35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Seizure Prediction with Bipolar Spectral Power Features using Adaboost and SVM Classifiers

Mojtaba Bandarabadi, Antonio Dourado, Cesar A. Teixeira, Theoden I. Netoff, Keshab K. Parhi

Abstract—**This paper presents the results of our study on finding a lower complexity and yet a robust seizure prediction method using intracranial electroencephalogram (iEEG) recordings. We compare two classifiers: a low-complexity Adaboost and the more complex support vector machine (SVM). Adaboost is a linear classier using decision stumps, and SVM uses a nonlinear Gaussian kernel. Bipolar and/or time-differential spectral power features of different sub-bands are extracted from the iEEG signal. Adaboost is used to simultaneously classify as well as rank the features. Eliminating the low discriminating features reduces computational complexity and power consumption. The top features selected by Adaboost were also used as a feature set for SVM classification. The outputs of classifiers are regularized by applying a moving-average window and a threshold is used to generate alarms. The proposed methods were applied on 8 invasive recordings selected from the EPILEPSIAE database, the European database of EEG seizure recordings. Double-cross validation is used by separating data sets for training and optimization from testing. The key conclusion is that Adaboost performs slightly better than SVM using a reduced feature set on average with significantly less complexity resulting in a sensitivity of 77.1% (27 of 35 seizures in 873h recordings) and a false alarm rate of 0.18 per hour.**

Index Terms – Epilepsy, Seizure Prediction, Classification, Adaboost, Support Vector Machine, Power Spectral Density.

I. INTRODUCTION

Epilepsy is the second most prevalent brain disorder. About 60 million people worldwide suffer from epileptic seizures. In the United States, the cost to cover direct/indirect expenses related to epilepsy is estimated to be approximately $15.5 billion per year [1]. A major disadvantage of the disease to patients is the random nature in which seizures appear anytime, anywhere, and often without predictive symptoms, resulting in social difficulties or even injuries. Antiepileptic drugs as well as brain surgery are common therapies for epilepsy treatment; however, they often have critical side effects, are only effective in a fraction of the population, and many patients are not eligible for surgery [2]. An accurate seizure prediction algorithm implemented on an implantable device would allow for novel reactive therapies acting on time scales of seconds to minutes prior to the seizure onset. Any success in real-time seizure prediction could improve the living conditions of epileptic patients.

There is a need for seizure prediction algorithm with high sensitivity and specificity that can be implemented on an implantable device for closed-loop therapies. Such a device should be of low-power-budget and small enough to be implantable inside the skull cavity or worn as a portable pack. Unfortunately seizure prediction algorithms have not matured enough to be implemented in a standalone portable pack [3]. The needs for a device remain as follows: 1) A seizure prediction algorithm that has high enough sensitivity and specificity for medical purposes, and 2) This algorithm must be implemented on a device within a power budget of about 50μW as required by experts in the medical industry [2]. Power consumption is determined by the features to be extracted, the type of features, the number of channels monitored, sampling rate, and the classifier. In general, nonlinear features as well as nonlinear classifiers consume more power than linear ones. We suggest that an efficient classifier, such as Adaboost, using a subset of spectral power features, may achieve these goals.

A vast range of features has been tested with machine learning algorithms to predict seizures during the last decade [4-6], but often these findings cannot be reproduced in long-term EEG recordings [3, 7]. It has been shown that transient changes of power spectral density (PSD) of EEG signal preceding the seizure has significant predictive power [8-10]. In [11] it was argued that the spectral power in certain sub-bands of the iEEG, specifically in higher frequency sub-bands, could play a key role in seizure prediction. In [8] a patient-specific seizure prediction algorithm was proposed using a SVM classifier and bipolar iEEG recordings with 9 spectral bands from six EEG electrodes achieving a sensitivity of 97.5% (78 of 80 seizures, in 433.2 hours), and a false positive rate per hour (FPR) of 0.27.

This paper makes two contributions. First, it compares the capabilities of SVM and Adaboost classifiers for patient-specific seizure prediction task using spectral power features extracted from space differential recordings. SVM is a powerful classifier, but Adaboost is computationally much more efficient. Since the Adaboost algorithm has less computational cost, therefore, it is more suitable for low-power implantable devices. Second, it evaluates the efficiency of previously proposed bipolar spectral power features on new long-term continuous iEEG recordings. Results of prior studies were based on the Freiburg dataset [8, 11], which has fewer channels, where signals are sampled at a lower rate, and which are much shorter than those in the EPILEPSIAE database used in this paper.

* MB would like to acknowledge the Portuguese Foundation for Science and Technology (SFRH / BD / 71497 / 2010). This work was partially supported by the EPILEPSIAE project (EU FP7 211713), and by iCIS project (CENTRO-07-0224-FEDER-002003).

M. Bandarabadi, A. Dourado, and C.A. Teixeira are with the Centre for Informatics and Systems (CISUC), University of Coimbra, Portugal (email: [mojtaba, cteixei, dourado]@dei.uc.pt).

T.I. Netoff and K.K. Parhi are with the University of Minnesota, Minneapolis, MN 55455, USA (email: tnetoff@umn.edu, parhi@umn.edu).

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

II. METHODOLOGY

A. Subjects

Long-term continuous multichannel iEEG recordings of 8 patients with refractory partial epilepsy were chosen from EPILEPSIAE, the European database on epilepsy [12]. From among 70 candidate patients available in the database, only those with at least 6 seizures were considered with sampling rate of 400Hz. The iEEG signals were recorded at the epilepsy unit of the Pitié-Salpêtrière Hospital of Paris, France. Patient characteristics are summarized in Table I. For each patient, 9 channels were selected; six channels were selected from focal channels, and three from a non-focal area. The seizure focus was identified by a neurologist and annotated in the database.

Table I. Information for the 8 studied patients

ID	Sex	Pat. age	Onset age	Rec. time (h)	# seiz.	Seizure type	Localize
1	f	33	21	458.6	7	CP(4), UC(3)	Right T
2	m	27	16	234.9	6	SP(2), SG(2), CP(1) ,UC(3)	Left T
3	m	34	15	335	7	CP(3), SP(3) ,UC(1)	Right T
4	f	31	10	159.2	7	CP(2), SG(2), SP(1) ,UC(2)	Left T
5	m	38	27	342.1	9	CP(9)	Right T
6	m	40	12	221.2	8	CP(4), SG(1), UC(3)	Left T
7	m	36	1	211.7	12	UC(12)	Right F
8	m	36	6	253.9	6	SG(6)	Right P
A/T		34.4	13.5	2216	62		

- Seizure type: Type of the clinical seizures; SP: Simple Partial, CP: Complex Partial, SG: Secondarily Generalized, UC: Unclassified.
- Localize: Localization of seizures; T: Temporal lobe, F: Frontal lobe, P: Parietal lobe (between frontal and occipital).

B. Feature extraction

The spectral power features in different frequency bands are extracted from bipolar and/or time-differential windowed iEEG signals [8]. Spectral bands are determined by the standard Berger EEG bands: delta 'δ' (0.5-4Hz), theta 'θ' (4-8Hz), alpha 'α' (8-13Hz), beta 'β' (13-30Hz), except for the gamma 'γ' band which was divided into four smaller sub-bands of 30-48Hz, 52-75Hz, 75-98Hz, 102-200Hz. The total power also is considered as a feature. Power line noise is eliminated by excluding 48-52Hz, 98-102Hz, 148-152Hz, and 198-200Hz frequency bands from power spectral density (PSD). The *normalized spectral power* (NSP) feature is used here, which for a given sub-band is computed by dividing the spectral power of that sub-band to the total power using:

$$NSP_i = \frac{\sum_i PSD(x)}{\sum_{tot} PSD(x)} \quad (1)$$

where x is a portion of the raw EEG, and i and tot index values represent the i-th frequency sub-band and total frequency band, respectively. *PSD(x)* is the power spectral density which is estimated using the Welch's method [13]. The advantage of using the *NSP* values instead of absolute values is that they decrease the effect of changes in total power on the values of spectral power of sub-bands [5]. Length of the each window is chosen as 20 seconds, with an overlap of 50%, providing a seizure prediction every 10 seconds. By employing space-differential method, common-mode interferences such as movement artifacts and power line noises can be efficiently rejected [8]. Features were extracted from 9 iEEG channels selected from each of the 8 invasively recorded patients. For bipolar features, all possible pairwise combinations of 9 channels were considered. Thus we have 9*36=324 features for each of bivariate measures (Bipolar and/or Time-differential).

C. Feature preprocessing

Features were preprocessed prior to classification in order to reduce the effect of noise and to remove outlying samples. Outliers were detected using Grubb's methods [14], and replaced by values obtained by interpolating the values of their neighbors. Each resulting feature was then scaled to a range of [0 1] by dividing by the maximum value of that feature. All features were subsequently divided into training and test sets: features related to the part of signal containing first three seizures were used as training samples. The remaining signals and their related features were used for test. As supervised classifiers were investigated here, each sample had to be labeled. Seizure prediction was considered as a two-class problem: distinguishing between preictal and non-preictal states. Preictal samples were defined operationally as any recording within 30 minutes preceding a recorded seizure. All remaining data excluding data occurring 30 minutes following seizure onset were labeled as non-preictal. In such classification problems facing unbalanced classes, the classifier tends to produce higher accuracy for the dominant class [15]. To prevent this issue, the number of non-preictal samples of the training set was reduced by sub-sampling to achieve a balanced number of samples between the two classes. After training (optimizing the classifier parameters) using balanced number of training samples, the trained classifiers were tested using original non-reduced test samples.

D. Feature selection

Determining discriminative features from the 324 potential features extracted reduces computational complexity and can increase sensitivity and specificity. Classifiers such as Adaboost try to reduce the effective weight of features that are less discriminative during the training phase. However, this increases the computational cost exponentially with the number of features. On the other hand, the SVM classifier assigns same weights to all the features; therefore, inclusion of non-discriminating features degrades the classification. Non-discriminative features can also increase extensively the number of support vectors, thus reducing robustness of the SVM classifier. This, in turn, can increase the necessary time for training and testing the SVM classifier. In this study, feature ranking using Adaboost, first used in [16] for seizure prediction, was used.

E. Classification

The preictal and non-preictal classes generally cannot be separated by a single linear discriminator in the feature space. Therefore, linear classifiers usually perform poorly. Employing a non-linear classifier generally greatly improves prediction, but at a computational cost. To classify datasets with nonlinear boundaries, SVM uses kernel functions representing the data in a higher feature space where linear boundaries may separate data. The popular Gaussian Radial Basis Function (RBF) kernel (2) is used,

$$K(x,y) = exp(\frac{-|x-y|^2}{2\sigma^2}) \quad (2)$$

where σ is the scale parameter (openness of the Gaussian), and x, and y are feature vectors in the input space. The

Gaussian kernel has two hyper parameters to control classification performance: the cost C and the scale parameter σ. SVM with Gaussian kernel acts as a non-linear classifier. However its computational cost is very expensive. Furthermore, for best operation, some parameters should be optimized. Therefore, two SVM parameters of soft margin (C) and scale parameter (σ) are optimized here using a grid search. SVM algorithm used here is the one implemented in the LibSVM [17] toolbox.

Adaptive boosting approaches, such as Adaboost, use many weak classifiers together to obtain good classification results. Adaboost uses linear decision stumps (base classifiers), which are less computationally demanding, very fast and quite suitable for large classification problems such as in seizure prediction. The Adaboost classifier used is the one implemented in the 'Statistical Pattern Recognition Toolbox' (STPRtool) [18].

To reduce the number of false alarms, the classification outputs of both classifiers were subjected to regularization by the firing power method [19] that accounts for the classification dynamics in the preictal class. The length of moving average window of the firing power method was selected as 10 minutes, which covers 60 consecutive feature samples.

III. Experimental Results

First, Adaboost was trained on the full data set, and the features were then ranked by feature weight and average rank through bootstrapping. Afterward Adaboost was trained on the training samples using different sized feature sets, ranging from one feature (the highest ranked) to the top 30 most predictive features. The features and number of features that produced the highest sensitivity and lowest false positive rates were selected. The same features were then used for SVM to allow for direct comparison of results.

A. Performance Analysis

Sensitivity (SS) and false prediction rate (FPR) of the raised alarms were used to compare the two classifiers. The SS is the fraction of correctly predicted seizures within the total seizures, and the FPR is the number of false predictions per hour. Table III summarizes the SS, FPR and F_2-score [20] for 8 patients from the EPILEPSIAE database.

On average, SVM provided seizure predictions with a sensitivity of 75.7% and FPR of 0.17h^{-1}, whereas Adaboost generated sensitivity of 75.7% and FPR of 0.21h^{-1}. The results obtained from SVM and Adaboost are quite similar. SVM exhibited slightly better performance in terms of FPR. Also bipolar spectral power features provided lower FPR in comparison to bipolar time-differential ones. Although the average results of all patients are not surprising, however, for 5 patients we achieved prediction with SS of 100% and FPR of 0.16h^{-1} on average.

B. Complexity Analysis

An advantage with the Adaboost method over the SVM is that it has much lower complexity, which makes it more suitable for designing implantable devices with low power budgets. The hardware complexity and thus the power consumption of an SVM implementation with Gaussian kernel is directly proportional to $d * N_{sv}$, where d and N_{sv} are the dimension of feature space and the number of support vectors generated during training process, respectively [16]. The number of support vectors depends on separability of the features; the more nonlinear the separation boundary between the classes in feature space the greater number of support vectors is needed. The number of required features depends on how well the data is separated by the top-ranked features. In some patients classification can be achieved with a few features, while in others many more are needed. Thus the overall complexity of an SVM based seizure prediction algorithm depends on the discriminative quality of the features. On the other hand, the hardware complexity of Adaboost depends on the required numbers of comparison operations, which in turn is equal to the number of selected decision stumps. 60 decision stumps were considered in this work. The average optimal feature set was 14.6. The four first high ranked features for each patient are listed in Table IV.

Table III.a. Results of bipolar spectral power features

ID	#Select feature	Adaboost				SVM			
		SS	FPR	F_2-score	P-time	SS	FPR	F_2-score	P-time
1	11	100	0.16	0.26	937	75	0.04	0.21	1367
2	29	100	0.19	0.13	1605	100	0.23	0.1	575
3	6	100	0.18	0.1	843	100	0.16	0.08	1263
4	30	100	0.14	0.5	1581	100	0	0.43	1178
5	15	50	0.19	0.11	1103	50	0.04	0.17	573
6	19	100	0.07	0.31	1076	60	0.17	0.09	976
7	12	55.5	0.17	0.14	1556	66.6	0.27	0.15	1075
8	2	66.6	0.41	0.2	700	100	0.31	0.23	1386
A/T	15.5	77.1	0.18	0.22	1209	74.3	0.16	0.18	1031

Table III.b. Results of bipolar time-differential spectral power features

ID	#Select feature	Adaboost				SVM			
		SS	FPR	F_2-score	P-time	SS	FPR	F_2-score	P-time
1	6	75	0.24	0.16	1357	100	0.14	0.16	1110
2	13	100	0.09	0.14	480	100	0.16	0.11	1065
3	30	100	0.32	0.06	1583	66.6	0.12	0.06	1225
4	15	100	0.13	0.61	1068	100	0	0.3	1305
5	13	66.6	0.2	0.16	405	66.6	0.3	0.1	1210
6	20	80	0.29	0.18	1555	80	0.14	0.11	1475
7	9	55.5	0.32	0.14	702	55.5	0.21	0.18	558
8	4	66.6	0.24	0.23	1660	100	0.36	0.21	1337
A/T	13.75	74.3	0.25	0.21	1024	77.1	0.19	0.15	1080

- P-time: Average time distance between true alarms and their corresponding seizure onsets in second

Table IV. Four first high ranked features

ID	Bipolar				Bipolar time-differential			
	Feat1	Feat2	Feat3	Feat4	Feat1	Feat2	Feat3	Feat4
1	nf2-nf3-4	nf1-nf3-5	f2-f3-8	f2-nf2-7	f3-nf2-4	nf1-nf3-9	nf2-nf3-5	f2-nf1-7
2	f1-f3-8	f3-f4-4	f1-f3-9	f4-nf1-4	f5-nf1-7	f3-f4-6	nf2-nf1-4	f5-f2-9
3	f1-f2-9	f2-nf2-6	f5-f6-6	f2-f4-4	f1-f6-9	nf2-f5-3	f5-f6-9	f2-nf2-9
4	f2-nf2-8	f5-nf3-1	nf3-nf4-3	f4-f5-6	f2-nf2-3	f4-nf2-2	f3-nf2-6	f3-f5-3
5	nf2-nf3-4	f1-nf3-4	f2-nf1-2	f3-nf3-3	nf2-nf1-4	nf2-nf3-3	f3-nf3-2	f2-f3-8
6	nf2-f2-9	nf2-f3-8	f1-f5-9	nf1-nf2-4	nf2-f1-5	nf1-f4-4	nf3-f6-5	nf3-f5-6
7	f1-nf3-6	f3-f5-7	f1-f4-1	f1-f5-4	nf2-f1-7	f3-f2-9	f1-f5-5	nf2-f3-6
8	f3-f5-2	f1-f2-9	f3-f5-3	f6-f2-3	nf1-nf2-9	f3-f6-3	f4-f1-4	f6-f2-8

- ch1-ch2-freqBand, f: focal, nf: non-focal
- Frequency bands 1: delta 'δ' (0.5-4Hz), 2: theta 'θ' (4-8Hz), 3: alpha 'α' (8-13Hz), 4: beta 'β' (13-30Hz), 5, 6, 7, 8: gamma 'γ' band of 30-48Hz, 52-75Hz, 75-98Hz, 102-200Hz. 9: The total power

C. SVM parameters

We examined the relation between optimum C and σ values of the SVM classifier, by studying F_2-score of the classifier output. The best alarm sensitivity and FPR are obtained from the SVM outputs having the highest F_2-score. This measure was calculated for different C and Sigma (σ) values using a grid search, and an inverse relation between optimum C and σ values was found. Parameter C controls the tradeoff between maximization of the margin width and the minimization of the number of misclassified samples in the training set [21]. Also, the scale parameter (σ) controls the width of the Gaussian surface of the RBF kernel. Fig. 3 shows the F_2-score achieved for different combinations of C and σ, for one of the patients.

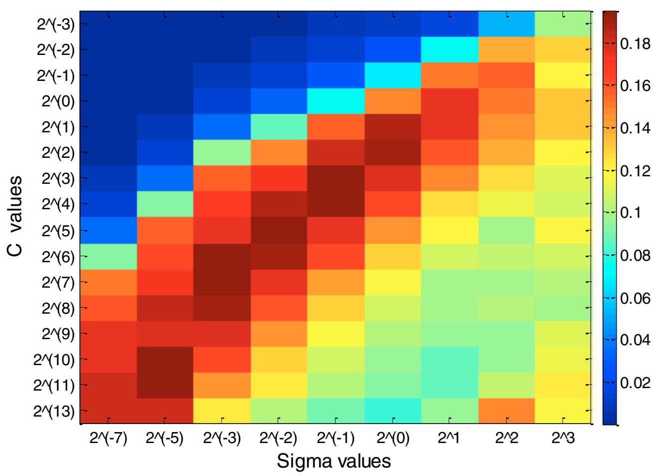

Figure 3. F_2-score of SVM classifier across different C and σ values for one of the studied patient.

Based on this finding it is suggested to limit the extensive grid search area to a diagonal region in the C-σ diagram only. This can significantly increase training speed. Furthermore it may be possible to govern the optimum C-σ relation by an equation or an inequality.

IV. Conclusion

A comparison between Adaboost and SVM classifiers for patient-specific seizure prediction was presented. Discriminative features were selected from amongst the bipolar and/or time-differential features by a boosting feature selection method. The lower-complexity Adaboost algorithm produced comparable results to those of SVM. As a result of this study, we find that linear-spectral power features are good features for classification, corroborating previous findings [8, 16].

References

[1] Epilepsy foundation. *Epilepsy foundation website*. Available: http://www.epilepsyfoundation.org/about/statistics.cfm

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

[3] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure prediction: the long and winding road," *Brain,* vol. 130, pp. 314-333, February 1, 2007.

[4] Suarez C. C. B., E. T. Fonoff, M. A. Munoz G, A. C. Godoi, G. Ballester, and F. J. Ramirez-Fernandez, "Wavelet transform and cross-correlation as tools for seizure prediction," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pp. 4020-4023, 2010.

[5] L. Chisci, A. Mavino, G. Perferi, M. Sciandrone, C. Anile, G. Colicchio*, et al.*, "Real-Time Epileptic Seizure Prediction Using AR Models and Support Vector Machines," *Biomedical Engineering, IEEE Transactions on,* vol. 57, pp. 1124-1132, 2010.

[6] W. Stacey, M. Le Van Quyen, F. Mormann, and A. Schulze-Bonhage, "What is the present-day EEG evidence for a preictal state?," *Epilepsy research,* vol. 97, pp. 243-251, 2011.

[7] R. G. Andrzejak, D. Chicharro, C. E. Elger, and F. Mormann, "Seizure prediction: Any better than chance?," *Clinical Neurophysiology,* vol. 120, pp. 1465-1478, 2009.

[8] Y. Park, L. Luo, K. K. Parhi, and T. Netoff, "Seizure prediction with spectral power of EEG using cost-sensitive support vector machines," *Epilepsia,* vol. 52, pp. 1761-1770, 2011.

[9] B. Direito, F. Ventura, C. Teixeira, and A. Dourado, "Optimized feature subsets for epileptic seizure prediction studies," in *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE*, pp. 1636-1639, 2011.

[10] M. Bandarabadi, C. A. Teixeira, B. Direito, and A. Dourado, "Epileptic seizure prediction based on a bivariate spectral power methodology," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pp. 5943-5946, 2012.

[11] Y. Park, T. Netoff, and K. K. Parhi, "Seizure prediction with spectral power of time/space-differential EEG signals using cost-sensitive support vector machine," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pp. 5450-5453, 2010.

[12] J. Klatt, H. Feldwisch-Drentrup, M. Ihle, V. Navarro, M. Neufang, C. Teixeira, C. Adam, M. Valderrama, C. Alvarado-Rojas, A. Witon, M. Le Van Quyen, F. Sales, A. Dourado, J. Timmer, A. Schulze-Bonhage, and B. Schelter, "The EPILEPSIAE database: An extensive electroencephalography database of epilepsy patients," *Epilepsia,* vol. 53, pp. 1669-1676, 2012.

[13] P. D. Welch, "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms," *IEEE Transactions on Audio and Electroacoustics,* vol. 15, pp. 70-73, 1967.

[14] F. E. Grubbs, *Procedures for Detecting Outlying Observations in Samples*, Defense Technical Information Center, 1974.

[15] N. V. Chawla, N. Japkowicz, and A. Kotcz, "Editorial: special issue on learning from imbalanced data sets," *SIGKDD Explor. Newsl.,* vol. 6, pp. 1-6, 2004.

[16] M. Ayinala and K. K. Parhi, "Low complexity algorithm for seizure prediction using Adaboost," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pp. 1061-1064, 2012.

[17] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.,* vol. 2, pp. 1-27, 2011.

[18] V. Franc and V. Hlavac. *Statistical Pattern Recognition Toolbox*. Available: http://cmp.felk.cvut.cz/cmp/software/stprtool/

[19] C. A. Teixeira, B. Direito, H. Feldwisch-Drentrup, M. Valderrama, R. P. Costa, C. Alvarado-Rojas, S. Nikolopoulos, M. Le Van Quyen, J. Timmer, B. Schelter, and A. Dourado, "EPILAB: A software package for studies on the prediction of epileptic seizures," *Journal of Neuroscience Methods,* vol. 200, pp. 257-271, 2011.

[20] S. Köknar-Tezel and L. J. Latecki, "Improving SVM Classification on Imbalanced Data Sets in Distance Spaces," presented at the Proceedings of the 2009 Ninth IEEE International Conference on Data Mining, 2009.

[21] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*: Mit Press, 2002.