Support vector machines for seizure detection in an animal model of chronic epilepsy

Manu Nandan1,7, Sachin S Talathi2,3,7,8, Stephen Myers2, William L Ditto4, Pramod P Khargonekar5 and Paul R Carney2,3,6

1 Department of Computer and Information Science and Engineering, University of Florida, FL 32611, USA
2 J Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, FL 32611, USA
3 McKnight Brain Institute, University of Florida, FL 32611, USA
4 School of Biological and Health Systems Engineering, Arizona State University, AZ 85287, USA
5 Department of Electrical and Computer Engineering, University of Florida, FL 32611, USA
6 Department of Pediatrics, University of Florida, FL 32611, USA
7 M Nandan and S Talathi made equal contributions.
8 Author to whom any correspondence should be addressed.

1. Introduction

Support vector machines (SVMs) [3, 7] are a general class of machine learning algorithms with diverse applications ranging from handwriting recognition [1] to time series analysis [4]. SVMs have become very powerful and widely accepted as they have several advantages over other machine learning techniques [2]. Recently, SVMs have also been utilized in the EEG data classification problem for seizure detection [10, 24, 28]. The use of traditional SVMs for seizure detection however poses a unique problem of unbalanced datasets, i.e. the amount of ictal data, if available at all, is usually minuscule compared to the amount of inter-ictal data for training of SVMs. Thus motivated, Gardner et al [10] reformulated the problem of seizure detection as a time series novelty detection problem. They proposed the usage of one-class SVM [20], a popular algorithm for outlier (‘novelty events’) detection, which can be trained using only the inter-ictal EEG datasets. Seizure detection is then reduced to the problem of identifying outliers in the EEG datasets. Although this approach presents...
several advantages over the traditional approach [10] of binary classification, a significant shortcoming of this approach is the lack of flexibility in allowing for the inclusion of information from EEG seizure epochs, when available, to improve the performance of novelty detection. This lack of flexibility is exacerbated particularly for the seizure detection application in long-term EEG recordings from an in vivo animal model of chronic epilepsy. Chronic recordings from experimental animal models of epilepsy are increasingly used to investigate the dynamics of progression of epileptic disease following brain trauma such as status epilepticus [19, 22]. EEG signals in this case contain several noise artifacts, including movement of jaw muscles while chewing, scratching of headstage and clipping of EEG signals during chronic data recording sessions. The likelihood of misclassifying these artifacts as novelty events is higher when specific knowledge from EEG seizure morphology is not utilized to distinguish the artifact-induced novelty events from the true seizure-induced novelty events. Support vector data description (SVDD) [23] is a related novelty detection technique, which unlike one-class SVM, can be used to make the detector more sensitive to EEG seizures when ictal data are available.

In order to evaluate our hypothesis that the inclusion of information from the knowledge of seizure morphology can indeed improve performance of seizure detection under the paradigm of novelty detection, we report on a rigorous comparison study evaluating the performance of one-class SVM, SVDD and a traditional SVM technique for seizure detection using EEG data collected from an in vivo animal model of chronic epilepsy. The paper is organized as follows: in section 2, we first describe the experimental procedure used to create an animal model of chronic epilepsy and the method for EEG data acquisition. In order to determine how the performance of SVM varies depending on the location of recording electrode placement for EEG data acquisition [21], the EEG data were collected from animals in two groups labeled group A and group B. The difference between the two groups was in the location of surgical implantation of the recording electrode from the site of electrical stimulation for brain injury (section 2.1.1). We then provide a brief introduction to the optimization problem solved by each SVM type and demonstrate their utility in the classification of a binary unbalanced toy dataset. Next, we present the mathematical details of the univariate measures derived from the underlying EEG data that form the feature space for data classification used by the SVMs. The statistical metrics used to compare the performance of each SVM type are then presented. Results are described in section 3 where we present two distinct strategies to identify optimal control parameters for the SVMs through a twofold cross-validation technique. We found that SVDD is a more efficient and reliable seizure detector than the traditional SVM and one-class SVM. The results of our analysis are summarized in table 3.

All the data analysis was performed using Matlab running on a 32 node Mac (PowerPC) cluster available to researchers at the University of Florida, Department of Biomedical Engineering. The SVM computation was done using LIBSVM [5]. All the data analyzed in this work and the Matlab M-files are available from the author (SST) upon request.

### 2. Methods

#### 2.1. Experimental setup

##### 2.1.1. Animals

All experiments were performed on 2 month old Sprague–Dawley rats (\(n = 5\)) weighing 210–265 g. The experimental protocols and procedures were approved by the Institution of Animal Care and Use Committee of the University of Florida.

##### 2.1.2. Surgery and electrode implantation

The top of the head was shaved and chemically sterilized with iodine and alcohol. The skull was exposed by a mid-sagittal incision that began between the eyes and extended caudally to the level of the ears to expose the bregma and lambdoidal suture. A peroxide wash was applied to the skull to remove excess soft tissue.

In three animals representing group A, dual purpose recording/stimulating bipolar twist teflon-coated stainless steel wires of 330 \(\mu\)m diameter, labeled as the macro electrodes, were placed bi-laterally in the ventral hippocampal area (\(-5.3 \text{ mm AP, } \pm 4.9 \text{ mm ML, and } -5 \text{ mm DV}\)). Two animals which formed group B were implanted with the macro electrodes bi-laterally in the CA1 area of the hippocampus (\(-4.3 \text{ mm AP, } \pm 2.0 \text{ mm ML, and } 2.8 \text{ mm DV}\)) for EEG recordings, while for stimulation a macro electrode was placed in the right ventral hippocampus region. In all of the animals, a reference screw electrode was placed on the midline 2 mm caudal to lambda and a ground screw electrode was placed just rostral to bregma (AP 2 mm, lateral 1 mm). Two additional anchoring, 0.8 mm stainless-steel screws were placed in the skull. All the electrodes were then connected to an A-M Systems (Sequim, WA) connector. A headstage cap was created by cementing the connector to the skull and anchoring the screws with cranialplast cement (Plastics One, Inc., Roanoke, VA). The schematic diagram for the layout of electrodes in the two groups of rats is shown in figure 1.

![Figure 1. Schematic diagram demonstrating the electrode placement in two animal groups.](Image)
2.1.3. Induction of status epilepticus. Status epilepticus was induced in each of the five animals following 1 week of recovery post-surgery. During stimulation animals were placed in specially constructed recording chambers. The stimulating electrode was connected to an A-M Systems Isolated Pulse Stimulator (A-M Systems Inc., Carlsborg, WA). The stimulation consisted of a 50 Hz square wave pulse. The current ranged from 100 to 400 mA. For each animal, the current level was selected by the experimenter in response to the animal’s reaction to the stimulation. The pulse trains were delivered in a 10 s on and 2 s off duty cycle [13]. During the stimulus, the usual behavioral response of the animal was the display of ‘wet dog shakes’ and increased exploratory activity. After approximately 20–30 min of stimulation, convulsive seizures in the form of ‘wet dog shakes’ (up to 1 min in duration) were usually observed about every 10 min. At the end of the stimulus period which lasted around 60–90 min, continuous EEG recordings were observed for evidence of slow waves in all recorded channels. If slow waves were not demonstrated, then the stimulus was re-applied for 10 min intervals for one to three more times until continuous slow waves appeared following termination of the stimulus. Upon termination of continuous hippocampal stimulation, the EEG continued to demonstrate activity below 5 Hz for 12–24 h and intermittent spontaneous 30–60 s electrographic seizures for 2–4 h. Animals were observed for seizure activity and provided with adequate food and water intake for 12–24 h after stimulation.

2.1.4. Animal housing. Following recovery from surgery, the animals were moved to the vivarium and maintained in a general purpose rat housing in a climate (21 °C) and light controlled (12 h light/dark cycle) environment with free access to food and water. During periods of EEG recordings, the animals were moved to specially constructed recording chambers. These chambers allowed for continuous video monitoring and the passage of recording wires out of the cage via a commutator. The housing consisted of a 10 in diameter, 12 in tall clear acrylic tube. The floor of the housing was a soft plastic mesh that allowed for animal waste to fall below the cage to a collection tray. The top of the housing was also constructed of clear acrylic and designed to allow for adequate airflow and support of the commutator above the cage.

2.1.5. EEG/video acquisition. Three weeks post-induction of status epilepticus, the animals were moved into the recording chambers. Each animal was connected to a Grass Amplifier (Grass Technologies, West Warwick, RI) via a 12-channel commutator. The amplifier band-pass filtered the EEG from 0.3 to 300 Hz. The resulting output of the amplifier was digitized with a 12-bit National Instruments A/D card and custom written software, which time locked the EEG and video data. The EEG was recorded continuously at 1 kHz and were screened daily for the presence of seizures. If a seizure was not seen within the first 3 days of screening the animals EEG, the animal was disconnected and returned to general housing. A week later the animal was again connected and rescreened for the appearance of seizures. Once seizures were detected, the EEG was recorded for a minimum period of 4 weeks.

2.1.6. Data characteristics. A total of 24 seizures from animals in group A and 33 seizures from animals in group B comprised the seizure dataset analyzed in this study. Although the EEG recordings were performed using two electrodes placed bilaterally in the hippocampal area of rat brain (figure 1), we analyzed the data from one electrode (contra-lateral to the brain stimulation site) using univariate measures for seizure detection as described in section 2.3. All the seizures were confirmed and classified by a board-certified epileptologist and experimental neurologist (PRC) and a biomedical engineer (SST) who made an independent determination of the timing of each electrographic seizure onset. The consensus on seizure onset time was reached upon by focusing on a well-defined time-horizon for seizure onset by each of the independent reviewers, and then choosing the minimal time window of the two independently obtained seizure horizons and assigning the mid-point of the seizure horizon as the seizure onset time. Details of the datasets considered for analysis are given in table 1.

We found that the EEG data from group B had more EEG artifacts induced by chewing and scratching as compared to the data from group A. Figure 2 shows a sample raw time trace

![Figure 2](image-url)
of the EEG with a chewing artifact from the data in group B. The significance of the larger amount of artifacts present in the data from group B in terms of its effect on the performance of the SVMs for seizure detection is described in section 3.

2.2. Support vector machines

Given data vectors \( \mathbf{x}_i \) \((i = 1, 2, \ldots, N)\), the standard SVM provides an optimal separating boundary (a hyperplane), to assign a label \( y_i \in [-1, 1] \) to each \( \mathbf{x}_i \) by solving the following optimization problem:

\[
\min_{\mathbf{w}, b, \xi} \left( \frac{1}{2} |\mathbf{w}|^2 + c \sum_{i=1}^N \xi_i \right),
\]

subject to the constraints \( y_i (\mathbf{w} \cdot \Phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \) and \( \xi_i \geq 0 \) where \( c \) is called the penalty parameter and \( \Phi(\mathbf{x}_i) \) is a nonlinear function that transforms \( \mathbf{x}_i \) to a higher dimensional feature space. Parameters \( \mathbf{w} \) and \( b \) define a hyperplane in the feature space consisting of data vectors \( \mathbf{x} \) that satisfy the constraint \( \mathbf{w} \cdot \Phi(\mathbf{x}) + b = 0 \). Variables \( \xi_i \) are slack variables that define a soft margin around the hyperplane. The optimization problem (1) is solved using Lagrangian optimization. The solution consists of a set \( M \leq N \) of data vectors called support vectors \( \mathbf{v}_i \) and the corresponding Lagrange multipliers \( \alpha_i \) \((i = 1, \ldots, M)\). These support vectors lie on or within the soft margin as shown in figure 3. Obtaining the solution requires use of kernel functions [16] and for all the work presented here the Gaussian radial basis function kernel was used.

Below we describe the specifics of three SVM extensions (types): weighted SVM [6], one-class SVM [10] and support vector data description (SVDD) [23], which are compared on the basis of their performance in the application to seizure detection.

- **Weighted SVM.** Here the soft margin for each class is controlled by a different penalty parameter. The hyperplane is obtained as the solution to the following optimization problem:

\[
\min_{\mathbf{w}, c} \left( \frac{1}{2} |\mathbf{w}|^2 + c_1 \sum_{i=1}^N \xi_i + c_2 \sum_{i=1}^N \xi_i \right),
\]

subject to the constraints \( y_i (\mathbf{w} \cdot \Phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \) and \( \xi_i \geq 0 \); \( c_1 \) and \( c_2 \) are the penalty parameters of each class and are the control parameters of weighted SVM, i.e. these are the parameters that control the performance of the SVM type. After the support vectors are derived, to check which class a data vector \( \mathbf{x} \) belongs to it is enough to check which side of the hyperplane the data vector lies on.

- **One-class SVM.** Here a hyperplane is derived with maximal separation between the data that belongs to only one class and the origin in feature space. This is achieved by constructing a hyperplane that is maximally distant from the origin with all data points lying on the opposite side from the origin and such that the margin is positive. The resulting hyperplane is obtained as the solution to the following optimization problem:

\[
\min_{\mathbf{w}, \rho} \left( \frac{1}{2} |\mathbf{w}|^2 + \frac{1}{vN} \sum_{i=1}^N \xi_i - \rho \right),
\]

subject to the constraints \( \mathbf{w} \cdot \Phi(\mathbf{x}_i) > \rho - \xi_i \) and \( \xi_i > 0 \) where \( v \) is a maximum bound on the fraction of outliers in the dataset and also a minimum bound on the fraction of support vectors; \( \mathbf{w} \) and \( \rho \) define the hyperplane as \( \mathbf{w} \cdot \Phi(\mathbf{x}) - \rho = 0 \) in feature space. \( v \) is the only control parameter of one-class SVM. Again, once the support vectors are derived, to check if a vector \( \mathbf{x} \) belongs to the single class it is enough to check which side of the hyperplane the vector is on. If it is on the side of the origin it is not in the class.

- **Support vector data description.** Here a hypersphere boundary of minimum radius enclosing data vectors from the predominant data class (positive samples) is derived in feature space. In addition to enclosing the data from the predominant class, it also excludes data from the other class (negative samples). The formulation of SVDD is as the following optimization problem:

\[
\min_{R, c, \xi} \left( R^2 + c \sum_{i=1}^N \xi_i + c_2 \sum_{i=1}^N \xi_i \right),
\]

subject to the constraints \( |\mathbf{x}_i - \mathbf{a}|^2 \leq R^2 + \xi_i \), \( |\mathbf{x}_i - \mathbf{a}|^2 \geq R^2 - \xi_i \), \( \xi_i \geq 0 \), \( \xi_i \geq 0 \) \( \forall i, l \), where subscripts \( i \) and \( l \) are for positive and negative samples, respectively. \( R \) is the radius of the hypersphere and \( \mathbf{a} \) is its center. The control parameters of SVDD are \( c_1 \) and \( c_2 \). After deriving the support vectors, all new data vectors that are found to be inside the hypersphere are considered to belong to the predominant data class.

SVDD can be considered as a modification to the one-class SVM algorithm in the sense that it can be applied to data with more than one class. As described above SVDD derives a hypersphere boundary of minimum radius that encloses most of the positive samples of data, i.e. data from the predominant class (exactly how many samples are outside the hypersphere is decided by the control parameters).
parameters). If negative samples are available, SVDD changes the radius of the hypersphere such that most of them are outside the region enclosing the predominant class. If negative samples are not available it still constructs a minimum radius hypersphere that encloses most of the positive samples. It is shown in [23] that for normalized one class datasets SVDD and one-class SVM formulations are comparable. Weighted SVM on the other hand can only be applied to two class datasets.

In order to demonstrate the difference between the performance of SVM types described above in discriminating unbalanced datasets, we have constructed a toy example comprising data from two classes (labeled 1 and 2) represented through color-coded data points in a two-dimensional Euclidean space referred to as the input space (see figure 4(a)). The design of the toy dataset was done so as to mimic the situation that exists in the application for seizure detection. For example, the number of data samples belonging to class 1 in the toy dataset is much greater than those belonging to class 2, analogous to the problem of seizure detection, where the amount of interictal data exceeds the ictal data by many folds. Our goal is to determine which SVM type performs better in identifying true class 2 data samples while misclassifying the least number of the data samples in class 1. The application of the three SVM types results in the construction of high density region of class 2 shown as the region in orange in figures 4(b)–(d).

We can see from figure 4(b) that the high density region for class 2 derived by the weighted SVM encompasses almost all the data samples of class 2, but some of the class 1 data samples are also present in this region. From figure 4(c) we see that the high density region of class 2 obtained by one-class SVM has missed some of the class 2 data samples. Figure 4(d) shows that the high density region of class 2 obtained by SVDD envelopes all the data samples of class 2 while including very few of the class 1 data samples unlike the case in figures 4(b) and 4(c) (notice the difference between the figures in the lower right corner). From this analysis we see that SVDD performs the best in terms of correctly identifying data samples in class 2 while minimally misclassifying data samples from class 1. It should be noted that one-class SVM and SVDD derive only high density regions for class 1 data and the area outside the class 1 high density region is represented manually to be the class 2 high density region.

2.3. Analysis procedure

We used the setup shown in figure 5(a) for our analysis of SVM performance for seizure detection. EEG data are read in windows of 4 s duration with an overlap of 3 s, as shown in figure 5(b). EEG data in this 4 s window are pre-processed to remove artifacts associated with isolated spike events as follows: we calculate the mean of the EEG data over a time epoch of 120 s starting 240 s before the current window and subtract it from the EEG data in the current window. The result $x(t)$ with a raw sampling rate of 1 kHz is then median filtered with a filter duration of 25 ms to generate a new time series $y(t)$, which now is effectively sampled at 40 Hz. This preprocessed data are then fed through a feature extractor that calculates the value of each feature for the given window. Hence each 4 s data window results in a single value for a given feature. The extracted features are normalized between [0, 1] and then used to train the SVM type under test. Normalization was performed by first subtracting the minimum value of each feature from all the values of the feature and then dividing by the maximum of the result. For training with one-class SVM, only the interictal data were used. For the other two SVM types the training data consisted of both interictal and ictal datasets.

2.3.1. Features. The choice of the following features for seizure detection was based on evidence from the existing
shown that EEG energy in the beta band (12–20 Hz) is particularly enhanced during a seizure epoch [14]. For seizure detection, since the localization of the energy content in a precise temporal window at or just before the start of clinical manifestation of seizure is very important, we used a continuous wavelet transform to determine wavelet energy (WE) in the beta band (wavelet scale corresponding to the Fourier frequency of 20 Hz). This was done by using the mean subtracted EEG data $x(t)$ since the median filtered data do not have enough samples for each window to apply the continuous wavelet transform for the chosen scale. The continuous wavelet transform of $x(t)$ is obtained as follows:

$$W(\tau, s) = \int_{-\infty}^{\infty} x(t) \frac{\Psi'(t/\tau)}{\sqrt{\tau}} dt,$$  

(7)

where $\tau$ and $s$ are translation and scale respectively. We used a Mexican hat mother wavelet

$$\Psi(t) = \frac{1}{\sqrt{2\pi}} (1 - t^2) e^{-t^2/2}.$$  

(8)

The WE in the beta band is then obtained as

$$WE = \frac{1}{N} \sum_{j=n-N+1}^{n} W(\tau_j, s_{\beta})^2,$$  

(9)

where $\tau_j = j\delta t$ and $\delta t = 1$ ms, and $s_{\beta}$ is the scale with Fourier frequency in the beta band and $N$ is the number of samples in the given time window.

### 2.4. Performance metrics

Performance of SVM types for seizure detection were gauged using the following statistical metrics:

(a) **Sensitivity**

$$S = \frac{N_{TP}}{N_{TP} + N_{FN}},$$  

(10)

$N_{TP}$ is the number of true positives and $N_{FN}$ is the number of false negatives. Here a true positive is the successful detection of a seizure within the detection horizon. The detection horizon is defined as the time interval starting from 90 s prior to the unequivocal time instance of seizure onset as defined by the expert personnel (PRC and SST) to 20 s after the seizure onset. The failure to report a seizure within the detection horizon is regarded as a false negative. $S$ indicates the fraction of seizures the classifier can be expected to detect.

(b) **Specificity**

$$K = \frac{N_{TN}}{N_{TN} + N_{FP}},$$  

(11)

$N_{TN}$ is the number of true negatives and $N_{FP}$ is the number of false positives. $K$ indicates the efficiency of the classifier in not triggering false detection events. In order to get a meaningful value of $K$, we calculate it by viewing the SVM output in 60 s bins as shown in figure 7. True negative bins are those bins that do not contain a true seizure event and are also not triggered by the SVM. Bins which contain SVM trigger but do not contain true seizure events are considered as false negative.
Comparison of the performance of each of the three SVM types (\(X = \text{weighted SVM, one-class SVM, SVDD}\)) was done using a twofold cross-validation technique [8] to identify the optimal set of control parameters that produced the best statistics for seizure detection in terms of sensitivity (\(S\)), specificity (\(K\)) and the mean seizure detection latency (\(\bar{\tau}\)). The choice of the twofold cross-validation technique instead of the more popular tenfold cross-validation is due to the limited computational power available, given the size of the data analyzed (see table 1). We note that only SVM training is computation intensive.

Let us consider the example of an EEG dataset from animals in group A to describe the procedure for twofold cross-validation. The total EEG data are divided into two equal subsets, labeled A1 and A2, each containing 12 seizures. We then choose a set of control parameters \(c^X_i\) \((i = 1, \ldots, N)\) for each SVM type X and train the SVMs using \(j\) to generate the support vectors \(\{v^X_j\}\) and the corresponding Lagrange multipliers \(\alpha_j^X\) \((j = A1, A2)\). The support vectors \(\{v^X_{A1}\}\) and the Lagrange multipliers \(\alpha_{A1}\) are then used to classify EEG from dataset A2 to produce the corresponding seizure detection statistics: \(S^X_{A2}, K^X_{A2}\) and \(\bar{\tau}^X_{A2}\). This procedure is reversed and the support vectors \(\{v^X_{A2}\}\) and the Lagrange multipliers \(\alpha_{A2}\) derived from dataset A2 are used to determine seizure statistics from dataset A1: \(S^X_{A1}, K^X_{A1}\) and \(\bar{\tau}^X_{A1}\). The mean statistics are calculated as \(Y^X = (Y^X_S + Y^X_K)/2\) where \((Y = S, K, \bar{\tau})\). These statistics are determined for each choice of control parameter \(c^X\) for a given SVM type.

We now describe the procedure followed to identify the optimal control parameter \(c^X\) that resulted in the best seizure detection statistics of a given SVM type. We adopt two different strategies to identify \(c^X\). The first strategy labeled as the blind strategy involves identifying the subset of control parameters \(\{c^X\}\) that produce 100% mean sensitivity for seizure detection. This is a pre-condition imposed by us with the assumption that a good seizure detector should be able to detect all true seizures. The subset of resulting control parameters are then ranked in the order of decreasing specificity. The control parameter with highest specificity is considered to be the optimal control parameter. In the event that there are more than one control parameters that result in the same value for highest specificity, we chose the control parameter with minimum detection latency. The second strategy to identify the optimal control parameter \(c^X\) is based on maximizing the optimality index defined below. As in the blind strategy, we again begin by selecting a subset of control parameters that result in 100% sensitivity for seizure detection. We then define a metric called the optimality index, which is a measure that combines the statistics \(K\) and \(\bar{\tau}\) for each of the seizures in a given dataset.

The optimality index is defined as

\[
O = \frac{G1(K) + G1(G2(\bar{\tau}))}{2},
\]

where

\[
G1(x) = \frac{e^x - 1}{e - 1}
\]

and

\[
G2(\tau) = \frac{\tau_{max} - \tau}{\tau_{max} - \tau_{min}}.
\]

The choice of function \(G1\) is motivated by our desire to give higher weightage to statistics closer to 1 while minimizing the contribution from statistics with values closer to zero. \(G2(\tau)\) converts the value of \(\tau\) to lie in the range \([0, 1]\), with the minimum \(\tau\) giving a value of 1. The value of \(O\) ranges from 0 to 1. The control parameter that results in the highest value of \(O\) is considered to be the optimal control parameter \(c^X\) for a given SVM type X. In table 2 we summarize the optimal control parameters for the two EEG datasets for each of the SVM type tested using the twofold cross-validation technique described above.

The results of performance comparison of each SVM type in seizure detection using the optimal control parameters obtained from blind strategy on each of the two EEG datasets are presented in figure 8. We see that SVDD performed better in terms of overall higher value of \(\hat{O}\) as compared to the other SVM types for both EEG datasets. For EEG data from animals in group A, SVDD gave a value of \(K\) that was more than 10% greater than those obtained for the other two SVM types. Although the \(\hat{\bar{\tau}}\) for SVDD was higher than for the others, the gain in \(K\) outweighs the loss in \(\hat{\bar{\tau}}\). As a result, the \(\hat{O}\) for SVDD is higher. For EEG data from animals in group B, SVDD and weighted SVM obtained comparable values of \(K\) but SVDD had a lower value of \(\hat{\bar{\tau}}\), and hence the \(\hat{O}\) for SVDD was higher. One-class SVM performed poorly for this dataset with a specificity less than
Figure 8. Results of cross-validation of SVM types with control parameters chosen using the blind strategy. All plots on the left are for dataset A while those on the right are for dataset B. The topmost histogram represents the $O$ and the error bar represents its standard error. The detection latency is represented using a standard box plot, with an additional dotted line representing the mean.

Table 2. Optimal control parameters derived for each SVM type using the different selection strategies. $\sigma$ is the width of the Gaussian radial basis function kernel.

<table>
<thead>
<tr>
<th>Selection strategy</th>
<th>SVM type</th>
<th>Optimal control parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset A</td>
<td>Weighted SVM</td>
<td>$\sigma = 0.14, c1 = 0.35, c2 = 118$</td>
</tr>
<tr>
<td></td>
<td>SVDD</td>
<td>$\sigma = 1.29, c1 = 0.01, c2 = 0.1$</td>
</tr>
<tr>
<td></td>
<td>One-class SVM</td>
<td>$\sigma = 0.19, \nu = 0.04$</td>
</tr>
<tr>
<td>Blind</td>
<td>Weighted SVM</td>
<td>$\sigma = 0.14, c1 = 0.35, c2 = 118$</td>
</tr>
<tr>
<td></td>
<td>SVDD</td>
<td>$\sigma = 1.29, c1 = 0.01, c2 = 0.1$</td>
</tr>
<tr>
<td></td>
<td>One-class SVM</td>
<td>$\sigma = 0.19, \nu = 0.04$</td>
</tr>
<tr>
<td>Optimality index</td>
<td>SVDD</td>
<td>$\sigma = 1.29, c1 = 0.01, c2 = 0.1$</td>
</tr>
<tr>
<td></td>
<td>One-class SVM</td>
<td>$\sigma = 0.19, \nu = 0.04$</td>
</tr>
<tr>
<td>Dataset B</td>
<td>Weighted SVM</td>
<td>$\sigma = 0.2, c1 = 0.74, c2 = 33.8$</td>
</tr>
<tr>
<td>Blind</td>
<td>SVDD</td>
<td>$\sigma = 0.91, c1 = 0.2, c2 = 0.4$</td>
</tr>
<tr>
<td></td>
<td>One-class SVM</td>
<td>$\sigma = 0.29, \nu = 0.08$</td>
</tr>
<tr>
<td>Optimality index</td>
<td>SVDD</td>
<td>$\sigma = 0.2, c1 = 0.92, c2 = 42$</td>
</tr>
<tr>
<td></td>
<td>One-class SVM</td>
<td>$\sigma = 0.2, c1 = 0.24, c2 = 0.4$</td>
</tr>
</tbody>
</table>

50%. One-way ANOVA with the post hoc Tukeys multicomparison test was applied to determine whether there was a statistically significant difference in the $O$ values of each of the SVM types. For EEG data in animal from group A, one-way ANOVA yielded $F(2, 69) = 4.94, p = 9.9 \times 10^{-3}$ and Tukey post-hoc comparison indicated that the $O$ values of weighted SVM and SVDD were significantly different while there was no statistical difference in the $O$ values of SVDD and one-class SVM. For dataset B, one-way ANOVA yielded $F(2, 96) = 28.33, p = 2.1 \times 10^{-10}$ and Tukey post hoc comparison indicated that the $O$ values of weighted SVM and SVDD were not significantly different while those of SVDD and one-class SVM differed significantly.

The results of performance comparison of each SVM type using the optimal control parameters obtained from optimality index-based strategy are presented in figure 9. The results for EEG data from animals in group A using both the strategies were the same and hence only the results for EEG data from animals in group B are shown for the optimality index-based strategy. Again, we see that SVDD performed better in terms of overall higher value of $O$ as compared to the other SVM types. Even though SVDD gave a value of $K$ that was slightly
Results of cross-validation of SVM types with control parameters chosen using the optimality index-based strategy for dataset B. The histogram represents the $O$ and the error bar represents its standard error. The detection latency is represented using a standard box plot, with an additional dotted line representing the mean.

<table>
<thead>
<tr>
<th>Selection strategy</th>
<th>SVM type</th>
<th>$S$</th>
<th>$K$</th>
<th>$\bar{\tau}$</th>
<th>$\sigma_{\bar{\tau}}$ (standard error)</th>
<th>$O$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dataset A</td>
<td>Weighted SVM</td>
<td>1</td>
<td>0.74</td>
<td>$-11.54$</td>
<td>1.25</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>SVDD</td>
<td>1</td>
<td>0.85</td>
<td>$-9.08$</td>
<td>1.45</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>One-class SVM</td>
<td>1</td>
<td>0.75</td>
<td>$-15.79$</td>
<td>1.58</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>Weighted SVM</td>
<td>1</td>
<td>0.74</td>
<td>$-11.54$</td>
<td>1.25</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>SVDD</td>
<td>1</td>
<td>0.85</td>
<td>$-9.08$</td>
<td>1.45</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>One-class SVM</td>
<td>1</td>
<td>0.75</td>
<td>$-15.79$</td>
<td>1.58</td>
<td>0.79</td>
</tr>
<tr>
<td>Dataset B</td>
<td>Weighted SVM</td>
<td>1</td>
<td>0.81</td>
<td>$-4.58$</td>
<td>0.77</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>SVDD</td>
<td>1</td>
<td>0.82</td>
<td>$-15.64$</td>
<td>0.94</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>One-class SVM</td>
<td>1</td>
<td>0.46</td>
<td>$-48.88$</td>
<td>1.09</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>Weighted SVM</td>
<td>1</td>
<td>0.81</td>
<td>$-9.21$</td>
<td>0.88</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>SVDD</td>
<td>1</td>
<td>0.79</td>
<td>$-26.18$</td>
<td>1.06</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>One-class SVM</td>
<td>1</td>
<td>0.46</td>
<td>$-48.88$</td>
<td>1.09</td>
<td>0.62</td>
</tr>
</tbody>
</table>

less than that for weighted SVM, the $\bar{\tau}$ of SVDD was much less than that for weighted SVM. Again, one-class SVM performed poorly with a specificity less than 50%. One-way ANOVA yielded $F(2, 96) = 26.33$, $p = 7.7 \times 10^{-10}$ and Tukey post hoc comparison indicated that the $O$ values of weighted SVM and SVDD were not significantly different while the $O$ values of SVDD and one-class SVM were significantly different.

In the optimality index-based strategy, equal weightage was given to both the specificity and latency for seizure detection. Hence it is expected that the detection latency is less for the results of twofold cross-validation using the optimal control parameters obtained with this strategy as compared to the performance resulting from usage of the optimal control parameters obtained with the blind strategy. The detailed results given in Table 3 confirm this expectation. It is seen that among the results of the blind strategy, the worst $\bar{\tau}$ is for weighted SVM on EEG data from animals in group B with a value of $-4.58$ s. In the result of optimality index-based strategy for weighted SVM on the same EEG data $\bar{\tau}$ improved to $-9.21$ s. In both cases the value of $K$ is the same.

Similarly for SVDD $\bar{\tau}$ reduces to $-26.18$ s from $-15.64$ s when the optimality index-based strategy is used instead of blind strategy with only a slight decrease in $K$.

One-class SVM performed poorly on EEG data from animals in group B with both the optimal control parameter selection strategies. As described earlier, this is primarily because of the presence of more EEG artifacts in this dataset that resemble the characteristics of an actual EEG seizure. As a result the features trigger more false positives. This means in feature space at least some of the inter-ictal data vectors corresponding to artifacts lie in the region occupied by the ictal data vectors. As the one-class SVM is trained with only inter-ictal data, it cannot fine tune its decision boundary to include these artifacts while excluding ictal data. To avoid misinterpretation of ictal data as inter-ictal data and thereby to attain 100% $S$, it is necessary for a one-class SVM to consider all the inter-ictal data vectors in the ictal data region in feature space as outliers. It might be possible to improve the performance of one-class SVM for this dataset by selecting
features that can more efficiently differentiate artifacts from ictal data, but that is beyond the scope of this study.

The maximum bound on outlier fraction is specified for one-class SVM by the control parameter $\nu$. Increasing the upper bound on outlier fraction results in more inter-ictal data vectors to be considered as ictal data and hence increases the number of false positive bins $N_{FP}$ (see section 2.4) and reduces the value of $K$. It can be seen in table 2 that the value of $\nu$ for dataset from group B is double the value of $\nu$ for dataset from group A for both control parameter selection strategies. Hence, one-class SVM has a $K$ value for dataset from group B that is lower than that for dataset from group A. A consequence of the increased number of inter-ictal data vectors that are misinterpreted to be ictal data when $K$ decreases is that at least some of these misinterpretations happen at the beginning of the seizure onset time frame thereby decreasing the value of $\bar{\tau}$ for that seizure. Thus, we observe a low value of $\bar{\tau}$ for one-class SVM on dataset from group B. Weighted SVM and SVDD are trained using both inter-ictal and ictal data and hence does not have a large variation in performance on the two datasets.

Finally, since the derivation of optimal control parameters for each SVM for each given dataset is a very time-consuming process, we tested for the performance of optimal control parameters derived for one dataset applied to the second dataset. We conducted such an experiment with the optimal control parameters derived using the optimality index-based strategy. The results shown in table 4 indicate that the optimal control parameters are specific to each dataset and that they do not give good results when used on other datasets.

4. Discussion

The utility for SVMs in seizure detection is relatively new and a number of recent studies have focused on the application of binary SVMs to classify ictal states from non-ictal states [24, 25, 28]. However as we describe in the introduction, there are several advantages to viewing seizure detection as a time series novelty detection problem. Our goal in this work therefore was to compare the performance of two novelty detection formulations of SVM: one-class SVM and SVDD, against the widely used binary formulation of SVM for unbalanced datasets, the weighted SVM, for the application of seizure detection in an animal model of chronic epilepsy. The binary weighted SVM classifier was chosen because it has been shown to perform well in classifying unbalanced datasets. Seizure data for the seizure detection problem provides an unique example of unbalanced sets as the amount of inter-ictal data available far exceeds the available ictal datasets. One-class SVM [20] was chosen because it is a popular algorithm for outlier detection and has found application in the seizure detection problem [10] wherein the seizure epochs are treated as outlier or ‘novelty’ events. SVDD [23] offers the advantages of both the weighted SVM and the one-class SVM. It is an outlier detection algorithm with the flexibility of incorporating information from outlier events if available, to improve upon the classifier performance. A comparison study to assess the merit of these SVM types for seizure detection using the same EEG datasets is still lacking and to the best of our knowledge, this is the first study in which we systematically investigate the performance of these distinct SVM types to merit their usage for seizure detection.

For the present study, we obtained long-term continuous EEG data from two groups of rats with chronic epilepsy comprising a total of 273 h and 91 h EEG recordings. Since our primary goal was to test the suitability of using novelty detection formulation of SVMs as classifiers for seizure detection, in particular, to test the hypothesis that performance of novelty detection can be improved by incorporating information from seizure data when available, we extracted well-established features—mean energy, mean curve length and wavelet energy. These features were selected because (i) they are non-parametric measures derived from univariate time series, (ii) they have been used in earlier works [9, 10, 12, 14, 17, 21] for applications in seizure detection, (iii) each of these measures is a metric for the energy content in the EEG data and (iv) these measures can be easily adapted to use with human EEG datasets obtained with traditionally lower sampling rates of 200–250 Hz.

We chose three statistical performance evaluation metrics—sensitivity $S$, specificity $K$ and detection latency $\tau$, in order to gauge the performance of each SVM. The values for these metrics were obtained through twofold cross-validation on the available datasets. Two independent strategies were employed to identify the optimal set of control parameters that produced best performance metrics for each SVM type. Optimality index ($O$) statistic, which provides a trade-off between $K$ and $\tau$, was defined in order to compare the performance of the three SVMs resulting from the choice of the optimal SVM control parameters. Statistical significance for the resulting values of $O$ was tested using the one-way ANOVA test followed by a post hoc Tukeys multi-comparison test. SVDD resulted in the highest values of $O$ for both datasets A and B. We found that for dataset from group A the $O$ values of SVDD and weighted SVM were statistically significantly

<table>
<thead>
<tr>
<th>Dataset</th>
<th>SVM type</th>
<th>$S$</th>
<th>$K$</th>
<th>$\bar{\tau}$</th>
<th>$\sigma_{\bar{\tau}}$ (standard error)</th>
<th>$O$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Weighted SVM</td>
<td>0.79</td>
<td>0.95</td>
<td>1.4</td>
<td>0.5</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>SVDD</td>
<td>0.83</td>
<td>0.81</td>
<td>-14.1</td>
<td>0.46</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>One class SVM</td>
<td>0.79</td>
<td>0.82</td>
<td>-12</td>
<td>0.66</td>
<td>0.86</td>
</tr>
<tr>
<td>B</td>
<td>Weighted SVM</td>
<td>1</td>
<td>0.25</td>
<td>-72.6</td>
<td>0.82</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>SVDD</td>
<td>0.99</td>
<td>0.47</td>
<td>-54.7</td>
<td>0.38</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>One-class SVM</td>
<td>0.99</td>
<td>0.41</td>
<td>-56.9</td>
<td>0.57</td>
<td>0.62</td>
</tr>
</tbody>
</table>
different while the $O$ values of SVDD and one-class SVM were not. For dataset from group B the $O$ values of SVDD and one-class SVM were statistically significantly different while the $O$ values of SVDD and weighted SVM were not. A related observation is that $O$ for SVDD was similar for both datasets for both strategies ($\hat{O}$ was in the range 0.87–0.89). One-class SVM performed poorly on the dataset from group B as its inter-ictal data had more artifacts than the inter-ictal data of the dataset from group A as described in section 3. Our results show that not only did SVDD outperform weighted SVM and one-class SVM in both datasets but also it has good reliability. Based on these results our conclusion is that SVDD is better suited for seizure detection than weighted SVM and one-class SVM. Furthermore, SVDD has the advantage that it is flexible with its requirements for training datasets as it can be trained on datasets consisting of only inter-ictal data or both inter-ictal and ictal data unlike the other two SVM types. Another conclusion from our analysis is that one-class SVM performs poorly for datasets with large amounts of noise artifacts and hence cannot be considered as an efficient technique for seizure detection. We found that the optimal control parameter sets derived for one dataset do not provide good results when used with other datasets (table 4). An important issue worth noting in this regard is that the generality of the derived optimal control parameters for a given dataset also depends on the choice of the underlying measure used to construct the high dimensional feature space. There are approaches that aim to automate the process of derivation of control parameter sets [26, 27], which merit further study and are beyond the scope of the present work.

With the availability of more computing power, it should be possible to test SVM performance on large datasets with more rigorous cross-validation (fivefold or tenfold) techniques. SVMs are scalable with respect to the number of features and hence several other features including multivariate measures [15] can be investigated for use with SVMs, to build better seizure detectors. Studies such as this are ultimately intended to be used in the development of seizure prevention devices for patients with intractable epilepsy. The seizure prevention devices that are most desired are those that require least adaptation for each patient by being independent of EEG characteristics, electrode implant location, etc. Our results from the first study of its kind indicates that SVDD has at least a few of the desired qualities for a suitable automated seizure detector, and we hope that with SVDD we can enable the development of seizure warning devices that can be used as implantable closed loop seizure intervention systems.

Acknowledgments

The National Institutes of Biomedical Imaging and Bioengineering (NIBIB) through Collaborative Research in Computational Neuroscience (CRCNS) grant numbers R01 EB004752 and EB007082, the Wilder Center of Excellence for Epilepsy Research, a grant from the Office of Naval Research (grant number N00014-02-1-1019) and the Children’s Miracle Network supported this research. SST was partially funded by the research grant to WD from Arizona State University. PPK was partially supported by the Eckis Professor Endowment at the University of Florida.

References


[26] Friedrichs F and Igel C 2005 Evolutionary tuning of multiple SVM parameters Neurocomputing 64 107–17
