IMPROVING KERNEL-ENERGY TRADE-OFFS FOR MACHINE LEARNING IN IMPLANTABLE AND WEARABLE BIOMEDICAL APPLICATIONS

Kyong Ho Lee, Sun-Yuan Kung and Naveen Verma

Department of Electrical Engineering, Princeton University, Princeton, NJ 08544
{kyonglee, kung, nverma}@princeton.edu

ABSTRACT

Emerging biomedical sensors and stimulators offer unprecedented modalities for delivering therapy and acquiring physiological signals (e.g., deep brain stimulators). Exploiting these in *intelligent, closed-loop* systems requires detecting specific physiological states using very low power (i.e., 1-10mW for wearable devices, 10-100μW for implantable devices). Machine learning is a powerful tool for modeling correlations in physiological signals, but model complexity in typical biomedical applications makes detection too computationally intensive.

We analyze the computational energy trade-offs and propose a method of *restructuring the computations* to yield more favorable trade-offs, especially for typical biomedical applications. We thus develop a methodology for implementing low-energy classification kernels and demonstrate energy reduction in practical biomedical systems.

Two applications, arrhythmia detection using electrocardiographs (ECG) from the MIT-BIH database [1] and seizure detection using electroencephalographs (EEG) from the CHB-MIT [1,2] database, are used. The proposed computational restructuring reduces energy by 2627x and 7.0-36.3x (depending on the patient), respectively.

***Index Terms*—** kernel-energy trade-off, energy efficiency, machine learning, biomedical devices

I. INTRODUCTION

The typical signals available for sensing in low-power, chronic biomedical systems are the result of complex and diverse physiological processes. Accurate physiology-based modeling thus faces severe challenges. *Data-driven* modeling, on the other hand, shows great promise for being both highly accurate and viable, particularly thanks to the recent availability of patient data in the healthcare domain [3]. Machine learning offers powerful methods for both the development and application of such models.

Biomedical sensors and stimulators have recently emerged showing unprecedented clinical efficacy for treating a broad range of diseases. Chronic implantable and wearable devices based on these can be made intelligent through embedded machine learning techniques. Such systems, however, face severe energy constraints.

The nature of data-driven modeling implies that the complexity depends on the characteristics of the data; most data-driven biomedical algorithms, however, have thus far paid little consideration to computational energy, especially in the context of micro- or milli-Watt devices.

The work of this paper is motivated by the energy constrained application of data-driven models through machine learning classifiers. The main contributions are summarized as follows:

- We propose a method of restructuring the computations of non-linear support vector machine (SVM) kernels to overcome energy scaling with the number of support vectors (which is representative of model complexity). Analysis shows model complexity to be a primary energy limitation. Polynomial kernels are exploited for retaining sufficient flexibility while also providing an opportunity to *precompute* the model decision matrix.
- We propose a methodology for low energy classifier implementation by characterizing the extensive trade-off space introduced by various kernels and their computational approaches. We show how the most favorable energy trade-off can be taken advantage of for an application, depending on the number of support vectors and features involved.
- We demonstrate the methodology using real patient data (from physiological databases) and a real processor model (i.e., MSPsim [4], a cycle accurate MSP430 instruction set simulator).

The paper is organized as follows. We describe two representative applications and analyze their computational energy in Section II. In Section III, we present the kernel-energy trade-offs and show how these can be altered for polynomial kernels through computational restructuring. We demonstrate the accuracy of using polynomial kernels in Section IV, and in Section V we show how the altered trade-offs lead to energy savings in the applications.

II. COMPUTATIONAL ENERGY ANALYSIS OF APPLICATIONS

Biomedical detection algorithms consist of biomarker extraction, to form a feature vector, and biomarker interpretation, using a classifier. Fig. 1 shows the applications that are used in this paper to illustrate the range of energy trade-offs affecting kernel-based classification. Both achieve state-of-the-art performance.

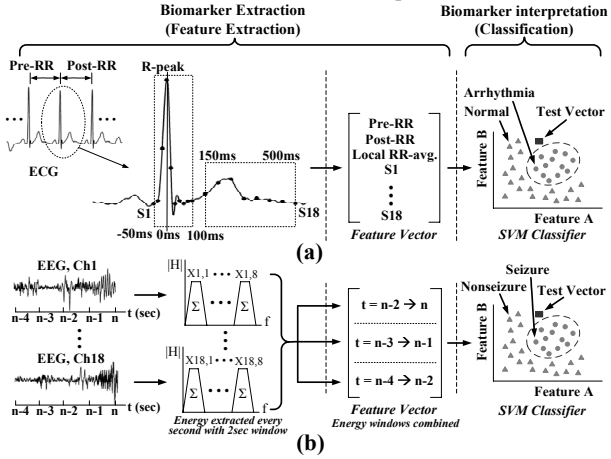


Fig. 1. (a) Arrhythmia detection [5] and (b) Seizure detection [6]. Physiological biomarkers form the classification features.

Arrhythmia detection [5] (Fig. 1a) uses waveform morphology features from each heart beat in the ECG. The classifier model is determined by training on a generic dataset derived from multiple patients. The correlations are thus diverse, leading to over 15k

support vectors. The feature vector has a dimensionality of 21.

Seizure detection [6] (Fig. 1b) uses spectral-energy features extracted from each EEG channel every second. The classifier model is determined by training on patient-specific data. Thus, fewer support vectors are needed (i.e. 84-625, depending on the patient), but feature-vector dimensionality can be as high as 432.

An SVM is used for classification in both cases. Most generally, radial basis function (RBF) kernels are used for their superior ability to handle non-linear decision boundaries. Feature extraction and RBF-SVM classification for the two applications is simulated using MSPsim, and the typical energy is provided in Table 1. As shown, classification energy dominates for both applications by well over an order of magnitude. As described in the next section, this is due to the complexity of the classification models, and, as a result, previous devices for chronic patient monitoring have been limited in their ability to incorporate the classifier [7].

Table 1. Feature extraction and classification energy. Classification energy dominates over feature extraction energy

Application	Feature Extraction Energy	Classification Energy	Ratio*
Arrhythmia	1.56 mJ per feature vector	49.52 mJ per classification	31.7
Seizure	1.44 mJ per feature vector	26.98 mJ per classification	18.7

*Ratio = (Classification Energy) / (Feature Extraction Energy)

III. KERNEL ENERGY TRADE-OFFS

Motivated by energy, this section analyzes how the computational complexity of classification scales and how the scaling trends can be altered favorably for the models required in biomedical applications. The classifier decision function is shown in Eq. 1, specifically for an RBF kernel.

$$\sum_{i=1}^{N} K(sv_i, x)\alpha_i y_i - b = \sum_{i=1}^{N} \exp\left(\frac{-|sv_i - x|^2}{\sigma^2}\right)\alpha_i y_i - b \quad \text{[RBF kernel]} \quad (1)$$

N is the number of support vectors, $K(\cdot)$ is the kernel function, sv_i are the support vectors, x is a test vector, and α_i, y_i, σ, and b are training parameters.

From Eq. 1, classification energy scales (roughly linearly) with both the number of support vectors (due to the summation) and the dimensionality of the feature vector (due to the vector arithmetic). The scaling is shown in Fig. 2 (derived from energy profiling using MSPsim), and it leads to the energy limitation seen in Section II, since typical biomedical applications require a high number of support vectors and/or a high number of features. In Section V we show how the high number of features can be managed; however, the high number of support vectors remains a primary concern that originates from the characteristics of the data.

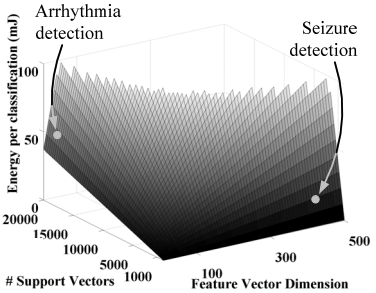


Fig. 2. Classification energy scaling with feature vector dimensionality and number of support vectors (using an RBF kernel). Arrhythmia and seizure detection energies are noted.

We thus look at how the computation can be restructured to overcome energy scaling with the number of support-vectors. Linear kernels, for instance, can overcome this scaling. Linearity in $K(\cdot)$ allows the test vector to be pulled out from the summation, enabling *precomputation* of a single decision vector, w, as shown in Eq. 2. As a result, even when the number of support vectors scales, the energy of linear kernels can remain constant:

$$\begin{aligned} &Let\ sv_i = [sv_{i1}\ sv_{i2} \cdots sv_{iM}]\ and\ x = [x_1\ x_2 \cdots x_M] \\ &\sum_{i=1}^{N} K(sv_i, x)\alpha_i y_i - b \\ &= \sum_{i=1}^{N} (sv_i \cdot x)\,\alpha_i y_i - b \qquad \text{[Linear kernel]} \\ &= \left(\sum_{i=1}^{N} sv_i\,\alpha_i\,y_i\right)\cdot x - b \;=\; w \cdot x - b \end{aligned} \quad (2)$$

N is the number of support vectors and M is the feature-vector dimensionality (once again sv_i are the support vectors, x is a test vector, and α_i, y_i and b are training parameters).

Such restructuring does not directly apply to non-linear kernels, such as RBF (Eq. 1). Fig. 3 shows how kernel energy scales with the number of support vectors, and RBF thus incurs much severer energy than linear kernels. Unfortunately, linear kernels have been shown, both in our experiments (Section IV) and in other reports [8], to be insufficient for typical biomedical applications.

When energy is not a primary concern, RBF kernels have been most prevalent due to their ability to provide a high degree of discrimination. Compared with RBF kernels, polynomial kernels offer only an intermediate level of discrimination, and they have thus been less widely used. As we show in Section IV, however, polynomial kernels can achieve detection performance very close to that of RBF kernels for biomedical applications.

Fig. 3 shows that the energy of a 2nd order polynomial kernel (POLY2) is lower than that of RBF kernels, based on conventional computations. Thus, even though the conventional computation cannot directly exploit linearity to overcome the energy scaling, polynomial kernels offer a good compromise. Their major benefit, however, emerges when we consider the computation restructuring proposed below. As illustrated for the case of a 2nd order polynomial, this allows precomputation in a manner similar to linear kernels, leading to the very desirable constancy in energy.

Computational restructuring of POLY2:

$$\begin{aligned} &Let\ sv_i = [sv_{i1}\ sv_{i2} \cdots sv_{iM}]\ and\ x = [x_1\ x_2 \cdots x_M] \\ &\sum_{i=1}^{N} K(sv_i, x)\alpha_i y_i - b \\ &= \sum_{i=1}^{N} (\beta\, sv_i \cdot x + \gamma)^2 \alpha_i y_i - b \qquad \text{[2nd order polynomial kernel]} \\ &= \sum_{i=1}^{N} \begin{bmatrix}1 & x_1 & x_2 \cdots & x_M\end{bmatrix} \begin{bmatrix}\gamma \\ \beta sv_{i1} \\ \beta sv_{i2} \\ \vdots \\ \beta sv_{iM}\end{bmatrix} \begin{bmatrix}\gamma & \beta sv_{i1} & \beta sv_{i2} \cdots & \beta sv_{iM}\end{bmatrix} \begin{bmatrix}1 \\ x_1 \\ x_2 \\ \vdots \\ x_M\end{bmatrix} \alpha_i y_i - b \\ &= \begin{bmatrix}1\ x_1\ x_2 \cdots x_M\end{bmatrix} \underbrace{\left(\sum_{i=1}^{N} \begin{bmatrix}\gamma \\ \beta sv_{i1} \\ \beta sv_{i2} \\ \vdots \\ \beta sv_{iM}\end{bmatrix} \begin{bmatrix}\gamma & \beta sv_{i1} & \beta sv_{i2} \cdots & \beta sv_{iM}\end{bmatrix} \alpha_i y_i \right)}_{Precomputed\ (M+1\ x\ M+1\ matrix)} \begin{bmatrix}1 \\ x_1 \\ x_2 \\ \vdots \\ x_M\end{bmatrix} - b \end{aligned} \quad (3)$$

(N is the number of support vectors, M is the feature vector dimensionality, sv_i are the support vectors, x is a test vector, and α_i, y_i, β, γ, and b are training parameters).

As with the decision vector precomputation in the case of linear kernels, computational restructuring allows precomputation of a *decision matrix* for polynomial kernels (approximately MxM for POLY2). Namely, the sum of squared inner vector products is converted to a single vector-matrix-vector multiplication. As shown in Fig. 3, computational restructuring thus overcomes energy scaling with the number of support vectors (N). The cost, however, is that more severe energy scaling is introduced with respect to the number of features (since the matrix now scales with M in two dimensions). Fig. 4 shows how energy scales with

the number of features. The resulting energy scaling is quadratic with the computational restructuring of Eq. 3. As discussed in Section V, however, (and as reported previously [9]) feature scaling is much more manageable in typical biomedical applications.

Consequently, polynomial kernels greatly enrich the trade-offs that are available for implementing low-energy classifiers. The extent to which they can be exploited, however, depends on whether the intermediate discrimination they provide leads to sufficient accuracy in the biomedical applications concerned.

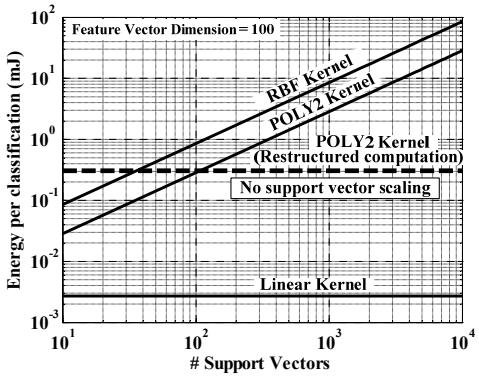


Fig. 3. Classification energy vs. number of support vectors for various kernels (based on energy profiling using MSPsim). Although conventional POLY2 computation involves energy scaling, the computation can be restructured to overcome this.

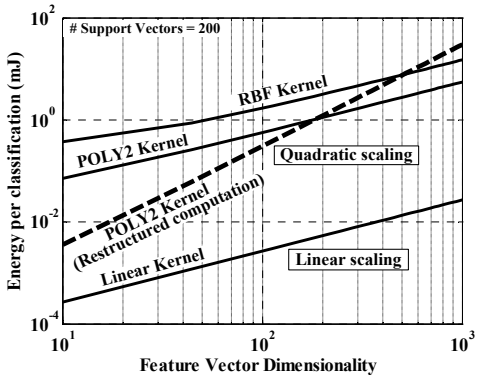


Fig. 4. Classification energy vs. feature vector dimensionality. Computational restructuring introduces more severe scaling.

IV. KERNEL PERFORMANCE AND LOW-ENERGY IMPLEMENTATION METHODOLOGY

This section investigates the performance of polynomial kernels in biomedical applications and suggests a methodology for low-energy classifier implementation.

The algorithms of Fig. 1 are implemented (using SVM-Light [10]), and their performance is characterized for various SVM kernels. Arrhythmia detection uses the physician annotated MIT-BIH ECG database [1], and seizure detection uses the physician annotated CHB-MIT scalp EEG database [1, 2].

Table 2 shows the results for arrhythmia detection, which involves patient-generic training, and thus the results are shown over all the patients. Although the linear kernel achieves insufficient performance, the POLY2 kernel achieves very similar performance to the RBF kernel.

Table 2. Arrhythmia detection performance using various kernels. Positive class corresponds to abnormal heartbeats.

	True Pos.	True Neg.	False Pos.
Linear	57.1%	90.6%	30.4%
POLY2	74.6%	89.0%	28.1%
RBF	75.9%	90.3%	25.4%

Table 3 shows the results for seizure detection, which involves patient-specific training, and thus results are shown for each patient separately. In Table 3, the kernel that yields the best performance for each patient is selected. Where the performance is similar, priority is given in order to linear, polynomial, and then RBF kernels. As shown, polynomial kernels can be used extensively, and even linear kernels can be used in some cases. In total only 5 out of 22 cases require RBF kernels. In two cases, POLY4 kernels are also selected. Similar computational restructuring can be applied to these, but with more severe energy scaling with the number of features.

Table 3. Seizure detection performance with various kernels. Patient results are compared against an RBF kernel (RBF performance is given in square brackets). The lowest order kernel exhibiting comparable performance to RBF is selected.

Patient #	Selected Kernel	Sensitivity (#detected / #total)	Latency (sec)	False Positive (# per day)
1	Linear	7/7 [7/7]	4.29 [2.00]	0.00 [1.19]
2	POLY2	3/3 [3/3]	5.67 [5.67]	1.36 [0.00]
3	POLY4	7/7 [7/7]	1.86 [1.86]	0.63 [0.63]
4	Linear	5/5 [4/5]	3.80 [3.00]	0.62 [0.00]
5	RBF	10/10 [10/10]	3.40 [3.40]	2.52 [2.52]
6	RBF	2/2 [2/2]	12.5 [12.5]	0.38 [0.38]
7	POLY2	5/5 [5/5]	4.20 [4.20]	2.40 [1.20]
8	RBF	4/4 [4/4]	5.50 [5.50]	0.00 [0.00]
9	POLY4	7/7 [7/7]	4.29 [1.58]	0.48 [0.96]
10	POLY2	3/3 [3/3]	1.67 [2.00]	0.69 [2.07]
11	POLY2	27/27 [27/27]	6.37 [4.82]	11.19 [10.17]
12	Linear	11/12 [11/12]	9.19 [9.55]	7.29 [8.02]
13	Linear	8/8 [8/8]	3.63 [3.13]	0.00 [0.93]
14	POLY2	20/20 [20/20]	8.95 [8.60]	9.02 [9.62]
15	Linear	9/10 [9/10]	2.67 [2.23]	5.05 [5.05]
16	POLY2	2/3 [2/3]	6.50 [5.50]	4.59 [4.59]
17	POLY2	5/6 [6/6]	6.80 [11.34]	2.70 [2.03]
18	POLY2	3/3 [3/3]	4.67 [6.34]	0.00 [0.00]
19	RBF	8/8 [8/8]	1.50 [1.50]	0.87 [0.87]
20	POLY2	4/4 [4/4]	1.75 [2.50]	2.94 [2.20]
21	RBF	3/3 [3/3]	9.67 [9.67]	1.55 [1.55]
22	POLY2	7/7 [7/7]	1.58 [2.29]	8.15 [7.25]
Average		**97.6% [97.6%]**	**5.26 [4.95]**	**2.53 [2.53]**

The energy trade-offs presented in Section III and the performance evaluation presented here, can be used as a classifier implementation methodology. The least computationally intensive kernel should be used, and polynomial kernels, which frequently provide a good compromise in discrimination ability, can selectively exploit computational restructuring to yield the most favorable energy trade-off for the application.

V. ENERGY SAVINGS

This section illustrates the energy savings that result in the applications by taking advantage of the approach proposed, particularly thanks to the computational restructuring of Section III.

Arrhythmia detection energy. Fig. 5 shows the energy improvements for arrhythmia detection. The POLY2 kernel requires 71% more support vectors than the RBF kernel, but its energy is still lower by 2.36x. The energy savings thanks to computational restructuring, however, are the most advantageous, corresponding to a further 1114.5x. These savings result due to the large number of support vectors and relatively low number of features in this application.

Seizure detection energy. The possibility of high feature vector dimensionality in seizure detection opposes the benefits of computational restructuring. In [9], however, it is shown that significant dimensionality reduction is possible without degrading performance by selecting the most relevant EEG channels.

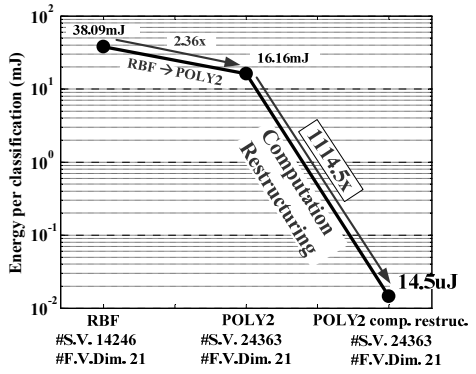
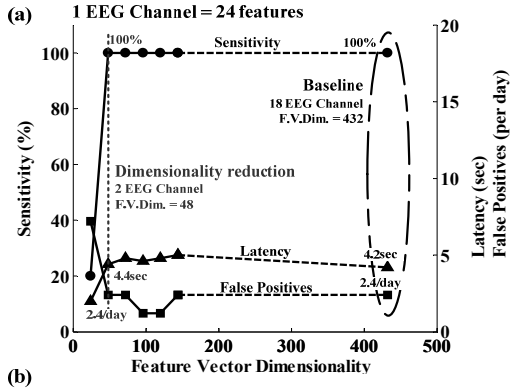


Fig. 5. Energy savings for arrhythmia detection. Computational restructuring yields 1114.5x energy savings for the POLY2 kernel due to a large number of support vectors.

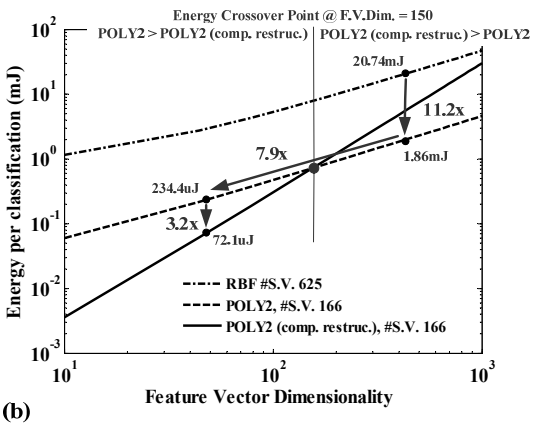


(b)

# EEG Ch.	# F.V.Dim.	Sensitivity	Latency	False Positives
18	432	100%	6.37sec	11.19 /day
4	96	96.30%	7.77sec	12.20 /day

Fig. 6. EEG channel selection for (a) Patient 7 and (b) Patient 11, whose results are summarized for simplicity.

(a)



(b)

Kernel	# EEG Ch.	#S.V.	Energy	Total Energy Saving
RBF	18	225	7.47 mJ	N/A
POLY2	18	195	2.34 mJ	3.2x
POLY2	4	227	620.7 uJ	3.8x
POLY2 CR	4	227	282.7 uJ	2.2x

Fig. 7. Energy savings of seizure detection for two patients: (a) Patient 7. (b) Patient 11. Only table is shown for simplicity.

To illustrate this, Patients 7 and 11 are considered. The most relevant channels are selected by first choosing the best performing single channel out of 18, and then successively adding a channel from the remaining set that yields the largest improvement. The process converges when minimal further improvement is observed. Fig. 6 illustrates the results, showing that two channels are required for Patient 7 and four channels are required for Patient 11. This leads to a total of 48 features and 96 features, respectively (since each EEG channel yields 24 features).

Fig. 7 shows the resulting energy savings for seizure detection. For Patient 7 (Fig. 7a), 11.2x energy savings result from using an 18 channel POLY2 kernel instead of an 18 channel RBF kernel. The energy savings here are in part due to the fewer support vectors required for the POLY2 kernel. In any case, further energy savings of approximately 7.9x result from channel down-selection to two. The resulting feature vector dimensionality makes computational restructuring compelling (with an energy cross over point at 150 dimensions). As a result, computational restructuring yields a further 3.2x energy savings. Corresponding numbers for Patient 11 are summarized in the table of Fig. 7b. Consequently, the use of POLY2 kernels and computational restructuring saves 36.3x and 7.0x energy, respectively, for the two patients.

VI. CONCLUSION

Data-driven models are emerging as a powerful means to detect specific physiological states of interest from patient signals. Their applicability in ultra-low-energy implantable and wearable systems, however, is limited by the computational complexity of applying the models involved. This paper analyzes the energy trade-offs introduced by machine learning SVM classification kernels. It proposes a method of structuring the computation of polynomial kernels to overcome energy scaling with number of support vectors. The efficacy of polynomial kernels in actual applications is demonstrated, and a methodology for low-energy kernel implementation is thus developed. The approach leads to 2627x energy reduction in a cardiac monitoring application and up to 36x energy reduction in a neurological monitoring application.

ACKNOWLEDGMENTS

The authors thank Dr. A. Shoeb (MGH, MIT) for valuable discussions and algorithm testing support. They also acknowledge the support of the Gigascale Systems Research Center, one of six research centers funded under the Focus Center Research Program (FCRP), a Semiconductor Research Corporation entity.

REFERENCES

[1] PhysioNet. *http://www.physionet.org*

[2] A.L. Goldberger, et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, 101(23):e215-e220

[3] G. Meyfroidt et al., "Machine learning techniques to examine large patient databases," *Best Practices & Research Clinical Anaestheiology,* vol. 23, no. 1, pp.127-143, March 2009.

[4] MSPsim, *http://www.sics.se/project/mspsim*

[5] P. de Chazal, et al., "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Transactions on Biomedical Engineering*, vol.51, no.7, pp.1196-1206, July 2004

[6] A. Shoeb and J. Guttag, "Application of machine learning to epileptic seizure detection," *Proc. of Int. Conf. on Machine Learning*, June 2010.

[7] N. Verma, et al., "A micro-power EEG acquisition SoC with integrated feature extraction processor for a chronic seizure detection system," *IEEE Journal of Solid-State Circuits*, vol.45, no.4, pp.804-816, April 2010

[8] A. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation. MIT, September 2009

[9] E. Shih and J. Guttag, "Reducing energy consumption of multi-channel mobile medical monitoring algorithms," in *Proceedings of the Second International Workshop on Systems and Networking Support for Healthcare and Assisted Living Environments*, June 2008

[10] SVM Light, *http:// http://www.cs.cornell.edu/People/tj/svm_light*