



This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.



**CC creative commons**  
COMMONS DEED

**Attribution-NonCommercial-NoDerivs 2.5**

**You are free:**

- to copy, distribute, display, and perform the work

**Under the following conditions:**

**BY:** **Attribution.** You must attribute the work in the manner specified by the author or licensor.

**Noncommercial.** You may not use this work for commercial purposes.

**No Derivative Works.** You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

**Your fair use and other rights are in no way affected by the above.**

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

For the full text of this licence, please go to:  
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

# SUPPORT VECTOR MACHINES FOR SEIZURE DETECTION

*Bruno González-Vellón, Saeid Sanei, and Jonathon A. Chambers*

Centre for Digital Signal Processing Research, King's College London, United Kingdom

## ABSTRACT

Development of a robust technique for automatic detection of the epileptic seizures is an important goal in clinical neurosciences. In this paper support vector machines (SVM) has been used for this purpose. The system presented detects and uses three features of the electroencephalogram (EEG), namely, energy, decay (damping) of the dominant frequency and cyclostationarity of the signals. The different types of epileptic seizures show some common characteristics in the feature space that can be exploited to distinguish them from normal activity in the brain or the nonepileptic abnormalities. The use of SVMs achieves high sensitivity and at the same time shows an improvement in terms of computational speed in comparison with other traditional systems.

## 1. INTRODUCTION

Analysis of the EEG is the primary method for diagnosis of epilepsy. Based on the EEG patterns during epileptic seizures, physicians can determine the type of epileptic syndrome. Long term monitoring is the common procedure to register the occurrence of epileptic events, which may last from seconds to hours.

Traditionally, seizure detection was done at hospitals by means of continuous observation and possibly the use of an alarm button activated by the patient. Otherwise, the patient reports it. Nevertheless, many epileptic seizures imply loss of consciousness, confusion or even loss of memory, making patient's registrations inaccurate. Moreover, some seizures do not present clinical symptoms and can be unnoticed even by the patient himself.

Automatic detection of epilepsy has been investigated for the past few years. Murro [1] proposed a method based on spectral analysis in which he used the dominant frequency, the relative amplitude and the power of the spectrum around the dominant frequency divided by the total power as features. Harding exploited the increase of spikes in the EEG in his proposal [2]. The use of neural networks has been common in the field of automatic detection since Webber [3] used a 31-30-8 network combining statistical features such as mean, variance and

skewness with morphologic features like amplitude, slope or duration of waves. Neural networks were also used by Pradhan [4]. He used raw EEG as the input. Weng et al. used those features proposed by Gotman as the inputs to their NN [5]. In 1996, Gabor et al. brought the dynamics of the brain insight with the help of wavelets [6]. In a recent paper, Qu and Gotman used a nearest neighbour classifier with features from both time and frequency domain extracted from overlapping EEG epochs of 2.56 sec length [7]. Finally, a popular method called Matching pursuit (MP) algorithm was introduced by Durka et al based on time-frequency decomposition, using Gabor functions called atoms [8].

Selection of the proper features plays a key role in design of the system. For classification, the most desirable features will be those that are easily separable (so errors in classification are less likely) and significant (so few features are required and computational costs are diminished). Thus, designing a feature extractor implies a deep knowledge of the problem domain, although fortunately no knowledge of the underlying laws that govern the brain is needed to recognize its manifestations. It is shown that only three features of the EEG are enough to detect epileptic seizures with sufficient accuracy. These are energy, frequency decay as the seizure evolves (i.e. the so-called damping factor) and the cyclostationarity.

The adequate properties of SVMs make them desirable against other alternative methods such as neural networks or decision trees. Effectively, SVMs involve optimisation of a cost function that is convex, so no false local minima can get the algorithm stuck during the learning process. Moreover, SVMs are simple to use and only require a kernel and a few parameters to be chosen by the designer. Besides, the results obtained by SVMs have the advantage of being stable and reproducible in the sense that for a given data set, several executions with the same kernel and the same parameters will produce the same result (unlike for neural networks).

## 2. FEATURES OF THE EEG

To analyse the EEG a window is slid over the EEG signal and the features of each frame of EEG are extracted. These windows overlap in order not to miss any possible

events happening at the end of a frame and prolonging to the next one. Moreover, overlapping windows offer better time resolution and can produce shorter delays in the detection.

Considering the frequency of the events involved, a frame length of 2 sec has been proved to be appropriate, resulting in a time resolution of 1 sec (since frames are half overlapped) and a frequency resolution equal to  $F_s/2M$ , where  $F_s$  is the sampling frequency of the EEG and  $M$  the number of points, is considered in the calculation of the spectrum. In our case,  $M=128$ .  $F_s$  is typically 200 Hz, yielding a frequency resolution of 0.78 Hz approximately. Obviously, a trade-off arises here between time precision and spectral resolution.

For majority of cases the amplitude of the signal increases during seizure, so, the first feature considered here is the average energy of the frame,  $E(t)$ , defined as:

$$E(t) = \frac{1}{L} \sum_{n=1}^L (x^t[n])^2 \quad (1)$$

where  $x^t[n]$  is the frame of EEG (containing  $L$  samples) centred at time  $t$ . Ictal activity is often well distinguished by high values of energy compared to background activity.

The decay (damping) in frequency during seizure is an interesting factor. It reflects variations in the dominant frequency and quantifies its decrease through time. For every frame the dominant frequency is calculated as:

$$f_d(t) = \arg \max_f (|X^t(f)|) \quad (2)$$

where  $X^t(f)$  is the estimated spectrum based on autoregressive modelling (AR) of the signals i.e.

$$X^t(f) = X^t(z) \Big|_{z=j2\pi f} = \frac{B}{1 - \sum_{k=1}^p a_k z^{-k}} \Big|_{z=j2\pi f} \quad (3)$$

where  $B$  is a constant set to  $B=1$ ,  $p$  is the prediction order, and  $a_k$  are the model prediction coefficients.  $a_k$  are calculated from the signal using Durbin algorithm. Then a first order polynomial fitting of  $f_d(t)$  is performed over a sliding window for several frames. The damping factor is represented by the slope of the polynomial.

Analytically, let the sliding window contain the frames at  $t_1, t_2, \dots, t_k$ , and  $f_d(t_1), f_d(t_2), \dots, f_d(t_k)$  be the dominant frequencies of those frames, respectively. The linear interpolation consists of fitting a linear polynomial so that we can express  $f_d(t) \approx at + b$ , with  $t = t_1, t_2, \dots, t_k$ .

The damping factor is thus represented by the coefficient  $a$ . If  $a < 0$  the pattern represents an onset of seizure. The size of the sliding window must be large enough for interpolation to be significant and small enough not to ignore short seizures and, in case of on-line detection, not to introduce a considerable delay;  $3f_s \leq L \leq 5f_s$  is usually adequate ( $f_s$  is the sampling frequency).

The third feature is a measure of cyclostationarity of the signal defined by the following indicator [9]:

$$I(t) = |c_2^0(0)|^{-4} \sum_{\alpha \neq 0} |P^\alpha|^2 \quad (4)$$

where  $P^\alpha = c_4^\alpha(0,0,0)$  represents the Fourier coefficient of the cyclic cumulant of fourth order at zero lag and can be estimated as follows:

$$\hat{c}_4^\alpha(0,0,0) = \frac{1}{N} \sum_{n=0}^{N-1} x_c^4[n] e^{\frac{-j2\pi n\alpha}{N}} - 3 \sum_{\beta} c_2^{\alpha-\beta}(0) c_2^\beta(0) \quad (5)$$

where  $x_c[n] = x^t[n] - \hat{M}_1[n]$ , with a synchronous average of  $\hat{M}_1[n]$  and

$$\hat{c}_2^\alpha(0) = \frac{1}{N} \sum_{n=0}^{N-1} x_c^2[n] \exp(-j2\pi n\alpha/N). \quad (6)$$

The indicator in equation (4) measures, for a frame centred at time  $t$ , how spread the energy is over the range of frequencies before seizure onset, the EEG has a chaotic behaviour and no frequency appears to control its trace. During seizure, the EEG becomes more rhythmic and therefore the spectrum appears dominated by a peak contrasting with the flatness of the remaining components. This results in large values of the indicator of cyclostationarity. The fact that the dominant frequency may vary (which makes necessary the use of the damping factor) does not affect the aspect ratio of the spectrum and therefore its cyclostationarity. Involvement of other features, i.e. decay in frequency and the signal energy, will be also effective due to the certain properties of the EEGs during the onset of seizures.

### 3. SYSTEM DESCRIPTION

The framework on which our system is based on is stated as follows. A set of known EEG recordings (for both seizures and normal states) is given. For each recording, a window is slid over time in overlapping steps. For each frame of EEG signal, three features are extracted. When all the frames within a recording interval have been processed, they are labelled accordingly (-1 for epileptic seizures, +1 otherwise). When all the recordings in the set

are processed, the supervised SVM algorithm is applied to the set of features to find the decision boundary that better separates the two classes or categories. At this point, the machine is said to be trained, i.e. it has "learnt" the common features that characterize the epileptic seizures and is prepared to identify them in the new recordings. The target values, i.e. -1 and +1, are presented to the system during the training process.

For the frequency analysis of the frames, the power spectral density must be estimated. This is done by means of an autoregressive (AR) model to overcome blocking effect in DFT and the restrictions induced by the nonstationarity of the signals. Moreover, AR filters use linear prediction of the output based on previous inputs, reinforcing the signal against the effect of noise, which makes the resulting spectra smoother and less affected by noise. The order of the AR filter represents a compromise between accuracy and computational speed. The prediction order can be estimated. For this application, values between 10 to 12 have been proved to be appropriate for our purposes. The convenience for AR models and its considerations have been discussed in more detail by other authors [10].

The feature extractor transforms portions of EEG into a space of higher dimension (i.e. the feature space,  $\mathfrak{R}^3$ ) where the significant information has been extracted to make the differences more apparent. Each transformed portion can be viewed as a point in the feature space. These points represent the input to the classifier.

Classification of the points is performed by means of SVMs [11]. This is a geometrically intuitive and elegant algorithm based on beautifully simple ideas from statistical learning theory that permit to construct complex models. Let every frame of EEG be described as a point  $x^t$  in the feature space, i.e.  $x^t = (x^t_I, x^t_D, x^t_E)$ , where  $x^t_I$ ,  $x^t_D$  and  $x^t_E$  are respectively the values of the indicator of cyclostationarity, damping factor and energy for the frame centred at time  $t$ . Thus, a recording of EEG comprising several frames is defined by a set of points  $\{x^t_1, x^t_2, \dots, x^t_N\}$ . During training the system, the set of targets  $\{y_1, y_2, \dots, y_N\}$  is also known (since it is supervised training), with  $y_i$  indicating if  $x^t_i$  corresponds to a seizure ( $y_i = -1$ ) or not ( $y_i = +1$ ). The purpose of the classifier is to find a boundary that separates points with different values of  $y_i$ . SVMs construct linear discriminant functions based on the criterion of maximisation of the separating margin between the two classes. Nevertheless, it is usual that the points are not linearly separable, in which case a linear function does not classify well. This is solved by the introduction of kernels and a relaxation of the margins so that some points are accepted to invade the opposite margin. The use of kernels

allows mapping the points into a space of higher dimensionality where they are indeed linearly separable. Any linear discriminant function traced in this space will yield a nonlinear boundary in the original feature space when the mapping is inverted.

SVMs minimize the overlap between the classes as:

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j \alpha_i \alpha_j K(x_i, x_j) - \sum_{i=1}^N \alpha_i \quad (7)$$

subject to the constraints

$$\sum_{i=1}^l y_i \alpha_i = 0 \quad (8)$$

$$\text{and } 0 \leq \alpha_i \leq C \quad i = 1, \dots, N. \quad (9)$$

The coefficients  $\alpha_i$  are the Lagrangian multipliers,  $K(x_i, x_j)$  is the kernel function and  $C$  is a parameter that determines how relaxed the margins are. In our case, a radial basis function (RBF) was chosen as a kernel, that is:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right). \quad (10)$$

Using RBF enables classification of a non-linear set of data with a quasi-Gaussian distribution, which perfectly matches the EEG signals. One of the advantages of using SVMs is that there exist fast computational algorithms to solve the quadratic problem stated by equation (6). The solution of the problem yields what are known as the support vectors, i.e. a subset of the data points that contains points relevant and closest to the decision boundary. Training the classifier means determining the support vectors for the training EEGs by solving the aforementioned quadratic equations. They provide the optimally encompassed regions corresponding to the two separated classes.

The output of the classifier consists of a time series of the values -1 or +1, depending on whether the corresponding frame is classified as a seizure or not, respectively. Post-processing is applied to these signals based on the knowledge of the problem domain. This avoids possible fluctuations in our detector due to the misclassified points and exploits the correlation between channels to obtain more information such as the localization of the seizure or its propagation.

#### 4. RESULTS

The performance of the system was tested with the choice of an RBF as a kernel. Thus, only two parameters must be

Table 1. Confusion table

		Output	
		Seizure	No Seizure
Seizure	Tonic-Clonic	100%	0
	Complex Partial	100%	0
No Seizure	Hyperventilation	18.75%	81.25%
	Nonepileptic Abnormalities	30%	70%
	Normal Activity	0	100%

chosen, namely,  $C$  and the variance of the Gaussian in equation (10). These parameters can be adjusted empirically but, in general,  $\sigma=1$  is a good option since the inputs to the classifier are previously standardized (that is, the mean removed and the variance normalized to unity) and even  $C=1$  provides enough relaxation of the separating margins, avoiding the classifier to excessively specialize in the training set and consequently perform poorly in the general cases (this effect is known as overfitting). However, we set  $C$  to 100 as a default value. The results obtained proved nonetheless to be stable for a wide range of values for both parameters.

The system was trained using the EEGs of a complex partial seizure. A set of 40 scalp EEGs from different subjects was used to test the performance. All the signals were 10 sec long and contained between 3 and 5 sec of preictal activity. 10 of the EEGs corresponded to epileptic seizures (7 tonic-clonic and 3 complex partial). The remaining EEGs contained no seizure (16 with hyperventilation, 10 with nonepileptic abnormalities and 4 normal). The results are summarized in table 1.

### 5. CONCLUSIONS

The proposed system has shown a sensitivity of 100%, i.e. all seizures were detected. With no abnormalities no false positives were produced (as shows the last row of table 1). However, the specificity of the system, i.e. its ability not to make erroneous detections, gets worse when some abnormal (although nonepileptic) activity is present, yielding a total specificity of 80%.

These figures would surely be improved if all the EEGs were corresponded to the same patient or they were intracranial EEGs, for which the features would present the differences more clearly. Moreover, a cascade of SVM classifiers could be used instead of only one, with each of them specialized in a particular type of seizure. Thus, sensitivity would be kept high without significantly

reducing speed, since the computational cost of performing additional classifications with SVMs is low compared to that incurred in the extraction of the features.

### 6. REFERENCES

- [1] Murro AM, King DW, Smith JR, Gallagher BB, Flanigin HF, Meador KJ. Computerized seizure detection of complex partial seizures. *Electroencephalography and Clinical Neurophysiology* 1991;79:330-333.
- [2] Harding GW. An automated seizure monitoring system for patients with indwelling recording electrodes. *Electroencephalography and Clinical Neurophysiology* 1993; 86(6): 428-437.
- [3] Webber WRS et al, An approach to seizure detection using an artificial neural network. *Electroenceph. and Clinical Neurophysiology* 1996; 98:250-272.
- [4] Pradhan N, et al. Detection of seizure activity in EEG by an artificial neural network: a preliminary study. *Computers and Biomedical Research* 1996;29(4):303-313.
- [5] Weng W, Khorasani K. An adaptive structure neural network with application to EEG automatic seizure detection. *Neural Networks* 1996;9(7):1223-1240.
- [6] Gabor AJ et al. Automated seizure detection using a self-organizing neural network. *Electroencephalography and Clinical Neurophysiology* 1996;99:257-266.
- [7] 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 Transactions on Biomedical Engineering* 1997;44(2):115-122.
- [8] Durka P. J, D. Ircha, and K. J. Blinowska, "Stochastic Time-Frequency Dictionaries for Matching Pursuit", *IEEE Trans. on Signal Processing*, 2001, 49(3), pp. 507-510,
- [9] A. Raad, et. al. "Indicators of Cyclostationarity: Proposal, Statistical Evaluation and Application to Diagnostics," *IEEE, ICASSP* 2003, pp. VI-757.
- [10] A. Schlögl, and G. Pfurtscheller, "Considerations on Adaptive Autoregressive Modelling in EEG Analysis," *Proc.Int. symposium on communication systems and digital signal processing*, Sheffield, UK, pp. 367-370, 1998.
- [11] C. Cortes, and V.N. Vapnik, Support Vector Networks, *Machine Learning*, vol. 20, pp. 273-297, 1995.