning and classification schemes. The various stages of the three algorithms are depicted in Figure 2, and are explained in detail below.

2.4. Early and late integration with least-squares support vector machines (LS-SVM)

2.4.1. Late integration (LI)

Traditional seizure detection systems classify the feature set of each channel separately, and then combine the channel outcomes in an independent step. There are several different strategies that can be followed. The outputs of the channel classifiers can be binary or probabilistic. Post-processing can be performed applying a moving average filter on the outputs from the consecutive epochs (Thomas et al. (2009)). A channel detection might automatically induce an alarm (Deburchgraeve et al. (2008)); or the channel outputs can be integrated via mean, max, min score, or majority vote (Greene et al. (2008)). The number of channels contributing to the global score might as well be limited (Saab and Gotman (2005)), or weighted, this way including prior spatial information. The final classifier output may be converted to posterior probabilities using a sigmoid function (Temko et al. (2011)). In the current study p number of LS-SVM classifiers (see 2.4.3) are trained, where p corresponds to the number of EEG channels. During monitoring the feature sets of each channel are classified independently. The continuous outputs of the single channel classifiers are integrated by taking their maximum. The final decision is made by comparing the integrated output to the detection threshold b_{LI}. The choice of the detection threshold is explained below in 2.6.

2.4.2. Early integration (EI)

In this approach the feature vectors extracted from each EEG channel are stacked into one long feature vector of length $p \cdot d$, where p is the number of channels and d is the number of extracted features. As explained above, stacking the channel vectors ensures that the synchronously recorded and inter-dependent nature of multichannel EEG is preserved in the input patterns, and therefore exploited in the learning procedure. Furthermore, the channels are concatenated in fixed order, aiming at including spatial information characteristic of the patient's seizures. The stacked feature vectors are used to train one global LS-SVM classifier (see 2.4.3). During monitoring the feature set of each new epoch is processed by the global classifier and the

continuous outputs are compared to the detection threshold b_{EI} (see 2.6) in order to obtain the final decision.

2.4.3. LS-SVM

A support vector machine (SVM) is a widely used universal learning machine, which works based on the following principle: Let $\{x_k, y_k\}_{k=1}^N$ be the training dataset with $x_k \in \mathbb{R}^d$ input samples and $y_k \in \{-1, 1\}$ class labels. In our case the input samples are the feature vectors extracted from the training EEG epochs, which are labelled as seizure or non-seizure examples. Our aim is to train a classifier based on the input samples, which will assign a class label to each new EEG epoch. The classifier is defined by

$$y(x) = \text{sign}[w^T \varphi(x) + b], \tag{1}$$

where w is a weighting vector and φ is the feature map which maps the input data to a higher-dimensional feature space. The objective in the SVM formulation is to construct a separating hyperplane in the feature space with maximal margin. This can be translated to the following optimization problem:

$$\begin{aligned} &\min_{w,e_k,b} \frac{1}{2} w^T w + C \sum_{k=1}^{N} e_k, \\ &\text{subject to} \\ &y_k[w^T \varphi(x) + b] \geq 1 - e_k,\ e_k \geq 0,\ k = 1, ..., N \end{aligned} \tag{2}$$

where C is a regularization constant. By constructing the Lagrangian and taking the conditions for optimality, one obtains the dual problem, which is the following quadratic programming problem:

$$\begin{aligned} &\max_{\alpha} \sum_{k=1}^{N} \alpha_k - \frac{1}{2} \sum_{k=1}^{N} \sum_{l=1}^{N} \alpha_k \alpha_l y_k y_l K(x_k, x_l), \\ &\text{subject to} \\ &\sum_{k=1}^{N} \alpha_k y_k = 0 \ \text{ and } 0 \leq \alpha_k \leq C \text{ for } k = 1, ..., N, \end{aligned} \tag{3}$$

where $x_k \in \mathbb{R}^d$ and $K(x, x_k) = \varphi(x)^T \varphi(x_k)$ is a symmetric and positive definite kernel function satisfying the Mercer theorem. The x_k input vectors corresponding to non-zero α_k values are called support vectors. The classifier in the dual space takes the form:

$$y(x) = \text{sign}[\sum_k \alpha_k y_k K(x, x_k) + b] \tag{4}$$

In the current paper we applied the least-squares support-vector machine introduced by Suykens and Vandewalle (1999). In this formulation equality constraints are used instead of inequality constraints in (2):

$$\begin{aligned} &\min_{w,e_k,b} \frac{1}{2} w^T w + \gamma \sum_{k=1}^{N} e_k^2, \\ &\text{subject to} \\ &\hat{y}_k[w^T \varphi(x) + b] = 1 - e_k, \end{aligned} \tag{5}$$

together with a squared loss, which simplifies the computation to solving a set of linear equations instead of a quadratic programming problem:

$$\left[\begin{array}{c|c} 0 & y^T \\ \hline Y & \Omega + \gamma^{-1} I \end{array}\right] \left[\begin{array}{c} b \\ \hline \alpha \end{array}\right] = \left[\begin{array}{c} 0 \\ \hline 1_N \end{array}\right] \tag{6}$$

with $y = [y_1, ..., y_N]^T$, $1_N = [1, ..., 1]^T$, $e = [e_1, ..., e_N]^T$, $\alpha = [\alpha_1, ..., \alpha_N]^T$, $\Omega_{kl} = y_k y_l K(x_k, x_l)$. In the LS-SVM case all values of α are non-zero, consequently, the model is based on the whole input dataset - which can be beneficial in the current application due to the small training set size. For the same reason, and due to the relatively high dimensionality of the input data, a linear kernel was applied. However, as RBF kernel is often used for seizure detection (e.g. Temko et al. (2011); Shoeb and Guttag (2010)), it has also been tested. The LS-SVMlab v1.8 toolbox implementation (www.esat.kuleuven.be/sista/lssvmlab, De Brabanter et al. (2011)) was used in this study, which performs automatic model selection, namely determining the tuning parameters by coupled simulated annealing (see Xavier de Souza et al. (2010)). As optimization criterion five-fold crossvalidated misclassification error was used.

2.5. Feature-channel matrix with nuclear norm learning (NNL)

2.5.1. Feature-channel matrix

This paper presents a novel approach, where the features extracted from each epoch are preserved in their original arrangement, in the form of a matrix of size $d \times p$, where d is the number of features and p is the number of channels. Each row of the matrix contains the feature set extracted from the corresponding channel. There is a large variability between the seizure patterns of different patients. However, synchronization between EEG channels is a generally occurring characteristic (Varsavsky et al. (2011)). Synchronization may be observed on a few channels in case of a partial seizure, or on a larger scale given a generalized seizure. Similar EEG pattern, therefore, similar feature values will be present on those channels and in the rows of the matrix corresponding those channels, which are involved in the seizure. Representing the data in matrix form allows to exploit the common information among the channels. Channels not involved in the seizure might show various patterns of background activity, not distinguishable from patterns of normal brain state. Thus, one can assume that the channels involved in the seizure and the features most descriptive about the seizure pattern will have the best discriminative power. As we expect common information on the seizure channels

2.5.2. Singular value decomposition

In order to establish the method we propose, the concept of singular value decomposition has to be introduced. Any matrix of real or complex values can be written as the following:

$$M = U\Sigma V^*, \tag{7}$$

where U and V are unitary matrices, V^* is the conjugate transpose of V, and Σ is a diagonal matrix. The diagonal elements of Σ, denoted as σ_i, are in decreasing order, and are called the singular values of M. The columns of U and V are called the left and right singular vector of M, and denoted as U_i and V_i, respectively. Alternatively, M can be written as the weighted sum of their outer products:

$$M = \sum_i \sigma_i U_i \times V_i, \tag{8}$$

The number of non-zero singular values is the rank of the matrix. Truncating the summation at the first r elements gives the best rank-r approximation of M in least-squares sense.

2.5.3. Convex optimization with nuclear norm penalty

For classification of the feature-channel matrix extracted from an EEG epoch X, we consider the following linear model:

$$\hat{y}(A,b) = \langle A, X\rangle + b, \tag{9}$$
$$\langle A, X\rangle = \sum_{i,j} A_{i,j} X_{i,j}$$

where $X \in R^{d\times p}$ is the input pattern, $A \in R^{d\times p}$ is the classifier matrix of the same size, and b is a bias term or threshold. Decisions are made according to $\text{sign}(\hat{y}) \in \{-1, 1\}$, where 1 and -1 correspond to seizure and non-seizure epoch, respectively. The classifier, namely the pair (A, b), is found by solving a non-smooth convex optimization problem using nuclear norm penalty:

$$\min_{A,b} F(A,b) = f(A,b) + \mu||A||_*, \tag{10}$$

$$f(A,b) = \sum_{k=1}^{N} (\hat{y_k}(A,b) - y_k)^2, \tag{11}$$

where $f(A, b)$ is the quadratic error function accounting for the misclassification. The choice of a quadratic error function was made specifically because the same loss function is used in LS-SVM classification. Further, $\mu \geq$ is the regularization parameter. Model selection, namely the selection of the μ was done according to the five-fold cross-validation of the misclassification error, similarly to the model selection of LS-SVM. Finally, $||A||_*$ is the nuclear norm of the matrix A with singular values σ_i :

$$||A||_* = \sum_i \sigma_i. \tag{12}$$

Regularization via nuclear norm conveys structural information and facilitates an approximately low-rank solution. It has been used to devise convex relaxation for various rank-constrained matrix problems (Candes and Recht (2009); Recht et al. (2007)), for matrix classification (Tomioka and Aihara

(2007)) and it has been extended to the case of higher order arrays as a heuristic for multilinear rank constrained problems (Signoretto (2011)).

The consecutive steps of the nuclear norm learning approach is summerized in Appendix B. The algorithm solving (10) can be implemented in `CVX` (Grant and Boyd (2011)), as given in Appendix C.

2.6. Optimal detection threshold

Seizures are rare events, consequently, there is a large amount of non-seizure data available for training compared to seizure data. Classification problems with skewed class sizes are called unbalanced problems and need special attention. The Bayesian formulation used by Saab and Gotman (2005) inherently incorporates seizure and non-seizure class probabilities. Gardner et al. (2006) bypasses this issue by reformulating the classification problem as a novelty detection problem, applying a one-class support vector machine (SVM). Nandan et al. (2010), applied support vector data description (SVDD), which can make use of ictal data as well, yielding better detection performance. Traditional SVM approaches addressing the same problem also exist. In the work of Temko et al. (2011) the classifier outputs are converted to posterior probabilities using a sigmoid function. This technique is proven to perform significantly better than applying a threshold on the original outputs (Platt (1999)).

Alternatively, a bias term correction can be used (Suykens et al. (2002)), by adjusting the model thresholds b_{LI}, b_{EI} and b_{NNL} in order to achieve optimal classification performance. Due to the small amount of available seizure data, the detection threshold was set using the training dataset, similarly to Lukas et al. (2002). In this work we used misclassification error as optimality criterion. First, the classification outputs obtained for the training samples were collected. From this discrete set, the value minimizing the number of falsely classified training datapoints was selected and assigned as detection threshold.

2.7. Evaluation measures

The receiver operating characteristic (ROC) curve reflects the discriminative power of a classifier by depicting its sensitivity versus its specificity for every possible threshold value. A good classifier has high sensitivity along with high specificity, resulting in a large area under the ROC curve (AUC).

In order to characterize the performance of the classifiers applied in a seizure alarm system, the outputs obtained for the individual epochs have to

be postprocessed and converted to alarms. The outputs were first converted to {0,1} values with a signum function. A detection was made, if a value 1 was obtained this way for at least five consecutive epochs. Detections occurring less than 30 seconds after each other were grouped together and one single alarm was set off.

The following event-based measures are used to evaluate the performances of the detectors:

1. Detection sensitivity

$$dS = \frac{\text{number of detected seizures}}{\text{total number of seizures}}$$

2. False alarm rate

$$R_{fa} = \frac{\text{number of false alarms}}{\text{duration of the EEG}}$$

3. Alarm delay

$$D = \text{time of alarm setoff} - \text{time of seizure onset}$$

4. Quality value

$$QV = \frac{1}{(R_{fa} + 0.2) \times (dS \times D + (1 - dS) \times 60)}$$

The quality value was introduced by Qu and Gotman (1997) in order to compare different seizure detectors based on one measure reflecting all the three above enumerated aspects of the performance. Sensitivity and detection delay are taken into account by creating a weighted average detection delay. When the weighted average detection delay or the false detection rate decreases, the quality value increases. Thus, the detector with the highest quality value is the best one.

3. Results

Figure 4(a), depicting the classifier obtained for Patient 1, illustrates the nature of the structural information conveyed by nuclear norm regularization and its physiological meaning. The 16 columns correspond to the different features extracted from the 23 channels, which constitute the rows of the matrix. The classifier is well approximated by a matrix of rank 1,

the first singular value carries 97.6% of the total energy (see Figure 4(b)). The values of the left and right singular vectors represent the relative discriminative power of the channels and features, respectively. Note that the highest channel entries (electrodes over the right temporal and parietal area), and highest feature entry (normalized power in the theta band) characterize well the seizure pattern and localization as observed on Figure 3. Nuclear norm regularization induced a solution where the feature entries follow a similar pattern in each channel involved in the seizure, in agreement with the considerations in 2.5.1.

In comparison, the classifier matrix obtained by EI (Figure 4(c)) is less structured. Its singular values decay slower, the first singular value carrying 47.6% of the total energy. There are more channel and feature entries present with high values, and the feature patterns observed on the channels are diverse. This intuitively means that more EEG pattern details are considered in the classifier. Subtle characteristics of the training seizures might not be present in other seizures. This leads to overfitting and decreased predictive performance. While it separates better the training data than NNL does(EI: AUC=0.9991 and NNL: AUC=0.9966), EI performs worse on the testing data (EI: AUC=0.9616 and NNL: AUC=0.9718).

The degree of structure in the classifier is controlled by the regularization parameter μ. This parameter is optimized for predictive performance in the crossvalidation. Hence, if a low-rank solution is not favorable, a low μ value would be set and the quadratic error function would dominate the expression in 10. In such cases the classifier would not be enforced to be approximately low rank. Nevertheless, in the majority (20 out of 23) of the cases the NNL approach results in more structured classifiers than the EI approach (Figure 3).

The overall performance of NNL, EI and LI detection approaches are compared in Table 2. Median and mean (in brackets) values of sensitivity, false detection rate and alarm delay were computed for the test data of the whole patient population. The false detection rate of NNL approach is three times lower than the EI approach, and its sensitivity and alarm delay is just slightly worse. The false detection rate of the LI approach lies between the one of NNL and EI. At the same time its sensitivity is lower than both other approaches, despite the fact that the OR function integrating the channel decisions facilitates the highest possible sensitivity.

It is worth to observe how the performance on the test data evolves due to additional training seizures. The detection sensitivity increases in all cases,

nevertheless, at the cost of a higher false alarm rate. It is important to note, though, that while the mean false alarm rates of the EI and LI approaches exceed 0.5/h (a value not acceptable in clinical applications see Qu and Gotman (1997)), the false alarm rate of NNL approach always remains below.

The inevitable trade-off between sensitivity and specificity explains why some seizures were missed by the NNL and not by the EI approach. Figure 6 shows the classifier outputs obtained by the EI 6(a) and NNL approach 6(b) during the first seizure of Patient 8. The output time courses of both classifiers follow a similar pattern, indicating comparable discriminative power. However, the seizure was missed by the NNL, due to the relatively high detection threshold compared to the one of the EI approach. The detection threshold was set according to the training data, and is only an estimation of the optimal threshold. Observing the classifier outputs of Patient 8, this detection threshold could be decreased, facilitating better sensitivity, but still preserving high specificity.

Due to the ambiguous results regarding the comparison of sensitivity and false detection rate of the NNL and EI approaches, a measure reflecting to both aspects is needed. Therefore, the quality value introduced by Qu and Gotman (1997) was applied to the results of each patient. The boxplot in Figure 7 shows the distribution of the test performance differences between the NNL and EI over the individual patients.

Wilcoxon signed rank tests were performed investigating whether any approach provides a significant improvement in the median quality value with respect to the other approaches. The test results are reported in Table 3, where the P-values below the significance level $\alpha = 0.05$ are indicated in boldface letters. The results show that the approaches utilizing multichannel information (NNL and EI) provide significant improvement compared to the LI approach, which processes the channels independently. Improved performance was obtained also in cases where the NNL and the EI was trained with less seizures than the LI approach. Moreover, the NNL approach performed significantly better than the EI given two and three training seizures; also when the EI approach was trained with one additional seizure.

Shoeb and Guttag (2010) report the performance of a seizure detector evaluated on the same EEG dataset, reaching a median of 2 false detections daily. Using one training seizure, their detector reaches 55% sensitivity on five randomly selected patients, claimed to be representative for the dataset. Given one training seizures the NNL algorithm presented here produces only 1.5 false detections daily, and reaches a median detection sensitivity of 75%,

confirming the superiority of our approach.

4. Discussion

We presented a novel seizure detection algorithm and compared it to two existing approaches. All the three solutions make use of the same feature set; they differ in how the individual channel information is integrated into a final decision. The late integration approach processes the channel information independently, and integrates the channel decisions with an OR function. The early integration approach stacks the features extracted from all the channels in one long feature vector, which is used to train the classifier. Finally, the nuclear norm approach constructs channel-feature matrices and exploits structural information from these during the learning procedure.

All the three approaches facilitated successful seizure detection. However, our results suggest that methods analyzing the multichannel EEG as one entity perform better than approaches analyzing the channels independently. This confirms the results presented by Greene et al. (2008), where the early integration and late integration approaches were compared in a non-patient specific setting. Moreover, enforcing a low-rank structure by nuclear norm regularization is proven to bring additional improvement compared to the early integration approach, where the inter-channel information is synthesized without postulating any underlying model.

Our seizure detection system performed better than a reference method reported by Shoeb and Guttag (2010), tested on the same online available EEG dataset. Direct comparison with other patient-specific studies in the literature is not feasible, as they use different datasets, moreover, those datasets often contain preselected patients (e.g. Qu and Gotman (1997); Minasyan et al. (2010)). In order to demonstrate the real-life clinical applicability of our method, we chose to avoid any preselection whatsoever.

The detector presented in this work is a patient-specific system. It achieves high performance by capturing spatial information from the available multichannel seizure recordings. The spatial information is incorporated by assuming a low-rank model underlying the seizure epochs, defined by the relative discriminative power of the features and their distribution over the channels. Our model and the learning algorithm is closely related to the concept of singular value decomposition. Similar structural assumptions were made by De Vos et al. (2007), where the epileptic seizures were successfully

localized by canonical decomposition, a possible tensor extension of singular value decomposition.

The localization can be expected to be consistent among the seizures of patients with unifocal epilepsy. However, seizures of multifocal epilepsy patients will have different spatial distributions and will not all be detected by our algorithm, unless seizures of all types occurring in the patient are available in the training.

A real-life application requires that the seizure detection system collects sufficient information already from a few seizures, and can start monitoring within a short time. Thus, our classifiers are trained using small amount of seizure datapoints. The consequence of this is twofold. This results in a highly unbalanced training set, where, in addition, the dimensionality of one datapoint is relatively large compared to the number of available positive samples. Such system is prone to overfitting, nevertheless, the representation and model assumption used within the nuclear norm learning approach helps overcoming this issue.

Due to the unbalanced input data, additional bias term correction has to be performed after the training to obtain the correct detection threshold. Considering the low amount of available seizure samples, the bias term correction was performed on the training data. The classification performance was assessed by computing the misclassification error. We have tested several alternative optimality criteria, which take into account the skewed class ratios, such as F2 score, balanced error rate, or a certain working point on the ROC curve. However, none of them showed significant improvement compared to the misclassification error.

The performance of our algorithm may be further improved by applying additional features, however, finding an optimal feature set is beyond the scope of this paper. The features used in this study are successfully applied in the literature, and they are developed to characterize repetitive, rhythmic seizure patterns. Nevertheless, seizures which are difficult to assess visually, due to the lack of such patterns or artifacts, will also be poorly described with these features. On the contrary, different data representations capturing more complex interactions between the channels might improve detection in these difficult cases as well.

Besides the morphology, spectral structure and spatial distribution, the temporal evolution is an important characteristic of the seizure pattern as well (Stern and Engel (2004); Shoeb and Guttag (2010)). A possible extension of the presented algorithm is to take into account temporal evolution

by analyzing consecutive EEG epochs together, forming a tensor from their feature-channel matrices. Both training and detection are performed on such tensorial data. The size of the tensors are $n \times m \times k$, with n and m are the number of channels and extaced features, respectively, and k is the number of consecutive EEG epoch analyzed at once. When choosing an appropriate value for k, one has to take into account that a large k will delay the detection. On the other hand, small k captures less information. The nuclear norm learning algorithm is directly applicable to tensorial data, as explained in Signoretto et al. (2011).

We presented successful seizure detection by using linear models, e.i. linear separation of seizure and non-seizure data. Although a non-linear separation might be better suited for this problem, the low amount of available seizure information does not allow the use of non-linear kernels. The early integration approach has been tested with an RBF kernel, however, it performed worse compared to the other approaches, and the test results are not shown here. Nonetheless, future work will be focused on structure-preserving tensorial kernel methods, which were shown to be relevant in solving such problems (see Signoretto et al. (2011)).

This paper presents for the first time the nuclear norm learning approach applied for seizure detection. There might be several related EEG applications where this technique can be beneficial. For instance, it is directly applicable in brain-computer interfaces based on event-related potentials or event-related synchronization and desynchronization. Such responses are consistent in their waveform, frequency content and spatial distribution among the trials, consequently, they can also be modeled with low-rank structures. Using the nuclear norm learning approach could reduce the training time necessary for reaching successful discrimination of these brain responses.

5. Conclusion

The strength of our method lies within conveying structural information from the multichannel EEG data. Such formulation allows to automatically include crucial spatial information characteristic to a patient's seizures. Our results show improved detection performance compared to alternative approaches, even if less seizure information is available for training. Therefore, the presented method potentially reduces the time needed to train the detection system before starting monitoring.

Acknowledgements

Research supported by Research Council KUL: CoE EF/05/006, GOA MaNet, PFV/10/002 (OPTEC); Flemish Government: projects: G.0427.10N (Integrated EEG-fMRI), G.0108.11 (Compressed Sensing); IWT: TBM070713-Accelero, TBM080658-MRI (EEGfMRI), TBM110697-NeoGuard; iMinds 2012; Belgian Federal Science Policy Office: IUAP DYSCO; JS acknowledges support from FWO: G.0377.12; Europan Research Council: ERC AdG A-DATADRIVEB; MDV acknowledges support from Alexander von Humboldt postdoctoral stipend.

Appendix A. Extracted features

Time domain features were extracted directly from the preprocessed EEG window:

- Number of zero crossings (#0), maxima (#max) and minima(#min)
- Skewness
$$\text{skew} = \mathbb{E}[(\frac{A-\mu}{\sigma})^3] \tag{A.1}$$
where A is the EEG time series, μ is the mean and σ is the standard deviation and $\mathbb{E}$ denotes the expected value.
- Kurtosis
$$\text{kurt} = \mathbb{E}[(\frac{A-\mu}{\sigma})^4] \tag{A.2}$$
- Root mean square amplitude:
$$\text{rmsa} = \sqrt{\frac{1}{N}\sum_{k=1}^{N} A(k)^2} \tag{A.3}$$

The frequency domain features were extracted from the power spectral density ($S(f)$) computed by Welch's method:

- Total power:
$$TP = \sum_{f_i=1}^{30} S(f_i), \tag{A.4}$$

- Peak frequency

$$PF = \arg \max_{f \in [1,30]} S(f) \tag{A.5}$$

- Mean power in delta, theta, alpha and beta frequency bands:

$$D = \sum_{f_i=1}^{3} S(f_i), \tag{A.6}$$

$$T = \sum_{f_i=4}^{8} S(f_i), \tag{A.7}$$

$$A = \sum_{f_i=9}^{13} S(f_i), \tag{A.8}$$

$$B = \sum_{f_i=14}^{20} S(f_i) \tag{A.9}$$

- Normalized power in delta, theta, alpha and beta frequency bands:

$$nD = \frac{\sum_{f_i=1}^{3} S(f_i)}{\sum_{f_i=1}^{30} S(f_i)}, \tag{A.10}$$

$$nT = \frac{\sum_{f_i=4}^{8} S(f_i)}{\sum_{f_i=1}^{30} S(f_i)}, \tag{A.11}$$

$$nA = \frac{\sum_{f_i=9}^{13} S(f_i)}{\sum_{f_i=1}^{30} S(f_i)}, \tag{A.12}$$

$$nB = \frac{\sum_{f_i=14}^{20} S(f_i)}{\sum_{f_i=1}^{30} S(f_i)} \tag{A.13}$$

Appendix B. Algorithmic steps of the nuclear norm learning approach

Algorithm 1: Nuclear norm learning

Input: multichannel EEG recording

Output: classifier matrix A and bias term b

1. create the input data $X \in R^{d \times p \times N}$ and labels $y \in R^{N \times 1}$
  preprocess EEG
  segment EEG to 2s windows
  for each N 2s window
    extract d features from each channel
    order the feature vectors in p rows to create $p \times d$ matrix
    assign corresponding label
    stack into X
  end

2. model selection: obtain μ
  for cv-loop
    split X into training and validation set
    for all μ_i values in a predifned range
      solve Eq. (10) on the training set to obtain A, b
      evaluate on the validation set to obtain misclassification
    end
  end

3. select μ with the best corresponding misclassification

4. solve Eq. (10) with with μ to obtain A, b

Appendix C. CVX implementation of the nuclear norm regularization

The `CVX` implementation of the algorithm solving equation 10 is as follows:

```
cvx_begin
    variables e(N) A(I,J)  b
    minimize(1/2*e'*e+reg_par*norm_nuc(A))
    subject to
    diag(y)*(X*A(:)+b)==ones(N,1)-e
cvx_end
```

where X is a $N \times (IJ)$ matrix (the generic row n is the vectorization of the $n-$th input data matrix of size $I \times J$), A is the $I \times J$ classifier matrix and b is a scalar bias term; y is the vector of labels. For consistency we have reported a constrained formulation of the problem that mimic the one used in the LS-SVM primal problem, see 5.

References

Candes, E., Recht, B.. Exact matrix completion via convex optimization. Foundation of Computational Mathematics 2009;9 (6):717–772.

De Boer, H.M., Mula, M., Sander, J.W.. The global burden and stigma of epilepsy. Epilepsy Behav 2008;12:540 – 6.

De Brabanter, K., Karsmakers, P., Ojeda, F., Alzate, C., De Brabanter, J., Pelckmans, K., De Moor, B., Vandewalle, J., Suykens, J.A.K.. LS-SVMlab Toolbox User's Guide version 1.8. Technical Report; ESAT-SISTA, K.U.Leuven; 2011.

De Clercq, W., Vergult, A., Vanrumste, B., Van Paesschen, W., Van Huffel, S.. Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram. IEEE Trans Biomed Eng 2006;53(12):2583 – 87.

De Vos, M., De Lathauwer, L., Vanrumste, B., Van Huffel, S., Van Paesschen, W.. Canonical decomposition of ictal scalp eeg reliably detects the seizure onset zone. NeuroImage 2007;37(3):844–854.

De Vos, M., Deburchgraeve, W., Cherian, P.J., Matic, V., Swarte, R.M., Govaert, P., Visser, G.H., Van Huffel, S.. Automated artifact removal as preprocessing refines neonatal seizure detection. Clinical Neurophysiology 2011;122(12):2345 – 2354.

De Vos, M., Rios, S., Vanderperren, K., Vanrumste, B., Alario, F.X., Van Huffel, S., Burle, B.. Removal of muscle artifacts from eeg recordings of spoken language production. Neuroinformatics 2010;8:135–50.

Deburchgraeve, W., Cherian, P., De Vos, M., Swarte, R., Blok, J., Visser, G., Govaert, P., Van Huffel, S.. Automated neonatal seizure detection mimicking a human observer reading eeg. Clin Neurophysiol 2008;119(11):2447 – 2454.

Engel, J.. Surgery for seizures. N Engl J Med 1996;334(10):647–53.

Gardner, A.B., Krieger, A.M., Vachtsevanos, G., Litt, B.. One-class novelty detection for seizure analysis from intracranial eeg. J Mach Learn Res 2006;7:1025–1044.

Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K., Stanley, H.E.. Physiobank, physiotoolkit, and physionet : Components of a new research resource for complex physiologic signals. Circulation 2000;101(23):215–20.

Grant, M., Boyd, S.. CVX: Matlab software for disciplined convex programming, version 1.21. `http://cvxr.com/cvx`; 2011.

Greene, B., Marnane, W., Lightbody, G., Reilly, R., Boylan, G.. Classifier models and architectures for eeg-based neonatal seizure detection. Physiol Meas 2008;29(10):1157.

He, X., Cai, D., Niyogi, P.. Tensor subspace analysis. In: Weiss, Y., Schölkopf, B., Platt, J., editors. Advances in Neural Information Processing Systems (NIPS). MIT Press; 2006. p. 499–506.

Lukas, L., Devos, A., Suykens, J.A.K., Vanhamme, L., Van Huffel, S., Tate, A., Majós, C., Arús, C.. The Use of LS-SVM in the Classification of Brain Tumors Based on Magnetic Resonance Spectroscopy Signals. In: Proceedings of the European Symposium on Artificial Neural Networks, (ESANN). Bruges; 2002. .

Meier, R., Dittrich, H., Schulze-Bonhage, A., Aertsen, A.. 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 2008;25(3):119–131.

Minasyan, G.R., Chatten, J.B., Chatten, M.J., Harner, R.N.. Patient-specific early seizure detection from scalp electroencephalogram. J of Clin Neurophysiol 2010;27(3):163–78.

Nandan, M., Talathi, S.S., Myers, S., Ditto, W.L., Khargonekar, P.P., Carney, P.R.. Support vector machines for seizure detection in an animal model of chronic epilepsy. J of Neural Eng 2010;7(3):036001.

Platt, J.. Probabilistic outputs for svm and comparison to regularized likelihood methods. Adv large Margin Classifiers 1999;:61–74.

Polychronaki, G., Ktonas, P., Gatzonis, S., Siatouni, A., Asvestas, P., Tsekou, H., Sakas, D., Nikita, K.. Comparison of fractal dimension estimation algorithms for epileptic seizure onset detection. J of Neural Eng 2010;7(4):046007.

Qu, H., Gotman, J.. A patient-specific algorithm for the detection of seizure onset in long-term eeg monitoring: possible use as a warning device. IEEE Trans Biomed Eng 1997;44(2):115 –122.

Recht, B., Fazel, M., Parrilo, P.. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Review 2007;52 (3):471–501.

Saab, M., Gotman, J.. A system to detect the onset of epileptic seizures in scalp eeg. Clin Neurophysiol 2005;116(2):427 – 442.

Shoeb, A., Guttag, J.. Application of machine learning to epileptic seizure detection. In: Fürnkranz, J., Joachims, T., editors. Proceedings of the 27th International Conference on Machine Learning (ICML). Haifa, Israel: Omnipress; 2010. p. 975–982.

Signoretto, M.. Kernels and Tensors for Structured Data Modelling. Ph.D. thesis; Katholieke Universiteit Leuven; 2011.

Signoretto, M., De Lathauwer, L., Suykens, J.A.K.. A kernel-based framework to tensorial data analysis. Neural Networks 2011;24(8):861 – 874.

Stern, J., Engel, J.. An Atlas of EEG Patterns. Lippincott Williams & Wilkins, 2004.

Suykens, J.A.K., Van Gestel, T., De Brabanter, J., De Moor, B., Vandewalle, J.. Least Squares Support Vector Machines. World Scientific, Singapore, 2002.

Suykens, J.A.K., Vandewalle, J.. Least squares support vector machine classifiers. Neural Process Lett 1999;9(3):293–300.

Tao, D., Li, X., Wu, X., Hu, W., Maybank, S.J.. Supervised tensor learning. Knowl Inf Syst 2007;13:1–42.

Temko, A., Thomas, E., Marnane, W., Lightbody, G., Boylan, G.. Eeg-based neonatal seizure detection with support vector machines. Clin Neurophysiol 2011;122(3):464–73.

Thomas, E., Temko, A., Lightbody, G., Marnane, W., Boylan, G.B.. A comparison of generative and discriminative approaches in automated neonatal seizure detection. In: Intelligent Signal Processing, 2009. WISP 2009. IEEE International Symposium on. 2009. p. 181–186.

Tito, M., Cabrerizo, M., Ayala, M., Jayakar, P., Adjouadi, M.. Seizure detection: An assessment of time- and frequency-based features in a unified two-dimensional decisional space using nonlinear decision functions. Journal of Clinical Neurophysiology 2009;26(6):381–391.

Tomioka, R., Aihara, K.. Classifying matrices with a spectral regularization. In: In proceeding of: Machine Learning, Proceedings of the Twenty-Fourth International Conference (ICML 2007). 2007. p. 895–902.

Varsavsky, A., Mareels, I., Cook, M.. Epileptic Seizures and the EEG: Measurement, Model, Detection and Prediction. CRC Press, 2011.

Xavier de Souza, S., Suykens, J.A.K., Vandewalle, J., Bollé, D.. Coupled simulated annealing. IEEE T Syst Man Cyb 2010;40(2):320–335.

Zandi, A.S., Javidan, M., Dumont, G.A., Tafreshi, R.. Automated real-time epileptic seizure detection in scalp eeg recordings using an algorithm based on wavelet packet transform. IEEE Trans Biomed Eng 2010;57(7):1639 –51.

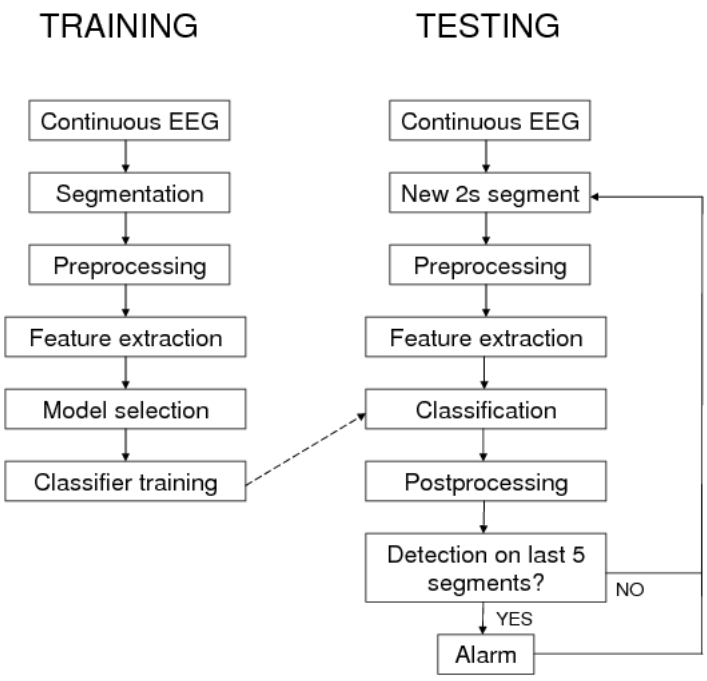


Figure 1: Operational structure of the classifier training and the testing of the seizure detector.

Table 1: Extracted Features

Time domain features	1-3. number of zero crossings (#0), maxima (#max) and minima (#min); 4. skewness (skew); 5. kurtosis (kurt); 6. root mean square amplitude (rmsa)
Frequency domain features	7. total power (TP); 8. peak frequency (PF); 9-16. mean and normalized power in frequency bands: delta: 1-3 Hz (D, nD) theta: 4-8 Hz (T, nT) alpha: 9-13 Hz (A, nA) beta: 14-20 Hz (B, nB)

Table 2: Event-based performance evaluation measures computed on the test data comparing the performance of the three detection approaches given different training sets. Sensitivities, false detection rates, detection delays and quality values are reported in terms of median and mean (in brackets) over all 22 patients. The number of seizures used for training is indicated in the first column of the table.

		NNL	EI	LI
1	dS (%)	75 (63)	93 (76)	75 (57)
	R_{fa} (/h)	0.06 (0.26)	0.31 (0.71)	0.09 (0.45)
	D (s)	7.6(10.4)	5.5 (8.1)	9.7 (12.1)
2	dS (%)	100 (74)	100 (76)	60 (52)
	R_{fa} (/h)	0.11 (0.29)	0.31 (0.71)	0.10 (0.60)
	D (s)	7.8 (9.8)	5.5 (8.4)	5.0 (8.2)
3	dS (%)	100 (78)	100 (81)	90 (64)
	R_{fa} (/h)	0.12 (0.40)	0.50 (0.76)	0.14(0.61)
	D (s)	10.5 (10.1)	7.0 (8.1)	6.8 (9.1)
4	dS (%)	100 (76)	100 (82)	93 (64)
	R_{fa} (/h)	0.11 (0.33)	0.47 (0.82)	0.43 (0.77)
	D (s)	9.8 (10.7)	7.0(8.4)	7.9 (9.8)
5	dS (%)	100 (80)	100 (83)	100 (67)
	R_{fa} (/h)	0.14 (0.41)	0.50 (0.88)	0.23(0.76)
	D (s)	9.0 (10.0)	7.0 (8.5)	8.5 (11.2)

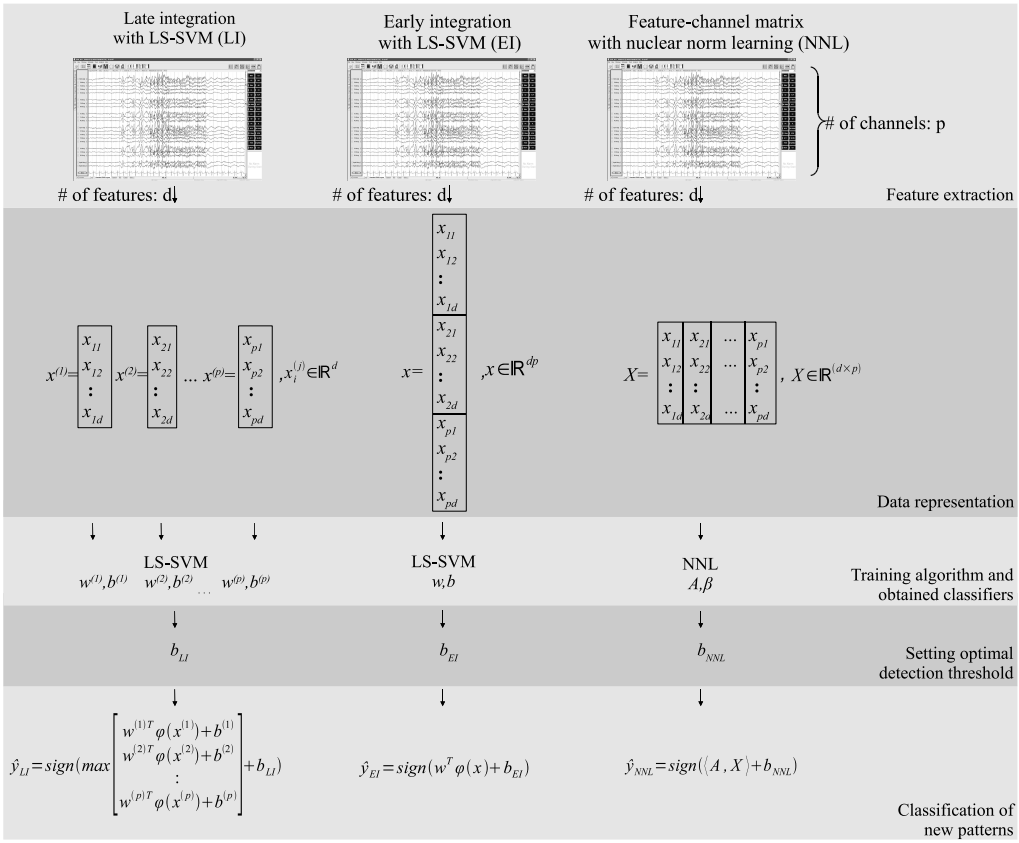


Figure 2: Diagram depicting the various steps of the different seizure detection approaches

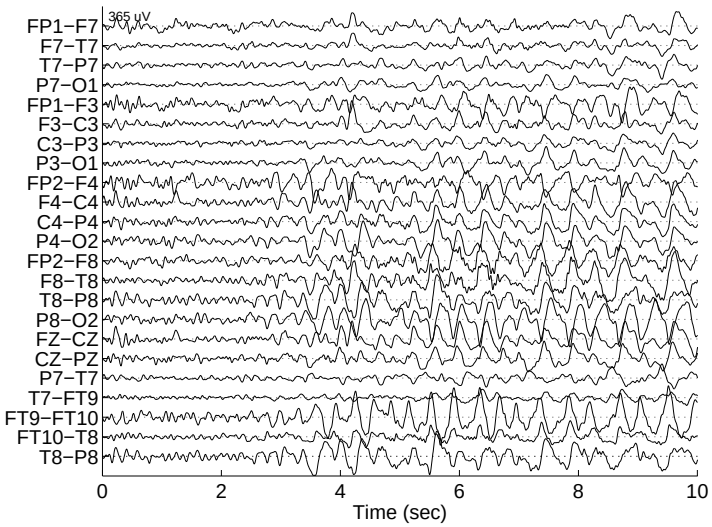
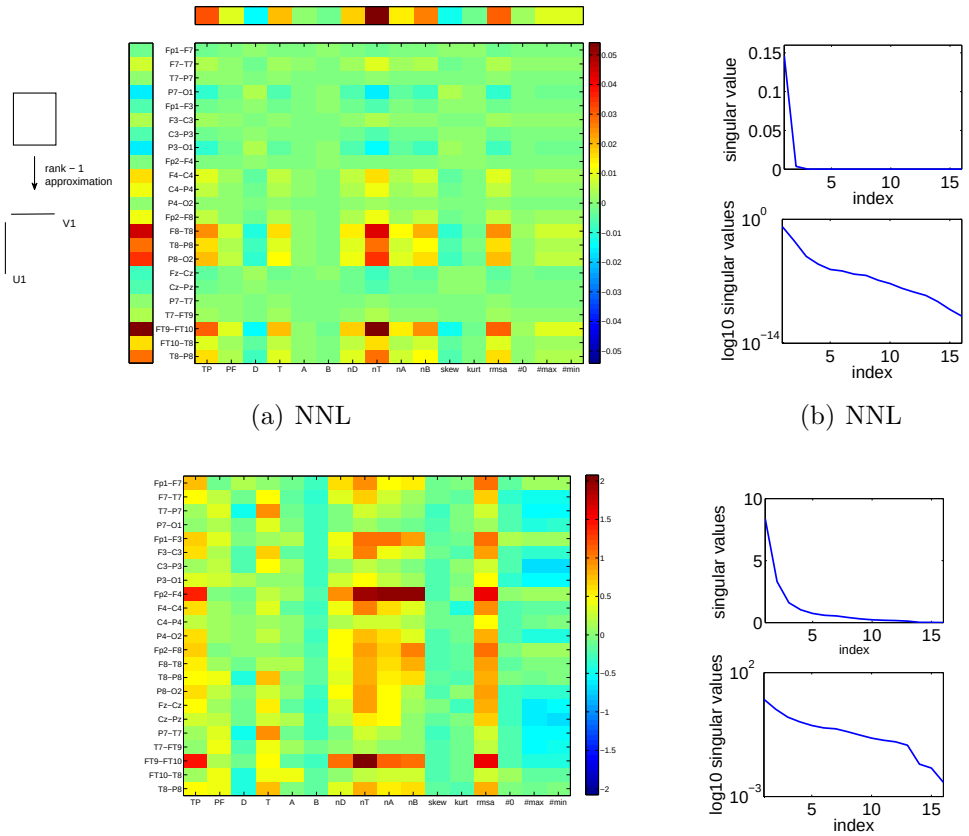


Figure 3: First training seizure of Patient 1.



(a) NNL (b) NNL

(c) EI (d) EI

Figure 4: Classifier matrices (4(a),4(c)) and their singular values (4(b), 4(d)) obtained by training with the first seizure of patient 1, using the NNL and EI approaches, respectively. 4(a) also shows the left and right singular vectors. It is visually observed that outer product of the singular vectors closely approximates the classifier matrix. 4(b) shows fast decaying singular values: the first singular value carries 97.6% of the total energy, indicating an approximately rank-1 structure. The values of the left and right singular vectors represent the relative discriminative power of the channels and features, respectively. Note that the highest channel entries (electrodes over the right temporal and parietal area), and highest feature entry (normalized power in the theta band) characterize well the seizure pattern on 3. In comparison, the classifier matrix obtained by EI (4(c)) is less structured. Its singular values decay slower, the first singular value carries 47.6% of the total energy.

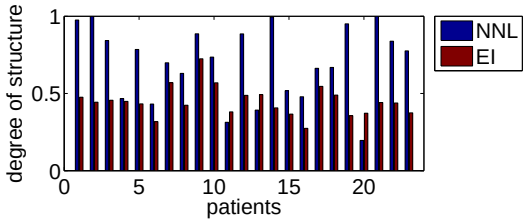
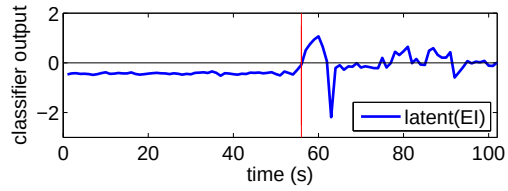
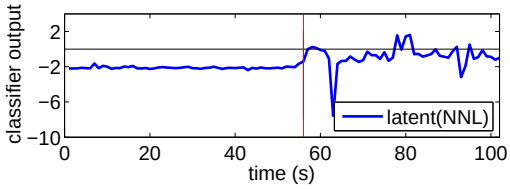


Figure 5: The bar plot compares the degree of structure in the classifiers obtained by NNL and EI. The degree of structure is expressed as percentage of energy carried by the first singular value of the classifier matrix. In 20 out of 23 cases the NNL approach results in more structured classifier.



(a)



(b)

Figure 6: Classifier outputs obtained by the EI approach 6(a) and the NNL approach 6(b) for the first test seizure of Patient 8. The vertical line indicates the onset of the seizure according to the labeling of an expert. This is one of the few examples where a seizure was missed by the NNL but detected by the EI approach. The output time courses of both classifiers follow a similar pattern, indicating comparable discriminative power. The seizure is missed by NNL due to the relatively high detection threshold.

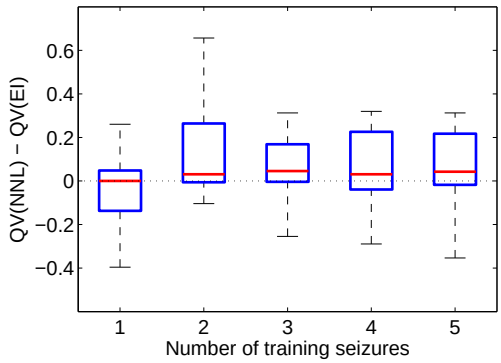


Figure 7: Boxplot depicting the patient-by-patient differences of NNL and EI-LSSVM performance in terms of quality value, evaluated based on the test data, given different number of training seizures

Table 3: P-values obtained by Wilcoxon signed rank tests comparing the detection approaches. The number of training seizure used to train train each method is shown after the name of the approach in the different rows and columns. Values corresponding to significant differences at level $\alpha = 0.05$ appear in bold.

NNL vs EI					
	EI 1	EI 2	EI 3	EI 4	EI 5
NNL 1	0.6389	0.6482	0.4842	0.5034	0.6051
NNL 2		**0.0459**	**0.0335**	0.0619	0.1997
NNL 3			**0.0392**	**0.0170**	0.1485
NNL 4				0.1886	0.1311
NNL 5					0.0918
NNL vs LI					
	LI 1	LI 2	LI 3	LI 4	LI 5
NNL 1	0.1396	0.1011	0.3221	**0.0221**	0.0777
NNL 2		**0.0061**	**0.0033**	**0.0014**	**0.0056**
NNL 3			**0.0002**	**0.0002**	**0.0010**
NNL 4				**0.0002**	**0.0004**
NNL 5					**0.0002**
EI vs LI					
	LI 1	LI 2	LI 3	LI 4	LI 5
EI 1	0.0630	0.0534	0.0617	**0.0156**	**0.0244**
EI 2		**0.0108**	**0.0045**	**0.0067**	**0.0108**
EI 3			0.0457	**0.0355**	**0.0420**
EI 4				**0.0078**	**0.0208**
EI 5					**0.0071**