# Real time epileptic seizure onset detection using approximate extreme points SVM

**Manu Nandan** MNANDAN@UFL.EDU
Department of Computer and Information Science and Engineering,
University of Florida, Gainesville, FL 32611, USA

**Pramod P. Khargonekar** PPK@ECE.UFL.EDU
Department of Electrical and Computer Engineering,
University of Florida, Gainesville, FL 32611, USA

**Sachin S. Talathi** TALATHI@UFL.EDU
Department of Pediatrics, Division of Neurology,
Department of Biomedical Engineering,
Department of Neuroscience,
University of Florida, Gainesville, FL 32611, USA

## Abstract

We present a novel algorithm for real time epileptic seizure onset detection, based on approximate extreme point support vector machine (Nandan et al., 2013). We demonstrate the utility of our method to this application, through theoretical and experimental results. We evaluated our approach using the large CHB-MIT database (846 hrs) of scalp EEG recordings of pediatric patients. Our method detected 93% of the seizures with a mean detection latency of 7.5 seconds and a false positive rate of 0.9 per 24 hours. We compared our method with an SVM based seizure detector (Shoeb & Guttag, 2010), and found the method to be on average 5 times faster in training and 2.5 times faster in classification. Our results indicate the suitability of our approach for use in closed loop seizure controllers.

## 1. Introduction

Epilepsy is a neurological disorder affecting roughly 1% of the world population. Current diagnosis protocol for epilepsy involves in-clinic video and EEG monitoring of patients for hours or days followed by an analysis of the recordings by a trained physician. Since this is a time consuming and labor intensive process, automated seizure detection tools have been researched for the last two decades. Naturally, many of these studies apply machine learning algorithms such as support vector machines (Gardner et al., 2006; Nandan et al., 2010; Shoeb & Guttag, 2010) to seizure detection.

The spontaneous nature of epileptic seizures is a serious impediment to the day-to-day activities of patients suffering from refractory epilepsy, and severely degrades their quality of life. A large portion of epileptic patients do not respond to medication and are not expected to benefit from surgery (Kahane & Depaulis, 2010). A modern approach to suppress seizures in such patients is to use neurostimulation methods such as deep brain stimulation (Velasco et al., 1995) and vagus nerve stimulation (Groves & Brown, 2005). These neurostimulation methods control seizures by continuously subjecting parts of the brain to electrical or optical stimulation (Zhang et al., 2009). However, there are concerns about the long-term side effects of the continuous brain stimulation used in these methods (Stacey & Litt, 2008; Kahane & Depaulis, 2010).

Recently, research studies (Verma et al., 2010; Liang et al., 2011) have addressed this problem by developing closed-loop seizure control methods (CLSCM) that apply neurostimulation only when there is an impending seizure. Such closed-loop systems use seizure detection techniques to detect the onset of seizures. The desired qualities of a seizure detection technique suitable for

*Proceedings of the 30$^{th}$ International Conference on Machine Learning*, Atlanta, Georgia, USA, 2013. JMLR: W&CP volume 28. 

use in CLSCM are:

- High accuracy and low latency in seizure onset detection
- Low computational complexity, enabling its use with low power hardware

SVMs have been found to result in highly accurate seizure detection (Shoeb & Guttag, 2010), with reasonably good detection latency. However, they suffer from high computational complexity in training and testing. High computational complexity while testing makes them unsuitable for use with low-power devices. High computational complexity while training makes them difficult to use by increasing the time gap from data collection to device deployment. This problem is compounded when parameter tuning requires grid search with cross validation. Shoeb et al. (2009) applied SVMs to CLSCM by post processing the SVM solution using reduced set methods to reduce the classification time. A problem with reduced set methods (Schölkopf et al., 1998) is that they generally do not have theoretical bounds on the approximation error.

In this paper we propose a new seizure detection method based on SVMs that addresses the key requirements of CLSCM. To reduce the training time of SVMs we used the recently proposed approximate extreme point support vector machine (AESVM) (Nandan et al., 2013), *and compared it with the standard SVM*. It has been shown (Nandan et al., 2013) that AESVM has much better performance in terms of training and classification time. AESVM achieves fast training and classification, by using only a subset of the training dataset called the representative set. The representative set is computed using a log-linear algorithm DeriveRS, proposed in Nandan et al. (2013). Fast AESVM classification is a direct consequence of the typically small size of the representative set compared to the training dataset. DrsPost is similar to reduced set methods, but has theoretical properties indicating its closeness to the SVM solution. Another difference between our method and the reduced set methods is that our method also reduces the training time.

The performance of our method was evaluated on the CHB-MIT database of scalp EEG recordings of pediatric patients. This database was used in Shoeb & Guttag (2010) to evaluate an SVM based seizure detector. They considered seizure detection as a binary classification task and presented the need for patient specific seizure detection, due to the variations in seizure characteristics between patients. We have used their method as a baseline, to illustrate the advantages of our method for application in CLSCM. The proposed method is on average faster in classification by 2.5 times and faster in training by 5 times, than an SVM based seizure detector, without any degradation in performance.

This paper is divided into the following sections. Section 2 gives a brief overview of DeriveRS and AESVM, and describes DrsPost. Section 3 describes the datasets, data pre-processing methods, seizure detection features and performance metrics. We compared our method with the method in Shoeb & Guttag (2010) and the results of the comparison are given in Section 4. Parameter tuning of SVM and AESVM can be problematic, especially for people not familiar with it. We present a novel method for parameter tuning of AESVM in Section 4, that is easy for clinicians to use. Finally, in Section 5 we present a few concluding remarks.

## 2. Methods

In this section we first briefly discuss properties of the representative set, followed by a brief review of the DeriveRS algorithm and the AESVM problem formulation. Next we present our algorithm, DrsPost, and describe its properties.

### 2.1. Properties of the representative set

Consider a two class training dataset of $N$ data vectors, $\mathbf{X} = \{\mathbf{x}_i : \ \mathbf{x}_i \in \mathbb{R}^D, i = 1, 2, ..., N\}$, and the corresponding target labels $\mathbf{Y} = \{y_i : \ y_i \in [-1, 1], i = 1, 2, ..., N\}$. Consider the set of transformed data vectors $\mathbf{Z} = \{\mathbf{z}_i : \mathbf{z}_i = \phi(\mathbf{x}_i), \forall \mathbf{x}_i \in \mathbf{X}\}$. The representative set $\mathbf{Z}^*$ of $\mathbf{Z}$, is any subset that satisfies the following properties:

$$\mathbf{Z}^* \subseteq \mathbf{Z} \tag{1}$$

$$\|\mathbf{z}_i - \sum_{\mathbf{z}_t \in \mathbf{Z}^*, y_t = y_i} \gamma_t^i \mathbf{z}_t\|^2 \leq \epsilon \tag{2}$$

$$\text{s.t. } 0 \leq \gamma_t^i \leq 1, \text{ and } \sum_{\mathbf{z}_t \in \mathbf{Z}^*, y_t = y_i} \gamma_t^i = 1$$

where $\epsilon$ is an approximation error bound. Any vector in $\mathbf{Z}$ can be approximately represented as a convex combination of vectors in its representative set $\mathbf{Z}^*$.The weights $\gamma_t^i$ are used to define $\beta_t$ as:

$$\beta_t = \sum_{i=1}^{N} \gamma_t^i \tag{3}$$

For ease of notation, we refer to the set $\mathbf{X}^* := \{\mathbf{x}_t : \mathbf{z}_t \in \mathbf{Z}^*\}$ as the representative set of $\mathbf{X}$ in the remainder of this paper.

### 2.2. Review of DeriveRS and AESVM

DeriveRS is a divide and conquer algorithm that computes the representative set of $\mathbf{X}$ in kernel space. The complexity of DeriveRS scales log-linearly with $N$, as $O(N(\log_2 \frac{N}{P} + \frac{P}{V} + \log_2 V))$. The parameters $P$ and $V$ are integers, with $P \geq V$. In DeriveRS, operations of increasing time complexity are executed on subsets of decreasing size. The algorithm proceeds by first dividing $\mathbf{X}$ into subsets of size $P$ and then into subsets of size $V$ each, using two different segregation schemes. The subsets are chosen such that the data vectors within the subsets are closer to each other, than to data vectors in other subsets. After data segregation, the resulting subsets of size $V$ are passed to a routine DeriveAE, using which $\mathbf{X}^*$ and $\beta_t$ are computed.

The time taken by DeriveRS to compute $\mathbf{X}^*$ is typically not an overhead, as it is reported to be much less than the time taken to train an SVM on $\mathbf{X}$ (Nandan et al., 2013). Moreover, when a grid search is performed to identify the optimum hyper-parameters, DeriveRS has to be run only once for each distinct kernel hyper-parameter value. This is because DeriveRS is independent of the penalty hyper-parameter $C$.

The result of DeriveRS is used to solve AESVM and obtain a solution that approximates the SVM solution. AESVM solves the following optimization problem:

$$\min_{\mathbf{w},b} F_2(\mathbf{w}, b) = \frac{\|\mathbf{w}\|^2}{2} + \frac{C}{N}\sum_{t=1}^{M}\beta_t l_{\mathbf{w},b}(\mathbf{x}_t) \tag{4}$$
$$\text{where } l_{\mathbf{w},b}(\mathbf{x}_t) = max\{0, 1 - y_t(\mathbf{w}^T\phi(\mathbf{x}_t) + b)\}$$
$$\text{and } \phi : \mathbb{R}^D \rightarrow \mathbb{H}, b \in \mathbb{R}, \text{ and } \mathbf{w} \in \mathbb{H}, \text{ a Hilbert space}$$

where $M$ is the number of vectors in the representative set, $\mathbf{X}^*$. AESVM training can be performed using any standard SVM solver such as SMO (Platt, 1999). It was shown that AESVM training was several times faster than SVM training, even with the additional requirement of running DeriveRS. It was also reported that as a direct consequence of the relatively small $\mathbf{X}^*$ a reduction in the number of support vectors in the AESVM solution was observed.

### 2.3. Post-processing the AESVM solution with DrsPost

The SVM discriminating function is specified by $\mathbf{w}$ and $b$, as $f(\mathbf{x}) = \mathbf{w}^T\phi(\mathbf{x}) + b$. For non-linear kernels, $\mathbf{w}$ is typically represented as a weighted sum

$$\mathbf{w} = \sum_{r=1}^{R}\alpha_r y_r \phi(\mathbf{x}_r) \tag{5}$$

Here $\mathbf{x}_r$ are the support vectors (SVs), $R$ is the number of SVs, and $\alpha_r$ are non-negative weights resulting from Lagrangian optimization. Here we propose a post-processing algorithm DrsPost, to reduce the number of support vectors in the solution computed by AESVM.

DrsPost is a simple extension of DeriveRS. In DrsPost, the set of transformed data vectors used is $\mathbf{V} = \{\mathbf{v}_j : \mathbf{v}_j = \alpha_j\phi(\mathbf{x}_j), \forall\mathbf{x}_j \in \text{SVs}\}$. This leads to the use of a modified kernel $k'(\mathbf{x}_i, \mathbf{x}_j)$, during algorithm execution, given by:

$$k'(\mathbf{x}_i, \mathbf{x}_j) = \alpha_i\phi(\mathbf{x}_i^T)\alpha_j\phi(\mathbf{x}_j) = \alpha_i\alpha_j k(\mathbf{x}_i, \mathbf{x}_j)$$

Here $k(\mathbf{x}_i, \mathbf{x}_j)$ is the kernel used in AESVM training. We define the post-processing representative set $\mathbf{V}^*$, as given below:

$$\mathbf{V}^* \subseteq \mathbf{V} \tag{6}$$
$$\|\mathbf{v}_r - \sum_{\mathbf{v}_s \in \mathbf{V}^*, y_s = y_r}\gamma_s^r \mathbf{v}_s\|^2 \leq \rho \tag{7}$$
$$\text{s.t. } 0 \leq \gamma_s^r \leq 1, \text{ and } \sum_{\mathbf{v}_s \in \mathbf{V}^*, y_s = y_r}\gamma_s^r = 1$$

The parameter $\rho$ is the maximum approximation error, analogous to $\epsilon$ in (2). DrsPost is identical to DeriveRS, except for the use of the modified kernel $k'(\mathbf{x}_i, \mathbf{x}_j)$. The result of DrsPost is an approximation to $\mathbf{w}$, defined as

$$\mathbf{w}' = \sum_{s=1}^{S}\delta_s y_s \phi(\mathbf{x}_s), \text{ where } \delta_s = \sum_{r=1}^{R}\alpha_s\gamma_s^r \tag{8}$$

where $\mathbf{x}_s$ are vectors in the post-processing representative set and $S$ is the size of $\mathbf{V}^*$. The error in approximation of $\mathbf{w}$ by $\mathbf{w}'$, is bounded as described in the following theorem.

**Theorem 1** *Let $\mathbf{w}'$ be as defined in (8) and $\mathbf{w}$, the solution of (4), be as defined in (5). Then,*

$$\|\mathbf{w} - \mathbf{w}'\| \leq R\sqrt{\rho}$$

*Proof.* From (8), we know that $\mathbf{w}' =$

$\sum_{r=1}^{R} y_r \alpha_r \sum_{s=1}^{S} \gamma_s^r \phi(\mathbf{x}_s)$. Using (5) and (7), we get:

$$\|\mathbf{w} - \mathbf{w}'\|^2 = \left\| \sum_{r=1}^{R} y_r \left[ \alpha_r \phi(\mathbf{x}_r) - \sum_{s=1}^{S} \gamma_s^r \alpha_s \phi(\mathbf{x}_s) \right] \right\|^2$$
$$= \left\| \sum_{r=1}^{R} y_r \left[ \mathbf{v}_r - \sum_{s=1}^{S} \gamma_s^r \mathbf{v}_s \right] \right\|^2$$
$$= \sum_{r=1}^{R} \sum_{p=1}^{R} y_r y_p \left[ \mathbf{v}_r - \sum_{s=1}^{S} \gamma_s^r \mathbf{v}_s \right]^T \left[ \mathbf{v}_p - \sum_{s=1}^{S} \gamma_s^p \mathbf{v}_s \right]$$
$$\leq \sum_{r=1}^{R} \sum_{p=1}^{R} y_r y_p \rho \leq R^2 \rho$$

□

## 3. Experimental setup

In this section, we first briefly describe the datasets used in this study. Next, we describe the data pre-processing methods and the seizure detection features used. Finally, we describe the performance metrics that were used.

### 3.1. Datasets

We conducted our experiments on the seizure database CHB-MIT (Goldberger et al., 2000), collected by Children's Hospital Boston and used in Shoeb (2009). The database is comprised of 24 sets of scalp EEG recordings from 23 pediatric patients with intractable seizures. We have used only 23 of the sets of recordings due to time constraints. In this paper, we refer to each of set of recordings from a patient as a dataset. A summary of the datasets' characteristics are given in Table 3.1. The patients had been monitored for up to several days after the withdrawal of anti-epileptic drugs. The datasets have a total duration of 846 hours, stored as one hour long records. Overall, 181 seizure episodes were recorded. The records contain 18 - 23 signals, corresponding to the scalp EEG electrodes, sampled at 256 Hz. The international 10-20 system of electrode placement was followed

*Table 1.* Characteristics of datasets. The first column represents the patient number in the CHB-MIT database

| P. | Dataset | Age | No. of seizures | Duration (hours) | Total Seizure duration (s) |
|---|---|---|---|---|---|
| 1 | 1 | 11 | 7 | 41 | 446 |
| 2 | 2 | 11 | 3 | 35 | 174 |
| 3 | 3 | 14 | 7 | 38 | 406 |
| 5 | 4 | 7 | 5 | 39 | 564 |
| 6 | 5 | 1.5 | 10 | 67 | 166 |
| 7 | 6 | 14.5 | 3 | 67 | 328 |
| 8 | 7 | 3.5 | 5 | 20 | 924 |
| 9 | 8 | 10 | 4 | 68 | 282 |
| 10 | 9 | 3 | 7 | 50 | 456 |
| 11 | 10 | 12 | 3 | 35 | 808 |
| 12 | 11 | 2 | 27 | 21 | 1018 |
| 13 | 12 | 3 | 12 | 33 | 544 |
| 14 | 13 | 9 | 8 | 26 | 180 |
| 15 | 14 | 16 | 20 | 40 | 2010 |
| 16 | 15 | 7 | 10 | 19 | 90 |
| 17 | 16 | 12 | 3 | 21 | 294 |
| 18 | 17 | 18 | 6 | 36 | 326 |
| 19 | 18 | 19 | 3 | 30 | 240 |
| 20 | 19 | 6 | 8 | 28 | 302 |
| 21 | 20 | 13 | 4 | 33 | 204 |
| 22 | 21 | 9 | 3 | 31 | 210 |
| 23 | 22 | 6 | 7 | 27 | 432 |
| 24 | 23 | N.A. | 16 | 12 | 526 |

### 3.2. Data pre-processing and seizure detection features

It is well known that epileptic seizures are accompanied by a sustained increase in energy in low frequency bands (Subasi, 2007; Talathi et al., 2008; Shoeb & Guttag, 2010). For this study, we followed the procedure in Shoeb & Guttag (2010), where the signal energy in the frequency range 0.5 - 25 Hz was used to train a seizure detector. The EEG channel recordings were read using a non-overlapped sliding window of 2 seconds duration. Using 8 equal bandwidth filters, the spectral components of each signal in the 0.5 - 25 Hz range was extracted. The signal energy in the output of each filter was used as a feature. The signal energy is defined for a signal $\mathbb{S} = s_1, s_2, ....., s_T$, as $E = \sum_{t=1}^{T} s_t^2$. For a set of $L$ channels, the procedure described above results in the extraction of a vector $x_T$ of 8*$L$ features from the 2 second data window at time T.

In addition to the above mentioned features, Shoeb & Guttag (2010) followed a stacked feature vector scheme, where the 8*$L$ features at time T was stacked on the features from T-2 and T-4. Thus the feature vector $\mathbf{x}_T$ is the 24*$L$ dimensional vector formed by stacking $x_T$, $x_{T-2}$, and $x_{T-4}$. This process captures the temporal characteristics of seizure onset and is reported to give good seizure detection performance. Upon computing the features of a dataset, each feature was scaled to be in the range $[0, 1]$. An illustration of the feature extraction scheme is given in Figure 1. The signal energy extracted from the data read from EEG channel i, by filter j is represented as $x_{i,j}$ in the figure. To reduce the rate of false positives, Shoeb & Guttag (2010) used only the first 20 seconds of each seizure in their method and discarded the remaining portions of the seizure episodes. Nonetheless, we did not remove any part of any recording as we found that more seizures were detected when entire recordings were used.

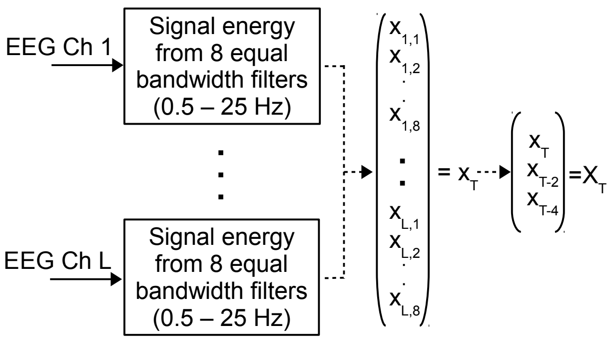


*Figure 1.* Overview of feature extraction.

### 3.3. Performance metrics

Three metrics were used to evaluate the proposed method.

- Sensitivity ($S$): the ratio of the number of seizures detected to the total number of seizures in a dataset.
- False positive rate ($FPR$): the average number of times in a 24 hour period that a seizure is wrongly reported for a dataset.
- Detection latency ($\tau$): the mean delay in seconds between seizure onset and its detection for a dataset.

The performance metrics were computed for *each dataset* using the leave-one-record-out cross validation method detailed in Shoeb & Guttag (2010). To estimate $S$ and $\tau$ for a dataset , each record $\mathbf{X}_s$ that has seizures in a dataset, was tested by a classifier trained on all the other records of the same dataset. We noted the number of seizures detected and the detection latency. This process was repeated on all records with seizures in the dataset, and their results were combined to obtain $S$ and $\tau$. To estimate $FPR$, each seizure free record $\mathbf{X}_{ns}$ in the dataset was tested by a classifier trained on all the other records in the dataset. The number of false positives reported for each $\mathbf{X}_{ns}$ was noted, and $FPR$ was calculated using the total number of false positives.

## 4. Results of comparison of methods

We started our experiments with a comparison of the seizure detection performance of our method with the SVM based method in Shoeb & Guttag (2010). The Gaussian RBF kernel, $k(\mathbf{x}_i, \mathbf{x}_j) = e^{(-g\|\mathbf{x}_i - \mathbf{x}_j\|^2)}$, was used in all the experiments. The DeriveRS parameter values were set as $P = 1000, V = 200$, and $\epsilon = 0.05$. The approximation error bound of DrsPost was set as $\rho = 0.01$. We performed a grid search of SVM and AESVM hyper-parameters to identify their optimum values. We chose a grid of all combinations of $C' = \{0.1, 1, 10, 100\}$ and $g = \{0.1, 0.5, 1\}$, where $C' = \frac{C}{N}$. Using each of these 12 hyper-parameter combinations we trained both the methods.

### Optimality index

To simplify the task of finding the hyper-parameter combination that gave the optimum values of $S, FPR$, and $\tau$, a combination metric optimality index ($O$), was used (Nandan et al., 2010). The optimality index is defined as:

$$O = G_1(S) * (K + \tau'); \text{where } G_1(a) = \frac{e^a - 1}{e - 1},$$

$$K = max(0, 1 - \frac{FPR}{F_{max}}), \text{ and } \tau' = max(0, 1 - \frac{\tau}{\tau_{max}})$$

We set $\tau_{max} = 20\ s$ and $F_{max} = 20$, as the maximum desired values of detection latency and $FPR$. The hyper-parameter combination that gave the highest value of $O$, was chosen as the optimum. It can be seen that $O$ has an exponential relationship with $S$, and a linear relationship with $\tau$ and $K$. We chose a formula for $O$ that gives highest weight to $S$, because the primary purpose of a seizure detector is to detect the maximum number of seizures.

AESVM was found to give performance competitive with SVM on all performance metrics. Overall, 94% of seizures were detected by SVM while AESVM detected 93%. The mean detection latency and mean $FPR$ of SVM was 7.2 seconds and 0.5 respectively, while the mean detection latency and mean $FPR$ of AESVM was 7.5 seconds and 0.9 respectively. The specificity, false positive rate, and detection latency observed for each dataset are shown in Figures 2 - 4. The most noticeable difference in performance of the two methods is for dataset 17, where SVM resulted in $\tau = 10.4$ s and $FPR = 2$, while AESVM resulted in $\tau = 16.5$ s and $FPR = 10.8$. However, this performance difference is not very significant when we consider that this dataset contains only 3.3% of the total number of seizures in the database. All the results for AESVM presented in Figures 2 - 4, were obtained after post processing the AESVM solution using DrsPost. There was no noticeable difference between the seizure detection performance of AESVM with and without post processing using DrsPost.

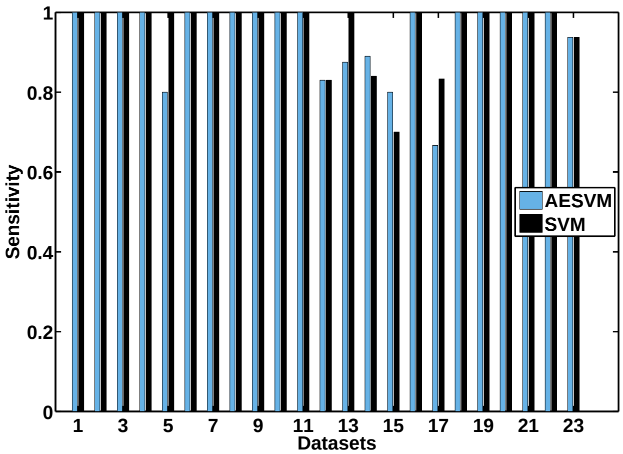


*Figure 2.* Sensitivity ($S$) of AESVM and SVM for each dataset.

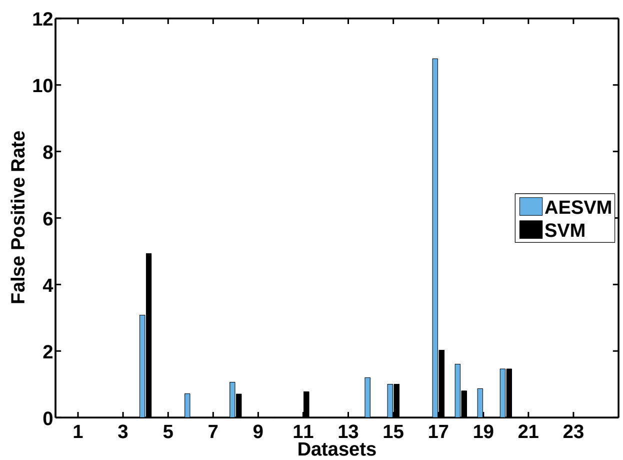


*Figure 3.* False positive rate ($FPR$) of AESVM and SVM for each dataset.

AESVM resulted in significant training and testing time speed ups in most cases. The average classification time speed-up attained by AESVM over all the datasets was 2.5, while the average training time speed-up over all the datasets was 5. The mean classification time speed-up for each dataset is illustrated in 5. The *mean classification time speed-up* is defined as the ratio of the total number of SVs computed by SVM, to the total number of SVs computed by AESVM (with and without DrsPost). Post processing the AESVM solution with DrsPost resulted in an average increase of mean classification time speed-up by 1.7 times.

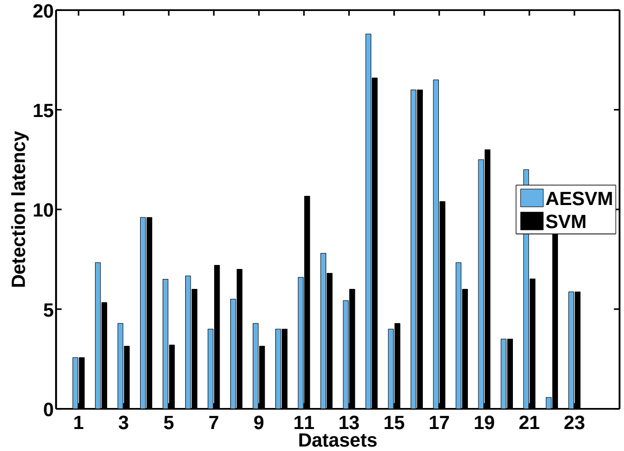


*Figure 4.* Seizure detection latency ($\tau$) of AESVM and SVM for each dataset.

Even though Figure 5 illustrates the classification time speed-up that can be expected by using our method, it is useful to observe the classification time speed-up at the optimal hyper-parameters. Figure 6 illustrates the classification time speed-up for AESVM (with and without DrsPost), when it was trained with the optimal hyper-parameters. The mean classification time speed-up at the the optimal hyper-parameters was found to be 3.5. The increase in classification time speed-up at the the optimal hyper-parameters, on post processing the AESVM solution using DrsPost was found to be on average 3 times. It should be noted that, by virtue of its log-linear computation complexity, the DrsPost computation time is negligible compared to the AESVM training time. Hence any increase in classification time speed-up is achieved with a negligible increase in computation time. A wide variation in the classification speed-up values for dataset 7 can be observed in Figure 6. This is due to the low optimal value of hyper-parameter $C' = 0.1$ for this dataset. On analyzing DrsPost we can see that it depends on the maximum value of $\alpha$ in (5), which in turn depends on $C'$. For small $C'$ values, DrsPost is found to result in large classification speed-ups.

The training time speed-up for each dataset is illustrated in Figure 7. The training time speed-up is defined as the ratio of the sum of training times of SVM to sum of training times of AESVM on a dataset. The

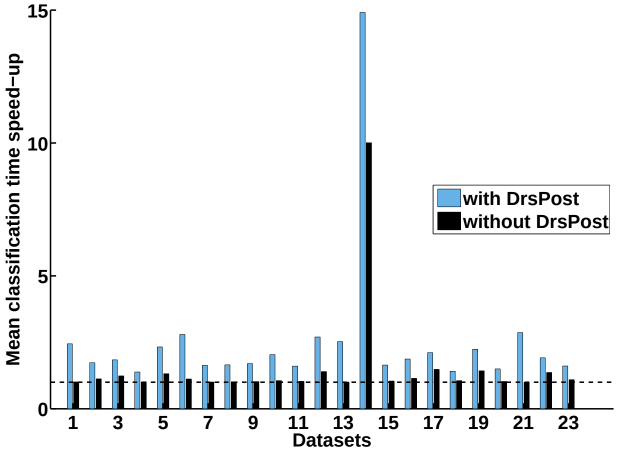


*Figure 5.* Mean classification time speed-up for each dataset. The dashed lined represents a speed-up of 1.

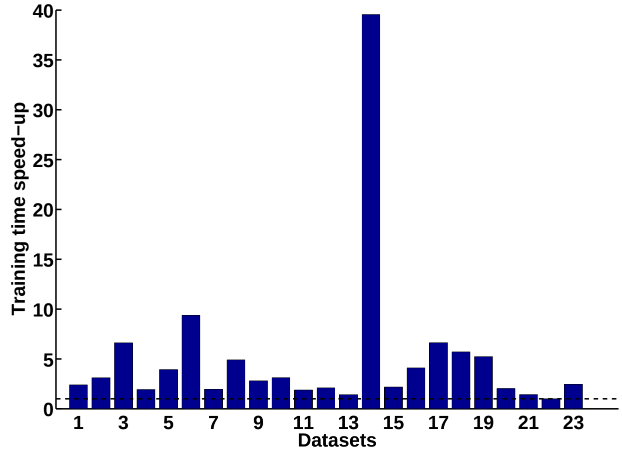


*Figure 7.* Training time speed-up for each dataset.

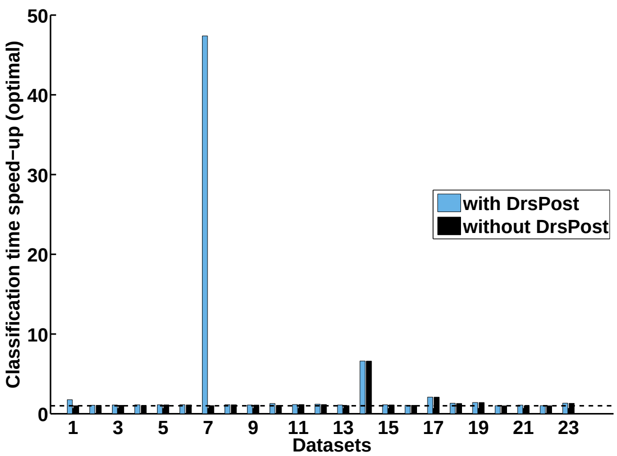


*Figure 6.* Classification time speed-up for each dataset with the optimal hyper-parameters.

AESVM training time includes the computation time of DrsPost. The training time speed-up varied from 1.1 to 39.5, with an average of 5.4. Hence our method can be expected to train and classify faster than the SVM based method on average. However, there is no significant change in seizure detection performance by using our method.

Nandan et al. (2013) reported much larger classification time speed-ups for some datasets. But, they had reported low classification time speed-ups for all high dimensional datasets. It was reported that DeriveRS resulted in comparatively large representative sets for high dimensional datasets. In this study, all the datasets used are high dimensional, with at least 432 dimensions. Hence it is reasonable to expect much higher speed-ups for datasets with lower dimensions.

## 5. Conclusion

In this paper we have proposed a method for real time seizure onset detection for use in closed loop seizure control systems. We have used the recently proposed AESVM classifier to give fast training and classification performance. In addition, we have proposed a novel algorithm DrsPost, to further reduce the classification complexity of our method. We have demonstrated that our method is competitive with an SVM based system, while being computationally more efficient. Our method was observed to be on average 2.5 times faster in classification and 5 times faster in training than the SVM based approach. We have also proposed an optimality index based method to select the optimum AESVM hyper-parameters, that simplifies the application of our method to clinical use.

To encourage the use of the proposed method by other researchers, we have made our software implementation publicly available[1]. The authors will next experiment on popular seizure detection features, such as wavelet energy, and extract fewer dimensions to further decrease the classification complexity of the method.

## Acknowledgment

Dr. Khargonekar acknowledges support from the Eckis professor endowment at the University of Florida. Dr. Talathi was partially supported by the Children's Miracle Network, and the Wilder Center of Excellence in Epilepsy Research. The authors acknowledge Mr.

[1] http://www.cise.ufl.edu/~mnandan/

Shivakeshavan R. Giridharan, for providing assistance with computational resources.

## References

Gardner, A.B., Krieger, A.M., Vachtsevanos, G., and Litt, B. One-class novelty detection for seizure analysis from intracranial EEG. *Journal of Machine Learning Research*, 7:1025–1044, 2006.

Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. Ch., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., and Stanley, H. E. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.

Groves, D. A. and Brown, V. J. Vagal nerve stimulation: a review of its applications and potential mechanisms that mediate its clinical effects. *Neuroscience & Biobehavioral Reviews*, 29(3):493 – 500, 2005.

Kahane, P. and Depaulis, A. Deep brain stimulation in epilepsy: What is next? *Current opinion in neurology*, 23(2):177, 2010.

Liang, S.F., Liao, Y.C., Shaw, F.Z., Chang, D.W., Young, C.P., and Chiueh, H. Closed-loop seizure control on epileptic rat models. *Journal of Neural Engineering*, 8, 2011.

Nandan, M., Talathi, S. S., Myers, S., Ditto, W. L., Khargonekar, P. P., and Carney, P. R. Support vector machines for seizure detection in an animal model of chronic epilepsy. *Journal of Neural Engineering*, 7, 2010.

Nandan, M., Talathi, S.S, and Khargonekar, P.P. Fast SVM training using approximate extreme points. *arXiv preprint 1304.1391 (http://arxiv.org/abs/1304.1391)*, 2013.

Platt, J.C. Fast training of support vector machines using sequential minimal optimization. In *Advances in kernel methods*, pp. 185–208. MIT Press, 1999.

Schölkopf, B., Knirsch, P., Smola, A., and Burges, C. Fast approximation of support vector kernel expansions, and an interpretation of clustering as approximation in feature spaces. In *Mustererkennung 1998*, Informatik aktuell, pp. 125–132. Springer Berlin Heidelberg, 1998. ISBN 978-3-540-64935-9.

Shoeb, A. and Guttag, J. Application of machine learning to epileptic seizure detection. In *Proceedings of the international conference on Machine learning*, 2010.

Shoeb, A., Carlson, D., Panken, E., and Denison, T. A micropower support vector machine based seizure detection architecture for embedded medical devices. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4202–4205. IEEE, 2009.

Shoeb, A. H. *Application of machine learning to epileptic seizure onset detection and treatment.* PhD thesis, Massachusetts Institute of Technology, 2009.

Stacey, W. C. and Litt, B. Technology insight: neuroengineering and epilepsydesigning devices for seizure control. *Nature Clinical Practice Neurology*, 4(4): 190–201, 2008.

Subasi, A. EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Systems with Applications*, 32(4):1084 – 1093, 2007.

Talathi, S. S., Hwang, D. U., Spano, M. L., Simonotto, J., Furman, M. D., Myers, S. M., Winters, J. T., Ditto, W. L., and Carney, P. R. Non-parametric early seizure detection in an animal model of temporal lobe epilepsy. *Journal of Neural Engineering*, 5:85–98, 2008.

Velasco, F., Velasco, M., Velasco, A. L., Jimenez, F., Marquez, I., and Rise, M. Electrical stimulation of the centromedian thalamic nucleus in control of seizures: Long-term studies. *Epilepsia*, 36(1), 1995.

Verma, N., Shoeb, A., Bohorquez, J., Dawson, J., Guttag, J., and Chandrakasan, A.P. A Micro-Power EEG Acquisition SoC With Integrated Feature Extraction Processor for a Chronic Seizure Detection System. *IEEE Journal of Solid-State Circuits*, 45 (4):804–816, 2010.

Zhang, J., Laiwalla, F., Kim, J.A., Urabe, H., Wagenen, R.V., Song, Y.K., Connors, B.W., Zhang, F., Deisseroth, K., and Nurmikko, A.V. Integrated device for optical stimulation and spatiotemporal electrical recording of neural activity in light-sensitized brain tissue. *Journal of Neural Engineering*, 6, 2009.