J Sign Process Syst
DOI 10.1007/s11265-012-0672-8

Low-energy Formulations of Support Vector Machine Kernel Functions for Biomedical Sensor Applications

Kyong Ho Lee · Sun-Yuan Kung · Naveen Verma

Received: 31 October 2011 / Revised: 27 March 2012 / Accepted: 29 March 2012

Abstract Although physiologically-indicative signals can be acquired in low-power biomedical sensors, their accurate analysis imposes several challenges. Data-driven techniques, based on supervised machine-learning methods provide powerful capabilities for potentially overcoming these, but the computational energy is typically too severe for low-power devices. We present a formulation for the kernel function of a support-vector machine classifier that can substantially reduce the real-time computations involved. The formulation applies to kernel functions employing polynomial transformations. Using two representative biomedical applications (EEG-based seizure detection and ECG-based arrhythmia detection) employing clinical patient data, we show that the polynomial transformation yields accuracy performance comparable to the most powerful available transformation (i.e., the radial-basis function), and the proposed formulation reduces the energy by over 2500× in the arrhythmia detector and 9.3-198× in the seizure detector (depending on the patient).

Keywords Kernel-energy trade-off · Energy efficiency · Machine learning · Biomedical devices

1 Introduction

The recent emergence of low-power ambulatory recording technologies for biomedical signals [1–3] raises unprecedented opportunities for devices to add substantial clinical value. Although the signals available through such sensors can potentially provide rich physiological information, the processes through which they manifest are extremely complex and subject to numerous interference and noise sources [4]. Many compelling applications are emerging, however, that require the ability to interpret specific physiological states in real-time; closed-loop (i.e., responsive) therapeutic and prosthetic devices [5, 6] and intelligent devices for chronic monitoring are important examples [7, 8]. As we describe in this paper, high-order signal models are required to accurately detect the specific states of interest, but the application of these models poses severe energy constraints in small-scale biomedical sensors.

Specifically, Fig. 1 illustrates the complexities typically encountered with physiological signals. The signals shown correspond to the electroencephalograph (EEG) of an epileptic patient. The EEG is a voltage signal, typically of amplitude less than 100 μV, that is available on the surface of the scalp. It originates due to electrophysiological activity of neuronal cells in the brain and thus exhibits correlations with seizures. However, several challenges are apparent. First, Fig. 1a shows that the signals are subject to numerous physiologic variances.

The authors thank Dr. A. Shoeb (MGH, MIT, now with WeatherBill) 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.

K. H. Lee (✉) · S.-Y. Kung · N. Verma
Princeton University, Engineering Quadrangle, Olden Street, Princeton, NJ 08544, USA
e-mail: kyonglee@princeton.edu

Published online: 22 April 2012

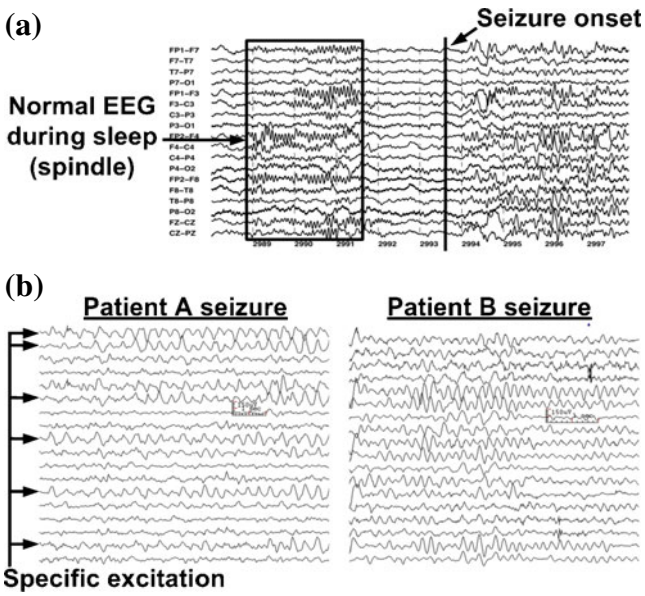

Figure 1 **a** EEG signals from a patient showing normal activity (during sleep) and seizure activity, which is difficult to discriminate. **b** EEG signals corresponding to seizures from two different patients, showing different excitation characteristics.

A transient burst is highlighted that corresponds to normal EEG activity during sleep (called spindles). Following this, a burst associated with a clinical seizure is labeled. Though distinct from the background, the challenge is that the two bursts are similar, necessitating signal models that are capable of making distinctions with high specificity. Second, Fig. 1b shows that the precise model that is required is variable. Two patient EEGs are shown, each exhibiting different seizure manifestations (both in terms of the affected channels and the excitatory response). Thus, the model must be selective to distinguish between seizure and non-seizure activity, but it must also be flexible and efficiently constructed to handle variabilities in seizure manifestations.

Considering these challenges, techniques based on data-driven modeling provide powerful capabilities for analyzing physiological signals [9, 10]. These techniques imply using the physical waveforms to construct the high-order models required. By leveraging modeling frameworks from the domain of supervised machine learning, efficient methods to construct such models with limited data are possible. In chronic devices for biomedical sensing, however, the power constraints are typically in the range of 10–100 μW for implantable devices and 1–10 mW for wearable devices [11]. As we attempt to incorporate these methods into such low-power devices, computational energy emerges as the critical challenge. In this paper, we make the following contributions:

- We perform energy analysis of representative data-driven algorithms for real-time biomedical detection. This identifies the need for high-order signal models as the key limitation.
- We present a new approach for formulating the kernel functions involved that alters the energy characteristics of the computations. This has the potential to dramatically reduce the energy when complex signal models are required.
- We demonstrate the energy savings in two biomedical applications (seizure detection and cardiac-arrhythmia detection) using clinical patient data. We show that the proposed approach retains the performance of the detectors while potentially yielding orders of magnitude energy savings.

2 Computational Energy Analysis

In this section, we first present details regarding the algorithmic framework that is considered. Following this, we present an energy analysis, using the seizure- and arrhythmia-detection applications as a case study.

2.1 Background on the Algorithmic Framework

A block diagram of a typical data-driven algorithm is shown in Fig. 2. There are two phases: (1) training and (2) real-time detection. In this work, we focus on the Support Vector Machine (SVM) [12], which is powerful supervised learning framework for realizing such algorithms. SVMs have gained wide-spread popularity due to their computational efficiency, and have proven particularly suitable for medical applications due in part to their robustness in the case of minimal training data [13]. SVMs analyze signal segments by deriving features from the signal and representing these as a vector. During training, feature vectors are assigned class labels, and an algorithm is used to select an optimal set of feature vectors at the boundary of the two classes in order to construct a decision function (as illustrated in Fig. 2b). The set of chosen vectors is called the support-vector model. The process described can be expanded to support multiclass problems [14].

Since the support-vector model can be derived offline, the energy for training is not of primary concern. For real-time detection, however, the support-vector model must be applied in an on-going manner to classify a continuous signal. The energy of real-time detection is thus a primary concern on the sensor platform.

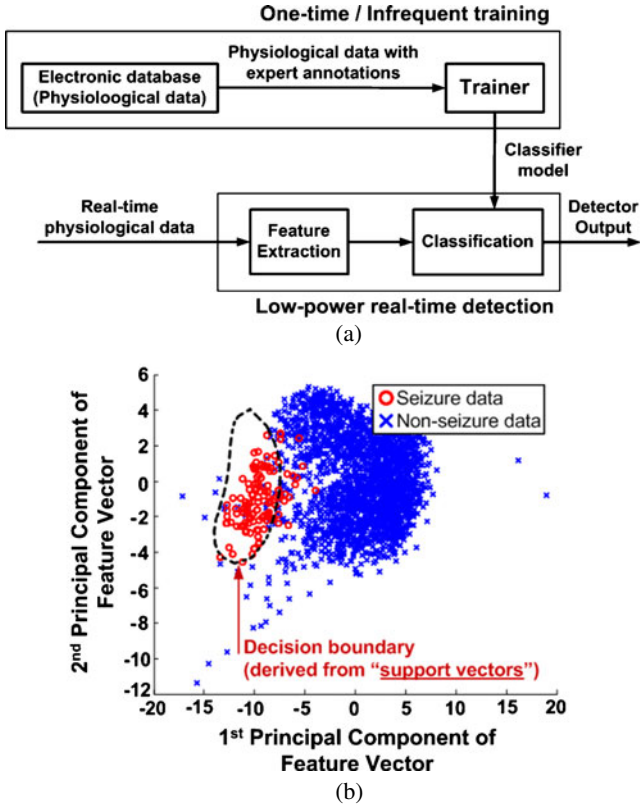

Figure 2 **a** In the SVM framework, training occurs infrequently based on previous observations with expert annotations, and, following this, real-time classification uses the derived model to perform detection. **b** SVM training conceptually involves deriving a decision boundary in the feature space that separates the two classes of training data; the decision boundary is represented by selecting feature vectors occurring at the edges of the data distributions (for illustration, the high-dimensional feature data is plotted in 2D using principal component analysis).

2.2 Application Case Studies

To analyze the energy limitations of detection, two biomedical applications have been considered: (1) arrhythmia detection based on electrocardiogram (ECG) signals and (2) seizure detection based on electroencephalogram (EEG) signals. Figure 3 shows the feature extraction computations required in the two applications. The features correspond to physiological *biomarkers*, i.e., parameters that exhibit some correlation (though the precise correlation may be unknown) with the physiological states we are interested in detecting. In the arrhythmia detector, the features correspond to the morphology of the ECG waveform. These are derived by sampling points as shown around the R-peak. In addition to the waveform samples, the time intervals between R-peaks provide three additional features. The total dimensionality of the resulting feature vector is 21 [15].

In the seizure detector, the features correspond to the spectral energies over a two-second window. These are obtained by applying 8 band-pass filters with different center frequencies to each EEG channel. Each filter is then followed by an accumulator. Up to 18 EEG channels can be used, giving a feature-vector dimensionality as high as 432 [16].

After deriving the support-vector model (**sv**) from the training phase, the feature vector (**x**) for the incoming data is classified based on the computed sign of the decision function shown in Eq. 1:

Let $f(x)$ denote the decision function

$$f(x) = \sum_{i=1}^{N} K(\mathbf{sv_i}, \mathbf{x})\alpha_i y_i - b$$

$$= \sum_{i=1}^{N} exp\left(\frac{-||\mathbf{sv_i} - \mathbf{x}||^2}{\sigma^2}\right)\alpha_i y_i - b$$

$$(\text{classification}) = sgn(f(x)) \quad (1)$$

A kernel transformation (K) is commonly used to transform the data to a higher dimensional space, essentially yielding much greater flexibility in the decision boundary [13]. The kernel, using a radial-basis-function (RBF) transformation, is shown (where α_i, y_i, and b are training parameters). The RBF provides very high flexibility, resulting in substantial performance enhancement. From Eq. 1, it can be seen that both

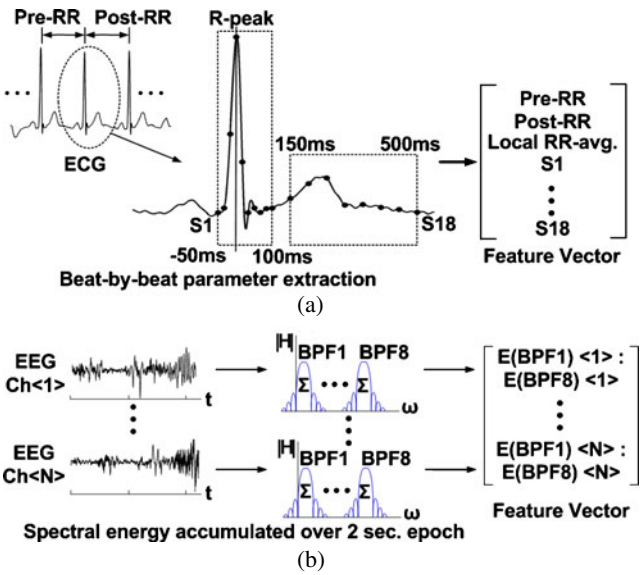

Figure 3 **a** Feature extraction for arrhythmia detection using ECG-morphology features and R-to-R interval features to form a feature vector. **b** Feature extraction for seizure detection using EEG spectral-energy features extracted using band-pass filters followed by accumulators operating over a two-second window.

Table 1 Feature-extraction and classification energy for seizure and arrhythmia detection using a cycle-accurate instruction-set simulator [17].

Application	Feature extraction energy	Classification energy	Ratio
Seizure detection	1.44 mJ/FV (5.3 Mcycles)	26.98 mJ/FV (99 Mcycles)	18.7
Arrhythmia detection	1.56 mJ/FV (5.7 Mcycles)	49.52 mJ/FV (181 Mcycles)	31.7

Classification energy dominates over feature extraction for the two representative applications. Cycle counts from the simulator are denoted in parenthesis (272pJ/cycle is assumed from lab measurements for energy calculation).

the number of support vectors (N) and the feature-vector dimensionality (due to the vector operations) will impact the computational energy.

Table 1 shows energy profiling results for the applications considered. The results are derived from cycle-accurate instruction-set simulation of a low-power microprocessor (MSP430) [17]. A key observation is that the classification energy dominates over that of feature extraction by a factor of nearly twenty and over thirty for the respective applications. This motivates the need for approaches that substantially reduce classification energy. In particular, the surface plot in Fig. 4 shows how the classification energy scales with respect to the number of support vectors and the vector dimensionality. These quantities are representative of the algorithmic complexity of an application, and, in particular, the number of support vectors reflects the complexity of the model involved. As shown in the figure, the applications chosen stress the two quantities respectively. The arrhythmia detector requires over 15k support vectors, while the seizure detector requires as many as 600, but has a high feature-vector dimensionality (of 432).

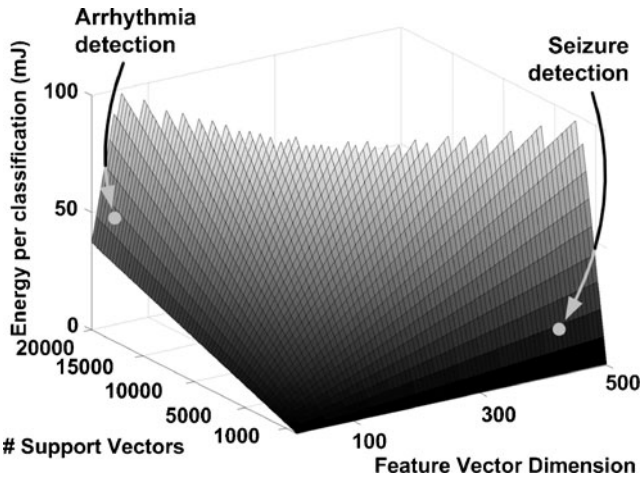

Figure 4 Surface plot showing how the classification energy scales with the number of support vectors and the feature-vector dimensionality. The chosen applications stress both of these parameters

3 Energy-driven Kernel-function Formulation

Given the severe classification energy observed in Section 2.2, our focus in this section is to find alternate formulations of the kernel function that lower energy while maintaining classifier performance. It is worth noting, that approaches have been suggested for reducing the feature-vector dimensionality [18–20] and reducing the number of support vectors [21], providing a potential means for overcoming the energy limitations seen in Fig. 4. In fact, several highly-effective application-specific and generic approaches for vector dimensionality reduction have been reported [18, 19]; as an example, such approaches are considered in the analysis of Section 4.2. The methods for support vector reduction, however, typically lead to corresponding degradation in classifier performance [8, 21]; the extent to which such methods can be applied is thus limited. This section presents alternate kernel-function formulations that address the energy scaling exhibited with respect to the number of support vectors. In fact, what results is a new kernel-energy trade space involving performance, support-vector scaling, and dimensionality scaling. As a result, the powerful dimensionality reduction methods can thus be leveraged to also manage the large number of support vectors that are required in many applications. This section starts by discussing the need for non-linear transformations in the kernel function, which is what leads to the problematic energy-scaling. Then, the new kernel-function formulation is presented. Last, the resulting energy tradeoffs are analyzed.

3.1 Need for Non-linear Kernel Functions

As previously mentioned, the transformation function (K) enhances the flexibility of the kernel by introducing the possibility of non-linearity, which effectively allows the data to be represented in a higher dimensional space. In fact, this non-linearity is precisely the source of the problematic energy scaling exhibited with respect to the number of support vectors. If a linear function is chosen for K, the input feature vector can be factored out of the kernel function summation, allowing the summation to be precomputed over all of the support

vectors into a single decision vector ($\mathbf{w}$), as shown in Eq. 2:

$$\begin{aligned} f(x) &= \sum_{i=1}^{N} K(\mathbf{sv_i}, \mathbf{x})\alpha_i y_i - b \\ &= \sum_{i=1}^{N} (\mathbf{sv_i} \cdot \mathbf{x})\alpha_i y_i - b \\ &= \left(\sum_{i=1}^{N} \mathbf{sv_i}\alpha_i y_i \right) \cdot \mathbf{x} - b \\ &= \mathbf{w} \cdot \mathbf{x} - b \end{aligned} \tag{2}$$

The problem with the linear kernel is that it does not yield sufficient classifier flexibility. In fact, both in the experiments of Section 4 and in previous reports [22], linear kernels are shown to be insufficient for biomedical detector applications. As an example, Fig. 5 illustrates the limitation using clinical data for a seizure detector employing intracranial EEG. In the two-dimensional feature space, the linear decision function results in the misclassification of several non-seizure points, implying a high rate of false positives. On the other hand, the RBF decision function results in superior performance. In biomedical sensing, the complex correlations between biomarkers and the physiological states of interest typically require the flexibility afforded by non-linear transformation functions.

The high degree of flexibility offered by the RBF kernel has made it a popular choice across applications. We, however, turn our focus to the *polynomial* kernel, which provides non-linearity for an intermediate level of flexibility. Because they are not as strong as the RBF kernel, polynomial kernels have been far less frequently used, particularly in medical applications. However, as we show in this section, they can provide an opportunity to overcome support-vector energy scaling through precomputation, similar to the linear kernel, but while also providing much higher flexibility. RBF kernels, due to their use of an exponential function, do not offer such an opportunity. As a result, polynomial kernels can introduce a valuable benefit when computational energy is a critical concern.

$$\begin{aligned} f(x) &= \sum_{i=1}^{N} K(\mathbf{sv_i}, \mathbf{x})\alpha_i y_i - b \\ &= \sum_{i=1}^{N} (\beta \mathbf{sv_i} \cdot \mathbf{x} + \gamma)^2 \alpha_i y_i - b \\ &= \sum_{i=1}^{N} \begin{bmatrix} 1 & x_1 & x_2 & \dots & x_M \end{bmatrix} \\ &\quad \times \begin{bmatrix} \gamma \\ \beta sv_{i1} \\ \beta sv_{i2} \\ \vdots \\ \beta sv_{iM} \end{bmatrix} \begin{bmatrix} \gamma & \beta sv_{i1} & \beta sv_{i2} & \dots & \beta sv_{iM} \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_M \end{bmatrix} \alpha_i y_i - b \\ &= \sum_{i=1}^{N} \mathbf{X}^T \mathbf{S_i} \mathbf{S_i}^T \mathbf{X} \alpha_i y_i - b \\ &= \mathbf{X}^T \left(\sum_{i=1}^{N} \mathbf{S_i} \mathbf{S_i}^T \alpha_i y_i \right) \mathbf{X} - b \end{aligned} \tag{3}$$

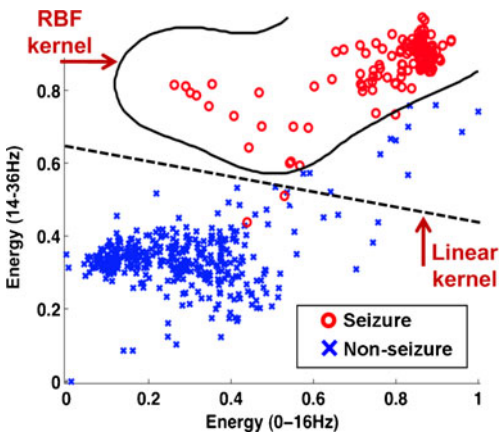

Figure 5 Feature-space plot using intracranial EEG data, showing the effectiveness of the linear and RBF kernels. The linear kernel misclassifies many non-seizure points whereas the RBF kernel yields high flexibility in the decision boundary, leading to much better performance.

3.2 Kernel Formulation to Overcome Energy Scaling

Equation 3 shows how a 2nd-order polynomial kernel can be reformulated to permit precomputation over the support vectors. The dot product and squaring computations in the polynomial kernel can be rewritten in vector multiplication form. The input vector, $\mathbf{x}$, is used to form a new vector, $\mathbf{X}$, and each support vector is used to form a new vector, $\mathbf{S}_i$, both having one additional constant term. $\mathbf{X}$ can then be factored out of the summation, yielding a vector-matrix-vector multiplication between $\mathbf{X}$ and a new *decision matrix*. The decision matrix corresponds to the sum of matrices formed by multiplying $\mathbf{S}_i$ with its transpose.

With this formulation, energy scaling with respect to the number of support vectors is overcome since the support-vector decision matrix can be precomputed. However, because the feature-vector computations are

converted to matrix computations, the number of arithmetic operations required scales more severely with the dimensionality of the vectors. Roughly speaking, the scaling is quadratic rather than linear. The same reformulation can be applied to higher even-ordered polynomial kernel functions, with correspondingly severe scaling due to feature-vector dimensionality.

3.3 Kernel-energy Tradeoffs

As discussed in Section 4, polynomial functions are effective in a broad range of applications. The new energy tradeoffs offered by the formulation of Eq. 3 can thus potentially be broadly exploited. This section analyzes the trade space that results thanks to the various kernel functions and their formulation options.

Based on instruction-set simulation of an MSP430 microprocessor, Fig. 6a shows how the energy scales with respect to the number of support vectors (for illustration a feature-vector dimensionality of 100 is assumed). With linear kernels, the formulation of Eq. 2 permits constant energy. As mentioned previously, and as shown quantitatively in Section 4, these kernels may not provide sufficient flexibility for performance comparable to the RBF kernel. Both the conventional RBF and polynomial kernels, however, exhibit the energy scaling shown. Utilizing the formulation of Eq. 3, this energy scaling is overcome. Of course, it remains to be seen (Section 4) whether the polynomial kernel leads to sufficient performance.

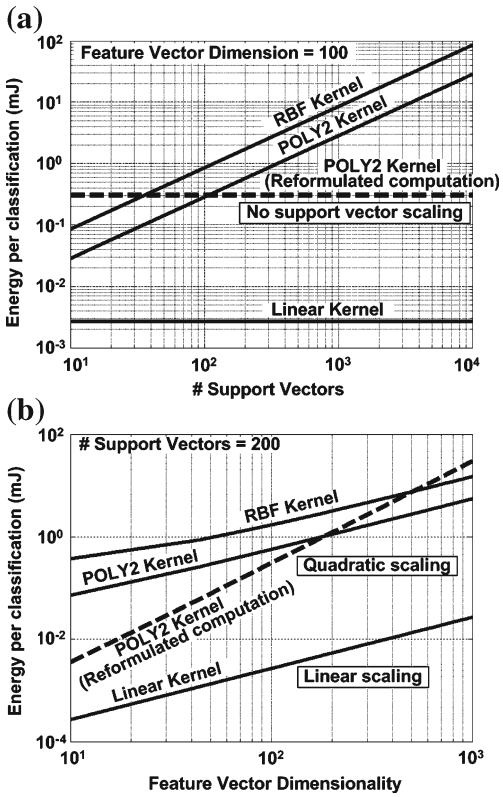

Figure 6 **a** Classification energy with respect to the number of support vectors. Computational reformulation of the polynomial kernel overcomes energy scaling. **b** Classification energy with respect to the feature-vector dimensionality. The quadratic scaling of the polynomial kernel with computational reformulation is due to conversion of vector multiplications into matrix multiplications.

Although the proposed formulation overcomes energy scaling in Fig. 6a, the y-intercept is higher than that of the conventional formulation. As mentioned, this results due to more severe energy scaling with the feature-vector dimensionality. Figure 6b shows the corresponding energy-scaling trend (for illustration, 200 support vectors are assumed). While the conventional formulations exhibit energy that scales linearly with feature-vector dimensionality, the proposed formulation exhibits energy that scales quadratically for a 2nd-order polynomial kernel, due to the matrix computations involved.

The tradeoffs suggest that applications requiring a large number of support vectors but a modest feature-vector dimensionality can substantially benefit from the formulation proposed. As discussed in Section 4.2, the availability of various methods for reducing the number of features makes this tradeoff potentially of high value. As mentioned, however, this tradeoff is only available for polynomial kernel functions that permit the formulation of Eq. 3. In the next section, the performance of the clinical applications using a polynomial kernel is validated.

4 Experimental Analysis

As described in Section 2.2, we have considered two representative biomedical applications for experimental analysis: arrhythmia detection based on clinical ECG data and seizure detection based on clinical EEG data. As shown in Fig. 4, arrhythmia detection requires a very large number of support vectors. On the other hand, seizure detection needs a modest number of support vectors, but the feature-vector dimensionality is high. In this section, we apply effective techniques including [18] to reduce the feature vector dimensionality so that the trade-off introduced by computational reformulation is highly beneficial.

4.1 Arrhythmia detection

The arrhythmia detector is based on the feature-extraction computations shown in Fig. 3a. The computations are implemented in Matlab, and the open source SVM-Light package is used to implement the classifier [23]. Clinical data for the experiments is obtained from the MIT-BIH arrhythmia database, PhysioNet http://www.physionet.org, which contains cardiologist-labeled ECG data over 46 patients.

Performance analysis The performance requirements depend on the clinical application, where the application could range from closed-loop stimulation of a therapeutic device to clinician monitoring of patient response to a medication plan. Since high performance is desired in all of these applications even though they have varied sensitivity and specificity requirements, our aim is to benchmark performance against the current state-of-the-art detectors. This corresponds to the RBF detectors described.

Table 2 shows the performance using three different kernel functions. The metrics used for this application are the true-positive rate (TPR), the true-negative rate (TNR), and the false positive rate (FPR). The linear kernel results in substantially lower performance, while the polynomial and RBF kernels result in comparable performance. Consequently, the formulation of Eq. 3 is a viable option.

Energy analysis The number of support vectors is also shown in Table 2. In all cases, a large number is required, and the polynomial kernel actually results in substantially more than the other two. As a result, the formulation of Eq. 3 leads to significant energy savings. Figure 7 first shows the energy savings going from the RBF to the conventional 2nd-order polynomial kernel; despite the higher number of support vectors, 2.36× energy savings are achieved thanks to the simpler transformation function. With the new formulation, however, over 1100× energy savings are achieved thanks to the ability to precompute the support-vector decision matrix. The energy numbers are based on instruction-set simulation of the MSP430 microprocessor.

Table 2 Performance of the arrhythmia detector with different kernels.

SVM Kernel	TPR	TNR	FPR	# of S.V.
Linear	57.1%	90.6%	30.4%	16066
Poly (2nd order)	74.6%	89.0%	28.1%	24363
RBF	75.9%	90.3%	25.4%	14246

The linear kernel shows poor performance whereas the polynomial kernel shows similar performance to the RBF kernel. (TPR = true positive rate, TNR = true negative rate, FPR = false positive rate).

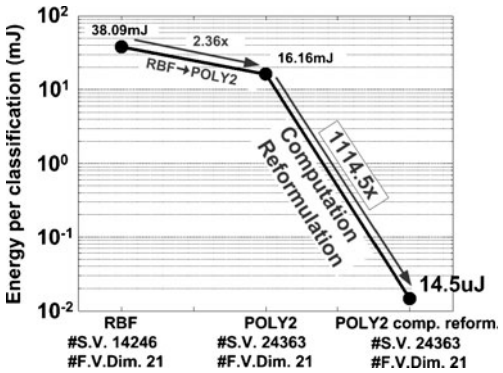

Figure 7 Energy saving of arrhythmia detection. Because of the large number of support vectors, the proposed polynomial kernel formulation saves energy by 1114.5×, leading to the total energy saving of 2627×.

4.2 Seizure Detection

The seizure detector is based on the feature-extraction computations shown in Fig. 3b. Clinical data for the experiments is obtained from the CHB-MIT seizure database (PhysioNet http://www.physionet.org, [24]), which contains neurologist-labeled EEG data over 22 patients. The algorithm used relies on constructing a patient-specific support-vector model. Results are thus shown individually for the 22 patients.

Performance analysis Table 3 shows the performance. The metrics used for this application are the sensitivity, latency, and rate of false alarms [16]. The linear, polynomial, or RBF kernel function is chosen in decreasing order of preference, such that minimal performance degradation is suffered for each patient. The chosen kernel is noted, and the performance of the RBF kernel is shown in brackets for comparison. For most patients, the polynomial kernel is sufficient, and in some cases, even a linear kernel can be used. In only five of the cases an RBF kernel is required.

Feature reduction via channel selection Because of the potentially large feature-vector dimensionality in this application, the formulation of Eq. 3 may not be directly viable. As mentioned, however, several methods for feature reduction are available. As an example, in [18] it has been shown that seizure detection can potentially require only a few EEG channels, since seizures typically originate in a particular region of the brain. For illustration, we select the channels by incrementally adding one channel at a time that results in the largest performance improvement, until performance similar

Table 3 Performance of seizure detection. Performance for each patient is reported with the average performance at the bottom. The performance of the RBF kernel is also shown in brackets for comparison. Poly4 (4th-order polynomial kernel) is used for two patients. Poly4 uses a inner-product computation similar to poly2, but it requires creating a four-dimensional matrix rather than the two-dimensional matrix of Eq. 3 to enable the precomputation.

Patient #	Kernel	Sensitivity (#detected/#total)	Latency (sec)	Rate of false alarms (# 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]	4.57 [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]
Overall average		97.6% [97.6%]	5.26 [4.95]	2.53 [2.53]

to the full 18-channel case is achieved. Figure 8 shows the results of such a process for one patient (Patient 7). In this case, only 2 EEG channels are required to achieve the full-channel performance. A similar process is applied for other patients, and the results are shown in Table 4.

Feature reduction via principal component analysis We apply another powerful yet generic technique for feature reduction, principal component analysis (PCA), to reduce the feature-vector dimensionality even further. PCA is a procedure that extracts uncorrelated components in the data and projects the data onto the uncorrelated basis. Figure 9 illustrates the concept using seizure data, where two-dimensional vectors are

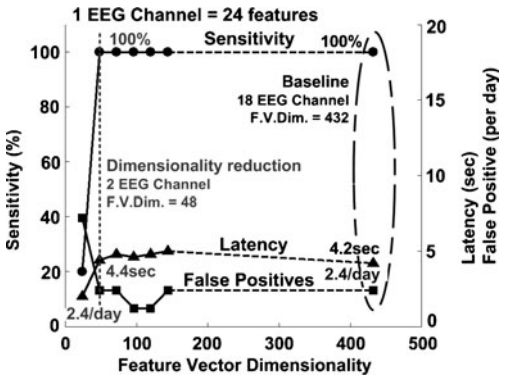

Figure 8 Channel selection in a seizure detector (for patient 7). With only two EEG channels, performance comparable to the full 18 EEG channel case is achieved.

Table 4 Channel selection and PCA of for feature reduction in the seizure detection algorithm.

F.V.Dim. (# EEG Ch)	Sensitivity	Latency	False positive
Patient 7			
432 (18)	100%	4.20 sec	2.4/day
48 (2)	100%	4.40 sec	2.4/day
20 (4)	100%	4.60 sec	3.6/day
Patient 11			
432 (18)	100%	6.37 sec	11.2/day
96 (4)	96.3%	7.77 sec	12.2/day
20 (4)	92.6%	7.80 sec	11.2/day
Patient 14			
432 (18)	100%	8.95 sec	9.02/day
72 (3)	90.0%	9.50 sec	10.23/day
20 (3)	90.0%	7.39 sec	8.42/day
Patient 20			
432 (18)	100%	1.75 sec	2.94/day
96 (4)	100%	2.50 sec	2.94/day
70 (4)	100%	2.50 sec	4.40/day

The number of channels are significantly reduced while the performance degradation is minimal, and further feature vector reduction is achieved using PCA. The feature vector dimensionality (F.V.Dim) and the corresponding number of EEG channels is shown in parenthesis. In each patient, the first line is the baseline performance with the full 18 channels, the second line is the performance after channel selection, and the third line is the performance after also applying PCA.

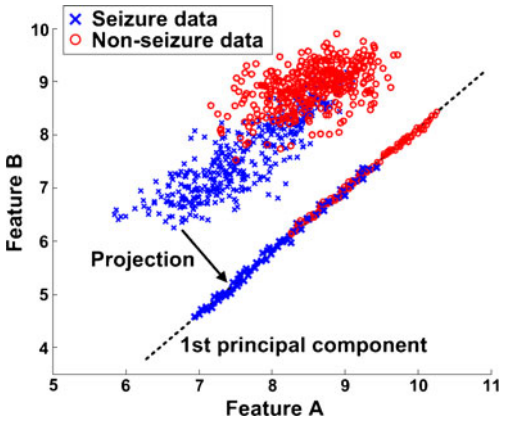

Figure 9 Illustration of principal component analysis (PCA) using two features from EEG data (feature A and B). PCA finds a principal component that data show the most variance and projects the data on the principal component. PCA, thus, returns n principal components from n-dimensional features in the order of decreasing signal variances.

projected onto a single dimension. The feature-vector dimensionality can thus be reduced with minimal performance degradation by selecting m principal components that exhibit the strongest correlations.

We applied PCA to the patient data after channel selection. The resulting detector performance is shown in Table 4. Most patients show comparable performance even with substantial feature reduction. Patient 14 shows improved performance after PCA, implying some noise rejection through the process. Patient 20, on the other hand, benefits from only modest feature reduction, implying that most of the original features contain important information for seizure detection.

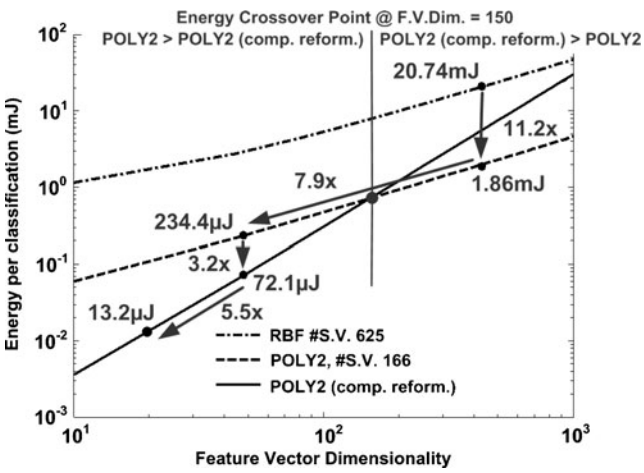

Figure 10 Energy savings for Patient 7 in the seizure detection application. Directly applying computational reformulation on the full 18-EEG channel detector does not save energy; after channel selection and PCA for feature reduction, going from the RBF kernel to the polynomial kernel and applying computational reformulation saves energy by a factor of over 198× (Corresponding energy saving results are shown in Table 5; channel selection yields 7.9× energy savings in addition to the approaches listed in Table 5).

Energy analysis Having substantially reduced the feature-vector dimensionality, the formulation of Eq. 3 becomes extremely advantageous. For example, the energy savings for Patient 7 (based on instruction-set simulation of the MSP430 microprocessor) are shown in Fig. 10. Going from an RBF kernel to the polynomial kernel saves energy by 11.2×. The formulation of Eq. 3 is not viable at this point due to the large number of features. After channel reduction, however, the formulation saves an additional 3.2× in energy (giving energy savings of 36.3× by combining the use of the polynomial kernel with the ability to precompute the support-vector decision matrix). By applying PCA to the computation-reformulated classifier, the computational energy is reduced by a further 5.5×, giving total energy savings of 198.5×. The energy savings for the other patients are shown in Table 5; after applying the feature reduction techniques, computational reformulation yields 9.3–198.5× energy savings.

Table 5 Energy saving in seizure detection.

Kernel	Energy/FV	Energy savings
Patient 7		
RBF Ch. Sel.	2.62 mJ	–
Poly2 Ch. Sel.	234.4 µJ	11.2×
Comp. Reform.	72.1 µJ	3.3×
PCA	13.2 µJ	5.5×
Total saving		198.5×
Patient 11		
RBF Ch. Sel.	1.66 mJ	–
Poly2 Ch. Sel.	620.7 µJ	2.7×
Comp. Reform.	282.7 µJ	2.2×
PCA	13.2 µJ	21.4×
Total saving		125.8×
Patient 14		
RBF Ch. Sel.	2.20 mJ	–
Poly2 Ch. Sel.	760.0 µJ	2.9×
Comp. Reform.	160.0 µJ	4.8×
PCA	13.2 µJ	12.1×
Total saving		166.7×
Patient 20		
RBF Ch. Sel.	1.41 mJ	–
Poly2 Ch. Sel.	323.0 µJ	4.4×
Comp. Reform.	282.0 µJ	1.2×
PCA	151.4 µJ	1.9×
Total saving		9.3×

Corresponding channel reduction and PCA are shown in Table 4.

5 Conclusions

Data-driven techniques based on an SVM provide a powerful framework for analyzing physiological signals in biomedical sensors. Particularly through the use of non-linear transformation functions, the SVM can enable detectors capable of accurately modeling the signals to derive clinically-meaningful outputs. Such non-linearity, however, introduces severe energy challenges when high-order models are involved, since it causes the computational energy of classification to scale with the model complexity. In this paper, we present how the polynomial kernel-transformation function yields an opportunity to overcome this scaling while still retaining non-linearity. As a result, the energy in practical clinical application is reduced by orders of magnitude. We show that the performance that results is comparable to that achieved using a radial-basis function kernel. The formulation presented for polynomial kernels thus offers substantial advantages in small-scale sensors where accurate yet low-energy signal analysis is required.

References

1. Yazicioglu, R. F., et al. (2007). A 60 μW 60 nV/$\sqrt{Hz}$ readout front-end for portable biopotential acquisition systems. *IEEE Journal of Solid-State Circuits, 42*(5), 1100–1110.
2. Verma, N., et al. (2010). A micro-power EEG acquisition SoC with integrated feature extraction processor for a chronic seizure detection system. *IEEE Journal of Solid-State Circuits, 45*,(4), 804–816.
3. Denison, T., et al. (2007). 2 μW 100 nV/$\sqrt{Hz}$ chopper-stabilized instrumentation amplifier for chronic measurement of neural field potentials. *IEEE Journal of Solid-State Circuits, 42*(12), 2934–2945.
4. Verma, N., et al. (2011). Data-driven approaches for computation in intelligent biomedical devices: a case study of EEG monitoring for chronic seizure detection. *Journal of Low Power Electronics and Applications, 1*(1), 150–174.
5. Schachter, S. C. (1998). Vagus nerve stimulator. *Epilepsia, 39*, 677–686.
6. Csavoy, A., et al. (2009). Creating support circuits for the nervous system: considerations for "brain-machine" interfacing. In *VLSI circuits, 2009 symposium on* (pp. 4–7).
7. Dishman, E. (2004). Inventing wellness systems for aging in place. *IEEE Computer, 37*, 34–41.
8. Shoaib, M., et al. (2011). A low-energy computation platform for data-driven biomedical monitoring algorithms. *DAC* (pp. 591–596).
9. Hau, D., & Coiera, E. (1994). Learning qualitative models from physiological signals (Vol. SS-94-01, pp. 67–71). AAAI Technical Report, Menlo Park, CA, U.S.A., AAAI Press.
10. Lucas, P. (2004). Bayesian analysis, pattern analysis and data mining in health care. *Current Opinion in Critical Care, 10*(5), 399–403.
11. Chandrakasan, A., et al. (2008). Ultralow-power electronics for biomedical applications. *Annual Review of Biomedical Engineering, 10*, 247–274.
12. Vapnik, V. N. (1995). *The nature of statistical learning theory*. New York: Springer.
13. Christianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge, U.K.: Cambridge University Press.
14. Hsu, C. W., & Lin, C. J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks, 13*, 415–425.
15. de Chazal, P., et al. (2004). Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering, 51*(7), 1196–1206.
16. Shoeb, A., & Guttag, J.(2010). Application of machine learning to epileptic seizure detection. In *Proc. of int. conf. on machine learning*.
17. Eriksson, J., & Finne, N. MSPsim. Swedish Institute of Computer Science. http://www.sics.se/project/mspsim.
18. Shih, E., & Guttag, J. (2008). 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*.
19. Glassman, E., & Guttag, J. (2006). Reducing the number of channels for an ambulatory patient-specific EEG-based epileptic seizure detector by applying recursive feature elimination. In *Proc. of the 28th IEEE EMBS annual international conference* (pp. 2175–2178).
20. Kung, S. Y., et al. (2010). Feature selection for genomic signal processing: Unsupervised, supervised, and self-supervised scenarios. *Journal of Signal Processing Systems, 61*(1), 3–20.
21. Shoeb, A., et al. (2009). A micropower support vector machine based seizure detection architecture for embedded medical devices. In *Proc. IEEE eng. med. biol. soc. conf.* (pp. 4202–4205).
22. Shoeb, A. (2009). Application of machine learning to epileptic seizure onset detection and treatment. Ph.D. Thesis, MIT.
23. Joachims, T. SVM Light. University of Dortmund. http://www.cs.cornell.edu/People/tj/svm_light.
24. Goldberger, A. L., et al. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiological signals. *Circulation, 101*(23), e215–e220.

Kyong Ho Lee received the B.S. degree in Electrical Engineering from Korea Advanced Institute of Science and Technology (KAIST), Korea in 2004 and the M.S. degree in Electrical Engineering from Stanford University in 2009. He is currently working towards a Ph.D. degree in Electrical Engineering at

Princeton University. His research interest includes energy-efficient signal processing for biomedical applications, and ultra-low energy biomedical circuit design employing machine learning techniques. He is a co-recipient of the 2011 Qualcomm Innovation Fellowship.

Sun-Yuan Kung is a Professor at Department of Electrical Engineering in Princeton University. His research areas include VLSI array processors, system modeling and identification, machine learning, wireless communication, sensor array processing, multimedia signal processing, and genomic signal processing and data mining. He was a founding member of several Technical Committees (TC) of the IEEE Signal Processing Society, and was appointed as the first Associate Editor in VLSI Area (1984) and later the first Associate Editor in Neural Network (1991) for the IEEE Transactions on Signal Processing. He has been a Fellow of IEEE since 1988. He served as a Member of the Board of Governors of the IEEE Signal Processing Society (1989-1991). Since 1990, he has been the Editor-In-Chief of the Journal of VLSI Signal Processing Systems. He was a recipient of IEEE Signal Processing Society's Technical Achievement Award for the contributions on "parallel processing and neural network algorithms for signal processing" (1992); a Distinguished Lecturer of IEEE Signal Processing Society (1994); a recipient of IEEE Signal Processing Society's Best Paper Award for his publication on principal component neural networks (1996); and a recipient of the IEEE Third Millennium Medal (2000). He has authored and co-authored more than 400 technical publications and numerous textbooks including "VLSI and Modern Signal Processing", Prentice-Hall (1985), "VLSI Array Processors", Prentice-Hall (1988); "Digital Neural Networks", Prentice-Hall (1993); "Principal Component Neural Networks", John-Wiley (1996); and "Biometric Authentication: A Machine Learning Approach", Prentice-Hall (2004).

Naveen Verma received the B.A.Sc. degree in Electrical and Computer Engineering from the University of British Columbia, Vancouver, Canada in 2003 and the M.S. and Ph.D. degrees in Electrical Engineering from Massachusetts Institute of Technology in 2005 and 2009 respectively. Since July 2009 he has been an Assistant Professor of Electrical Engineering at Princeton University. His research focuses on ultra-low-power integrated circuits including low-voltage digital logic and SRAMs, low-noise analog instrumentation and data-conversion, and energy-efficient processing algorithms especially for biomedical applications. He is co-recipient of the 2012 Princeton Innovation Forum 1st Prize, the 2008 ISSCC Jack Kilby Award for Outstanding Student Paper, and the 2006 DAC/ISSCC Student Design Contest Award.