Sensor selection for energy-efficient ambulatory medical monitoring

Citation: Shih, Eugene I., Ali H. Shoeb, and John V. Guttag. "Sensor selection for energy-efficient ambulatory medical monitoring." Proceedings of the 7th international conference on Mobile systems, applications, and services. Krakow, Poland: ACM, 2009. 347-358.

As Published: http://dx.doi.org/10.1145/1555816.1555851

Publisher: Association for Computing Machinery

Persistent URL: http://hdl.handle.net/1721.1/51688

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Sensor Selection for Energy-Efficient Ambulatory Medical Monitoring

Eugene I. Shih
Department of EECS
Massachusetts Institute of Technology
Cambridge, MA, USA
eishih@mit.edu

Ali H. Shoeb
Department of EECS
Massachusetts Institute of Technology
Cambridge, MA, USA
ashoeb@mit.edu

John V. Guttag
Department of EECS
Massachusetts Institute of Technology
Cambridge, MA, USA
guttag@mit.edu

ABSTRACT

Epilepsy affects over three million Americans of all ages. Despite recent advances, more than 20% of individuals with epilepsy never achieve adequate control of their seizures. The use of a small, portable, non-invasive seizure monitor could benefit these individuals tremendously. However, in order for such a device to be suitable for long-term wear, it must be both comfortable and lightweight.

Typical state-of-the-art non-invasive seizure onset detection algorithms require 21 scalp electrodes to be placed on the head. These electrodes are used to generate 18 data streams, called channels. The large number of electrodes is inconvenient for the patient and processing 18 channels can consume a considerable amount of energy, a problem for a battery-powered device.

In this paper, we describe an automated way to construct detectors that use fewer channels, and thus fewer electrodes. Starting from an existing technique for constructing 18 channel patient-specific detectors, we use machine learning to automatically construct reduced channel detectors. We evaluate our algorithm on data from 16 patients used in an earlier study. On average, our algorithm reduced the number of channels from 18 to 4.6 while decreasing the mean fraction of seizure onsets detected from 99% to 97%. For 12 out of the 16 patients, there was no degradation in the detection rate. While the average detection latency increased from 7.8 s to 11.2 s, the average rate of false alarms per hour decreased from 0.35 to 0.19.

We also describe a prototype implementation of a single channel EEG monitoring device built using off-the-shelf components, and use this implementation to derive an energy consumption model. Using fewer channels reduced the average energy consumption by 69%, which amounts to a 3.3× increase in battery lifetime.

Finally, we show how additional energy savings can be realized by using a low-power screening detector to rule out segments of data that are obviously not seizures. Though this technique does not reduce the number of electrodes needed, it does reduce the energy consumption by an additional 16%.

MobiSys'09, June 22–25, 2009, Kraków, Poland.

Categories and Subject Descriptors

I.5.4 [**Pattern Recognition**]: Applications—Signal processing; J.3 [**Life and Medical Sciences**]: Health

General Terms

Algorithms, Human Factors, Measurement

Keywords

ambulatory medical monitoring, epilepsy, reducing energy consumption, channel selection, electroencephalography (EEG)

1. INTRODUCTION

Epilepsy is a relatively common chronic neurological disorder characterized by recurrent unprovoked seizures. Epilepsy affects over three million Americans of all ages, at an estimated annual cost of $12.5 billion in direct and indirect costs. Despite recent advances in the management of epilepsy, more than 20% of individuals with epilepsy never achieve adequate control of their seizures [24]. A portable, non-invasive seizure monitor could benefit these individuals tremendously. For instance, it could be used to alert patients or caregivers of seizure onset before symptoms of the seizure cause injury [31]. In addition, an ambulatory seizure onset monitor could be used to automatically initiate delivery of a therapy to reduce both seizure intensity and duration [33]. Both applications require timely detection of seizure onset.

Seizure onset is defined as the point in time where a patient's EEG transitions from a non-ictal state to an ictal state. A seizure is the period in time following this event and typically lasts from one to two minutes. Note that the baseline EEG signals of a person with epilepsy are considerably different from the EEG signals of a person who does not have epilepsy [23]. Even among patients with epilepsy, the amplitude, spectral content, and spatial distribution of EEG signals can vary during both seizure and non-seizure periods.

Two key issues in designing an ambulatory seizure onset monitor are patient comfort and convenience. Several factors have an impact on these issues including:

(a) *Number of electrodes*. State-of-the-art seizure onset detection algorithms require 21 scalp electrodes to be placed on a patient's head to acquire EEG signals. The electrodes are used to generate 18 data streams called channels. Reducing the number of electrodes will help to improve the long-term wearability of a seizure monitor.

(b) *Number of wires*. Using wires to connect sensing components to computation components is cumbersome for a patient to

wear. The use of a small number of wireless integrated, small and lightweight devices is likely to be more comfortable.

(c) *Electrode interface*. Current electrodes used for scalp EEG monitoring require the use of electroconductive gel to provide a low impedance interface between the electrode and the surface of the patient's scalp. Wearing gel electrodes for a prolonged period of time is uncomfortable. One potential way to mitigate this problem is to use dry electrodes [18]. However, since dry electrodes are larger in size (≈ 2.5 cm in diameter) than conventional electrodes (≈ 0.5 cm in diameter), they may be uncomfortable for weight or size reasons.

(d) *Device weight.* Since the device is designed to be worn on the head, the weight of the overall device should be minimized as much as possible.

In this paper, we focus on the problem of channel reduction. We present a machine learning based algorithm to construct reduced channel seizure onset detectors. Starting from an existing algorithm for constructing 18 channel patient-specific detectors, we use our technique to build reduced channel detectors. While using fewer channels will improve wearability and reduce energy consumption and device weight, the potential downside is that detection performance may degrade.

To select channels, we use an instance of the wrapper approach [13], a feature selection technique. Because EEG signals can vary considerably among patients with epilepsy, a patient-specific seizure onset detector trained using data from a single patient tends to have better specificity and sensitivity than a non-patient specific detector [27]. Thus, for selecting a channel subset for seizure onset detection, we elected to use our technique on each patient separately.

Using our approach, we reduced the average number of channels needed to detect seizure onset from 18 to 4.6. For 12 out of the 16 patients, the subset method detected all of the seizures detected by the original method. When we ran these detectors on recorded EEG data, we observed only a small degradation in performance as compared to the original 18-channel detector. The mean fraction of seizures detected decreased slightly from 99% to 97%. While the average detection latency increased slightly from 7.8 s to 11.2 s, the average number of false events per hour decreased from 0.35 to 0.19.

We also describe a prototype implementation of a single channel EEG monitoring device built using off-the-shelf components. Using this implementation, we derived a model that was used to evaluate the energy consumed by our detectors. By using fewer channels, our algorithm reduced the energy consumption by an average of 69%, which amounts to an increase in battery lifetime of $3.3\times$.

Finally, we show how to further reduce energy consumption by adding a screening detector to the reduced channel detector. By reusing the wrapper approach, we constructed a patient-specific screening detector for each patient. The combination of the patient-specific screening detector with the original reduced channel detector consumes 16% less energy while decreasing the percentage of seizures detected by only 1%.

2. DEFINITIONS

To clarify the terminology used in this paper, we provide some definitions for a number of key terms and performance metrics.

- *window*: Two seconds of consecutively sampled EEG data. Consecutive windows have a one second overlap.
- *false negative rate*: The false negative rate refers to the fraction of true positive windows, as labeled by a human expert, that are misclassified as negative windows.
- *false positive rate*: The false positive rate refers to the fraction of true negative windows that are misclassified as positive windows.
- *seizure onset declaration*: The onset of a seizure is declared when four consecutive windows are labeled positive by the detector.
- *fraction of seizure onsets detected*: The fraction of seizure onsets detected is the number of seizures correctly declared over the total number of seizures present in a recording. Note that a high false negative rate may not necessarily translate into a low fraction of seizures detected.
- *false events per hour (false alarm rate)*: If a seizure onset declaration is made during a period that does not contain an actual seizure, we call this declaration a false alarm or false event. This method of counting false events is quite conservative as we would treat two runs of four positive windows separated by a single negative window as two events. Note that a high false positive rate may not necessarily result in a high false alarm rate.
- *detection latency*: The detection latency is the difference in time between when a detection algorithm declares seizure onset and the electrographic onset time of the seizure as marked by a human expert. Since a minimum of four consecutive positive windows needs to be found before a seizure is declared, the minimum latency is five seconds since the windows are two seconds in duration and overlap by one second. If we were to allow fewer consecutive positive windows before an event could be declared, the detection latency would decrease. However, this would probably increase the false alarm rate.

3. BASELINE ALGORITHM

In a standard EEG acquisition system, data is collected from scalp electrodes arranged as shown in Figure 1. Various researchers [28, 22] have used machine learning to construct algorithms to detect seizures from data collected using this montage. Using data collected from a specific patient, a detection algorithm is trained and then deployed for in-hospital real-time seizure detection. In [28], the authors show how this algorithm could be used to initiate a delay-sensitive functional neuroimaging procedure following the detection of the electrical onset of a seizure.

Shoeb *et al.* [28] use a support vector machine (SVM) [6] based classifier to determine whether an observed two second EEG window resembles an individual's non-seizure EEG or one of their seizure onset events. First, the detector passes a window from each of the 18 channels to a feature extractor. For each channel, the feature extractor assembles a feature vector whose seven elements correspond to the energies in seven overlapping frequency bands. The features from all channels are concatenated to form a 126 dimensional feature vector which is then assigned to the seizure or non-seizure class using an SVM classifier trained to differentiate between a patient's EEG containing seizure onset activity and EEG containing non-seizure activity. Epileptiform activity lasting more than four seconds is taken to indicate the start of a seizure. This classifier was able to achieve high specificity, high sensitivity and low detection latency for many of the patients in the study. We use

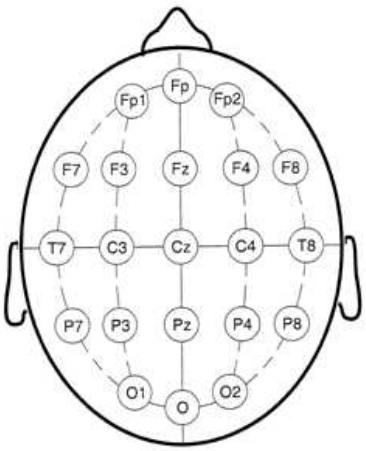
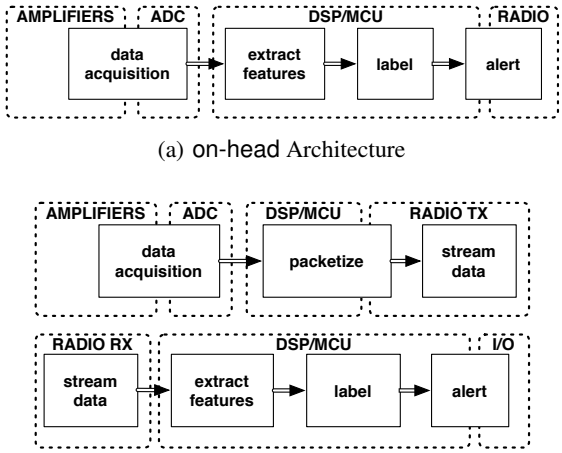


Figure 1: The 10/20 montage arrangement of scalp electrodes used for EEG monitoring. A channel is formed by measuring the difference between two adjacent electrodes (*e.g.* P7-O1).



(a) on-head Architecture

(b) off-head Architecture

Figure 2: A mobile seizure onset detection system can be realized as one of two different architectures.

this detector's performance and energy consumption as the baseline for comparison purposes.

The energy consumption of this algorithm was measured by Shoeb *et al.* in a subsequent paper [29]. However, the detector was implemented on a 32-bit floating point processor (TMS320C6711) operating at 150 MHz and consumed 2.5 W. A system that has such high energy consumption is not suitable for long-term ambulatory EEG monitoring.

4. SYSTEM OVERVIEW

In sensor-based detection algorithms, a sensor first converts a physical phenomenon into an electrical signal. A filtered version of the signal is then converted into digital samples. Often, samples are grouped to form windows that may overlap. Features are then extracted from the samples and are analyzed to produce a sequence of labels—each label indicates whether the features satisfy a specific set of criteria. For seizure onset detection, features corresponding to windows are labeled as containing seizure onset activity or not. We consider two basic architectures for implementing a seizure onset detector. Depending on the processing requirements of the detection algorithm, one architecture may be more suitable than the other.

Figure 2(a) shows a block diagram of an on-head architecture. The system is composed of an EEG data acquisition block that includes circuitry to amplify and filter the μV EEG signals. An analog-to-digital converter (ADC) converts each channel into digital samples. A microcontroller or DSP is then used to analyze the EEG channels in real-time using a patient-specific seizure onset detection algorithm. To perform detection in real-time, all operations must be complete before the next window of data arrives. When the onset of a seizure is detected, a message is sent to a remote device using a wireless radio to issue an alert to the patient or to initiate a therapy. Because the rate of seizures is fairly low (fewer than one per day for most patients), the use of the radio will be rare. Therefore, in this architecture, the predominant energy cost is in collecting and processing the data.

Figure 2(b) shows a diagram of an off-head architecture. This implementation includes circuitry to collect and sample the data, but no feature extraction or detection is performed locally. Instead, all the data is transmitted to a remote device where more computation and energy resources are available. In this architecture, the predominant energy cost is in transmitting the data.

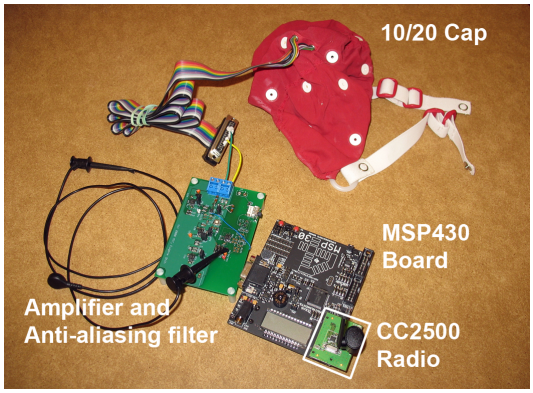
4.1 Prototype description

To understand the energy consumption of a mobile seizure onset detection system, we built a single channel EEG acquisition system using off-the-shelf components and measured the power consumption of the components using an oscilloscope. The prototype transmits acquired data to another device where final classifications can be performed. Thus, our prototype is an instance of the off-head architecture. While the system is capable of acquiring data from more channels, for the purposes of developing an energy model, building a single channel acquisition system is sufficient. Since the system is designed for wearable monitoring, we made an effort to select low-power, lightweight components.

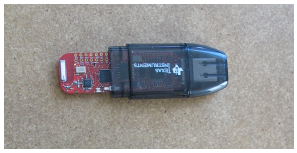
To initially amplify the EEG signal, we selected the Texas Instruments INA333 low-noise, low-power instrumentation op-amp and an anti-aliasing filter. Because EEG signals have a peak-to-peak amplitude of a few μV, a gain of approximately 1000 is applied to the signal. After the signal is properly filtered, it is sampled at 200 samples per second using the on-board 12-bit analog-to-digital converter (ADC) of a microcontroller, the Texas Instruments MSP430FG4618 [19], operating at 8 MHz at 3.0 V. In general, the interesting frequency range for scalp EEG is 0 to 150 Hz, but for detecting seizure onset, the maximum frequency of interest is around 75 Hz.

At the ADC, we truncate each sample to 8 bits. Once a full second's worth of data has been collected, *i.e.*, 200 bytes, the data is split into 50 byte packets and delivered to a CC2500 transceiver [4], which operates at 2.4 GHz. The data is split because the size of the transmit and receive FIFOs in the radio is 64 bytes. Along with the payload data, an 8-bit sequence number is included to help with detecting packet loss and ordering the incoming packets. Data is transmitted at 250 kbps using minimum shift-key (MSK) modulation and forward error correction. The prototype system is shown in Figure 3(a).

In our prototype, the data is delivered to a desktop computer via the EZ430-RF2500 module from Texas Instruments, which consists of a CC2500 transceiver as well as another MSP430 as shown in Figure 3(b). The prototype is powered by 2 AAA batteries as



(a) Acquisition hardware



(b) EZ430-RF2500 module

Figure 3: Prototype single channel EEG acquisition system. The EZ430-RF2500 module receives the data from the acquisition hardware. It is also connected to a PC where the data can be analyzed.

shown in Figure 4. The module stores a full second's worth of data before sending the data via a USB serial port to the desktop computer. In an actual deployment, the data would be transmitted to a device located at the patient's belt.

4.2 Energy model

Using an oscilloscope, we measured the energy consumption of the three main components of our system: the data acquisition circuit, the processor, and the radio. In our prototype, the data acquisition circuit remains fully powered at all times. Duty cycling the front-end is possible, however, aggressive duty cycling may lead to data corruption. When powered on, the data acquisition circuit has a current consumption of 100 μA at 3.0 V. The energy consumed to acquire N_c channels of data can be modeled as

$$E_{\text{sample}} = N_c(P_{\text{sensor}}^{(\text{on})} T_{\text{sensor}}^{(\text{on})}) \quad (1)$$

where $T_{\text{sensor}}^{(\text{on})}$ is the time the data acquisition unit is active.

As mentioned, the signal is collected at 200 samples per second using the on-board ADC of the MSP430. Since we configured the ADC to use an on-board timer clocked by the SMCLK signal to trigger conversions, we can place the CPU into Low Power Mode 0 (LPM0) only during idle periods, otherwise, the ADC will not function properly. Sampling a single EEG channel can be accomplished by operating the ADC in single-channel, repeated conversions mode; multiple channels can be sampled by using the ADC in multiple-channel, repeated conversions mode. Doing so will increase the on-time of the ADC, but the increase will have a negligible impact on the energy consumption because power consumed by the ADC is small compared to the CPU operating in LPM0 mode.

The energy consumed by the MSP430 can be approximated as:

$$E_{\text{CPU}} = P_{\text{CPU}}^{(\text{on})} T_{\text{CPU}}^{(\text{on})} + P_{\text{CPU}}^{(\text{idle})} T_{\text{CPU}}^{(\text{idle})}.$$

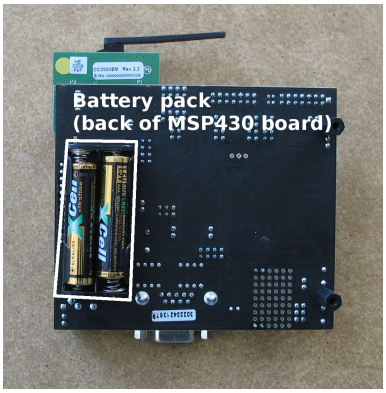


Figure 4: Battery pack used by the MSP430 Experimenter's Board. The MSP430 Experimenter's Board uses 2 AAA batteries.

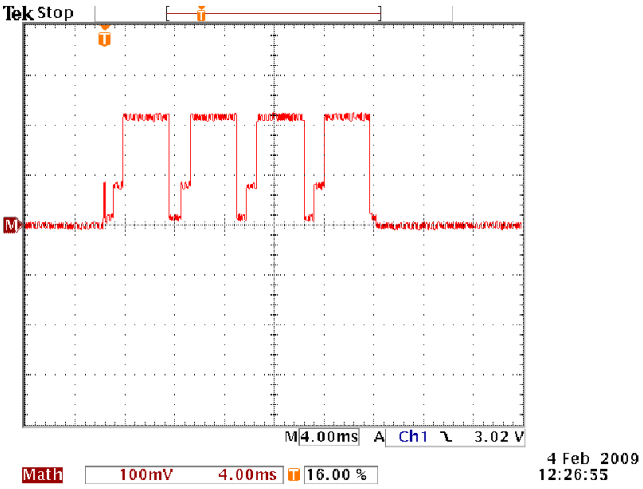


Figure 5: A trace of the power consumed by the CC2500 during radio transmission of a single channel collected over a one second window. The total amount of payload data is 200 bytes. Because of the limited size of the CC2500 transmit FIFO, we transmit 50 bytes at a time.

$T_{\text{CPU}}^{(\text{on})}$ and $T_{\text{CPU}}^{(\text{idle})}$ are the durations for which the CPU is active and idle respectively. In our system, the CPU is activated once per second. Thus, in a one second period, the CPU is idle for $1 - T_{\text{CPU}}^{(\text{on})}$ seconds. The CPU is only active when data needs to be transmitted. If T_{active} is the duration of time that the radio needs to be active in order to send one channel's worth of data for N_c channels, the energy consumed in one second can be written written as:

$$E_{\text{CPU}} = P_{\text{CPU}}^{(\text{on})}(N_c T_{\text{active}}) + P_{\text{CPU}}^{(\text{idle})}(1 - N_c T_{\text{active}}). \quad (2)$$

To transmit data to a nearby remote device (*e.g.*, on the belt), we use the CC2500 [4] transceiver. Figure 5 shows a trace of the radio sending 200 bytes of data. There are several things to note. First, the transmission of each 50 byte packet takes about 3.7 ms. We pause 800 μs between each packet to give the receiver a chance to process the previous packet of data. Also, notice that the radio transitions between various power states in order to transmit. We can identify five different power states from the diagram, which we will call: idle, transmit, listen, wake-up, and transfer. The operation

of the radio in idle, transmit, listen, and wake-up states is obvious. During the transfer state, data to be transmitted is written to the transmit FIFO on-board the CC2500. As data from more channels is sampled, the number of bits to transmit will increase. Thus, we can model the energy consumed by the radio as:

$$E_{\text{radio}} = \sum_{i \in R} P_{\text{radio}}^{(i)} T_{\text{radio}} = N_c \sum_{j \in S} P_{\text{radio}}^{(j)} T_{\text{radio}}^{(j)} + \sum_{k \in T} P_{\text{radio}}^{(k)} T_{\text{radio}}^{(k)} \quad (3)$$

where $S = \{\text{transmit}, \text{listen}, \text{transfer}\}$, $T = \{\text{idle}, \text{wake-up}\}$ and $R = S \cup T$. The duration of time the radio spends in the idle mode in one second is simply one minus the sum of the duration spent in the other modes.

Table 1 summarizes the values of the parameters of components used in our prototype system. Our model does not take into account the energy needed to switch state to state; we believe that this energy is negligible. Combining Equations (1), (2), and (3), and using the values in Table 1, we can compute the energy consumed to collect and transmit one second's worth of data as a function of N_c:

$$E(N_c) \approx 1.58 \text{ mJ} \cdot N_c + 2.12 \text{ mJ}. \quad (4)$$

Therefore, a system that uses 18 channels has a power consumption of 30.6 mW. If we are able to reduce the number of channels required by a seizure onset detector to one channel, we would consume 3.71 mW, a reduction of 88%. We use this model in the rest of the paper to evaluate the energy consumption.

One could build a significantly lower power on-head seizure onset detection system by using application-specific integrated circuits (ASIC) as shown by the authors of [32, 2]. The purpose of building our prototype was not to build the lowest power system possible, but rather to derive a reasonable energy model for an ambulatory seizure onset detection system. We use the derived model to demonstrate the usefulness of our channel reduction techniques. Moreover, our techniques are still applicable to ASIC-based systems.

Reducing the number of channels can improve wearability and reduce energy consumption, but there are at least two issues left to address. (a) How should the subset of channels be chosen? (b) Can a detector built using fewer channels achieve similar levels of specificity, sensitivity, and latency as those of the baseline detector that uses all the channels? We attempt to answer these questions in the rest of the paper.

5. CHANNEL SELECTION

To select channels, we use a feature selection algorithm from the machine learning literature called the wrapper approach [13]. Since EEG data from each patient during both seizure and non-seizure periods will differ from the EEG signals of other patients, we perform feature selection using the wrapper approach on a per patient basis.

5.1 Wrapper approach

The wrapper approach [13] works by using the performance of a learned algorithm to assess the usefulness of different subsets of features. When implementing the wrapper approach, one needs to decide (a) how to search the space of possible feature subsets, (b) how to assess the performance of each learned machine to guide the search, and (c) which learning algorithm to use to select features. One advantage of the wrapper approach is that it allows users to specify an objective function with which to evaluate feature subsets.

```
1: F ← [1, 2, ..., n] {Initialize list F to all features}
2: R ← [] {R is the list of models}
3: V ← [] {V is a value dictionary indexed by feature}
4: S ← (X, y) {Data on which to test performance}
5: repeat
6:    for i ← 1 to |F| do
7:       A ← [] {A stores trained models}
8:       for all s ∈ S do
9:          F' ← F − F[i] {Remove a feature}
10:         X' ← X(:, F') {Select features from data and train}
11:         α ← Train(X', y) {Train classifier}
12:         A ← [A, α] {Remember model}
13:      end for
14:      V[i] ← J(A, S) {Evaluate objective function J}
15:   end for
16:   k ← findBest(V) {Get index of worst feature}
17:   R ← [R, A[k]] {Store best model}
18:   F ← F − F[k] {Remove feature from list}
19: until F = []
20: Output R
```

Figure 6: Backward elimination algorithm

The simplest search for the best feature subset is an exhaustive search where a learned machine is built for all feature subsets. Unfortunately, for n features, there are $2^n - 1$ possible subsets. To train an SVM, a quadratic programming optimization problem is solved. In our experience, training an SVM for each subset takes 30 to 60 seconds on average. Therefore, for $n = 18$ channels, building an SVM for all subsets would take around 90 days assuming a single CPU. Thus, performing such a search is computationally impractical. An alternative is a greedy search. In a greedy approach, we still use machine learning to build our detectors, but we will decide which subsets to evaluate incrementally. There are two basic strategies: *forward selection* or *backward elimination*. In forward selection, we incorporate variables incrementally into larger and larger subsets. In backward elimination, we start with the set of all features and gradually remove or eliminate the least useful ones. To evaluate the performance of each feature subset, we use cross validation.

Figure 6 shows the pseudo-code for the greedy backward elimination algorithm. We assume that there are n total features. Essentially, in each round, we obtain the performance of a subset of features by training a detector using that subset and running it on the training data. Each set of features is generated by removing a different feature with replacement from a set of candidate features. After the performance of the detector built using each subset of features has been determined, we remove the feature that results in the smallest degradation in performance. Performance is defined by the user-specified objective function J. We use the training data to help us evaluate performance. At the same time, we store the detector that uses the best set of features in R. As a final step, the algorithm outputs R, a set of detectors that have the best performance for a given number of features.

We adapted the wrapper approach to select entire channels instead of individual features. Since seven features are extracted per channel, when a channel is removed, all seven features are removed from the feature vector. An SVM is trained using the features from the remaining channels. In the inner loop, we remove each channel with replacement until all remaining channels have been removed.

After the inner loop is complete, we must decide which of the detectors to keep before continuing with the algorithm. In this pa-

Table 1: Measured values for the parameters used in our energy model. The supply voltage for all the components was 3.0 V. The dominant source of power consumption is the radio during transmit and transfer modes.

EEG sensor	$P_{\text{sensor}}^{(\text{on})}$	sensor power consumption	300 μW @ 3.0 V
MSP430	$P_{\text{CPU}}^{(\text{on})}$	active power consumption	14 mW @ 8 MHz
	$P_{\text{CPU}}^{(\text{idle})}$	idle power consumption	1.6 mW @ LPM0 mode
	T_{active}	active time	23.2 ms
CC2500	$P_{\text{radio}}^{(\text{transmit})}$	transmit power consumption	64 mW @ 0 dBm
	$P_{\text{radio}}^{(\text{listen})}$	listen power consumption	4.8 mW
	$P_{\text{radio}}^{(\text{transfer})}$	transfer power consumption	24 mW
	$P_{\text{radio}}^{(\text{wake-up})}$	wake-up power consumption	24 mW
	$P_{\text{radio}}^{(\text{idle})}$	idle power consumption	0.5 mW
	$T_{\text{radio}}^{(\text{transmit})}$	transmit time	14.7 ms
	$T_{\text{radio}}^{(\text{listen})}$	listen time	0.64 ms
	$T_{\text{radio}}^{(\text{transfer})}$	transfer time	3.2 ms
	$T_{\text{radio}}^{(\text{wake-up})}$	wake-up time	0.048 ms

per, we select the detector that has the lowest false positive rate that correctly detects at least as many seizure onsets as the baseline detector. If there are no detectors that satisfy the seizure onset detection constraint, we simply pick the detector that has the lowest false positive rate. The output of the algorithm is a list of learned detection algorithms that use decreasing numbers of channels to perform detection. From this list, we then select the single best detection algorithm. For this last step, we retain the detector with the smallest number of channels instead of the detector with the lowest false positive rate.

5.2 Other methods

In addition to the wrapper method, there are a number of other feature selection methods that we could have used [11]. We also tried and evaluated AdaBoost [9] and SVM Recursive Feature Elimination [12] for selecting channels. Detectors constructed using the channels returned by those methods did not perform any better than the detectors constructed using the channels returned by the greedy wrapper method.

6. EVALUATION METHODOLOGY

To evaluate our channel selection and detection algorithms, we used pre-recorded pediatric EEG data from 16 patients collected in a previous study [27]. All patients were being evaluated for potential cerebral resection surgery, and had their medications stopped. Therefore, they had seizures more often than is typical.

Table 2 shows the number of seizures and the number of hours of EEG data recorded for each patient. The onset and duration of each of the seizures for each patient was labeled by a human expert. For each recording, labels for each window can be derived from the annotations made by the expert.

Those who are unfamiliar with collecting EEG data should be aware that obtaining a data set of this kind and size is a challenge. Collecting data from patients with seizures involves finding a neurologist with patients who are willing to participate in a study, obtaining Institutional Review Board (IRB) approval, and finally, going to the hospital to record the data.

6.1 Testing procedure

The EEG data belonging to each patient is organized into consecutive, one-hour records. Let N be the number of seizure-free, one-hour records and M be the number of one-hour records containing one or more seizures for a given patient. For each of the patients, we used data from that patient to construct two different patient-specific classifiers:

Table 2: Data from each patient was obtained from a previous study [27]. The table shows the number of seizures and number of hours of EEG data recorded for each of the 16 patients.

Patient Id	Number of Seizures	Recording Duration (h)
40	24	16.0
25	4	55.0
22	7	43.0
36	3	16.5
10	3	21.3
24	2	56.0
31	5	17.0
38	3	34.8
41	9	32.0
3	5	33.8
12	2	38.0
20	4	28.0
35	7	48.0
43	8	25.0
45	38	25.0
47	19	34.0

(a) original: an SVM-based classifier using 18 channels and

(b) subset: an SVM-based classifier using a subset of the channels, where the channels are chosen using the wrapper approach described in Section 5.1.

We used the `LIBSVM` package [5] to train the SVM classifiers. After conducting a grid search using a subset of the patients, we set $C = 10$ and $\gamma = 10^{-2}$.

When training a classifier for a patient, we input a subset of the windows from all N of the non-seizure records and windows from $M - 1$ of those records containing a seizure into the learning algorithm. We input only a subset of windows for practical reasons since using all the windows from the non-seizure records for training is time consuming. Experiments showed that including only a subset of windows versus including all the windows did not adversely affect the detection performance.

The detector that resulted was then tasked with detecting seizure onsets in all M seizure record including the record that was left out from the training set. For each detector, we recorded the follow-

Table 3: Channels selected for Patient 38 by the wrapper algorithm.

Training subset	Test file	Channels Selected
A	3	FP1-F7, F7-T7, T8-P8
B	2	FP1-F7, P3-O1, T8-P8, Fz-Cz
C	1	F7-T7, T8-P8, Fz-Cz

Figure 7: For Patient 38, the average rate of false alarms per hour is lowered from 0.1 to 0.03. A 0.03 false alarm rate per hour corresponds to one false alarm every 33 hours.

ing data: (a) the fraction of seizure onsets detected, (b) the number of falsely declared events per hour of monitoring, (c) the mean detection delay for each seizure in the recording, and (d) the energy consumption. This process was repeated M times so that each of the M seizure records was tested once. For each classifier, we aggregated the results from the M runs and reported averages of the results. To estimate the energy consumed by a classifier for the entire recording, we used the model developed in Section 4.2.

In our experiments, our primary goal was to determine whether "good" reduced channel detectors could be constructing using the approach outlined and to compare the labeling performance and energy consumption of the original 18 channel detector to the reduced channel detector.

7. CASE STUDY

In this section, we focus on the results from a single patient (Patient 38). Patient 38 experienced three seizures over a period of 34.8 hours; thus, we can construct three training subsets. As explained in the previous section, to construct a training subset, we included all the windows from two of the seizures and a subset of windows from the non-seizure data. The testing subset included all the windows from the seizure data and all the windows from the non-seizure data. Table 3 shows the channel subsets selected using the wrapper algorithm. From the table, we see that the wrapper approach output detectors that use fewer channels than the original 18 channel detector. Moreover, each of the reduced channel detectors detected all the seizures that the original detector detected in Patient 38's EEG recording. Figures 7, 8, and 9 show the false alarm rate, mean detection delay and energy savings achieved by the subset detection algorithm and original full montage detection algorithm for Patient 38.

In Figure 7 we see that using fewer channels reduced the rate of false alarms from an average of 0.1 to 0.03 false events per hour. A 0.1 false event rate corresponds to a false alarm once every 10 hours, while a 0.03 false event rate corresponds to a false alarm once every 33 hours. Whether an alarm is real or not, each alarm will cause emotional distress in a patient as he or she prepares for clinical onset. A false alarm puts patients in distress unnecessarily and a high false alarm rate will cause patients to distrust alarms to the point that they may ignore them altogether. Thus, a low false alarm rate is critical.

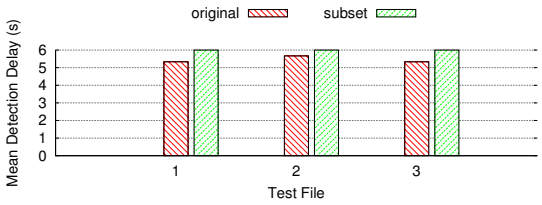


Figure 8: For Patient 38, the mean detection delay increases slightly to 6.0 s.

Figure 9: The energy savings for the different models trained for Patient 38.

The decrease in the false alarm surprised us; we had not expected that using fewer channels would reduce the number of false alarms. Even more surprising is that this result generalized to the other patients as we will see in Section 8. One explanation for the decrease is the following. The objective function used to evaluate channel subsets favors selecting a group of channels that has lower false positive rates. In particular, the objective function will tend to reject a channel that reliably exhibits epileptiform activity at seizure onset if that channel exhibits the same activity at other times. If the rejected channel exhibits this activity earlier than the channels in the subset, the latency will increase. Figure 8 compares the mean detection delay of the subset detection algorithm to that of the original detection algorithm. We see that the mean detection delay increased from 5.4 s to 6.0 s.

Figure 9 shows the average energy savings achieved by the subset detector assuming the energy model in Equation (4). For this patient, the energy consumption was reduced by an average of 76%.

We examined the EEG signals during the seizures to determine if the channels selected had a physiological explanation. Figure 10 illustrates the onset of a recorded seizure from Patient 38. For this patient, seizures appear to originate on the left-side of the brain. Seizure onset begins with the appearance of 8 Hz rhythmic activity most prominently on the left hemispheric channels (the top 8 channels in the figure; FP1-F7, FP1-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1). The onset has been marked with a vertical line. There is limited, or no, involvement of right hemispheric channels (the bottom 8 channels in the figure; FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8 ,F8-T8, T8-P8, P8-O2). Based on the EEG signals, the channels selected by the wrapper algorithm were reasonable; most of the channels selected were from the left hemisphere. In Table 3, we see that T8-P8 was also selected. During the seizure, there appears to be no activity on T8-P8, but its inclusion may have improve the detection rate since T8-P8 was reliably quiet during all the seizures in the recording.

One potential concern here is that the wrapper algorithm chooses different channel subsets when different seizure files are left out. This raises the question of how sensitive the detection performance is to the choice of channels. To address this question, we performed the following experiment. We took all the channel subsets reported

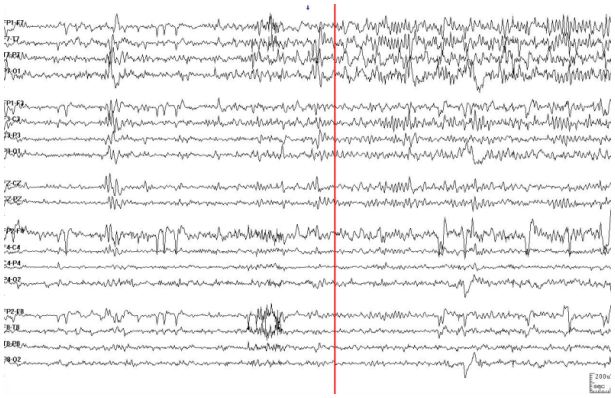


Figure 10: EEG signals of Patient 38 during a seizure. Seizure onset is denoted by the red line.

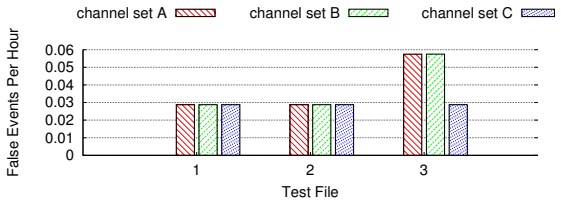


Figure 11: The false event per hour for Patient 38 remains low even when the channel subsets are changed.

by the detector when a specific file was left out and then trained a detector using those channels while leaving out a different file. For example, when we used Files 1 and 2 as the training set, channels FP1-F7, F7-T7, and T8-P8, were selected. What happens if we build a detector using those channels while using Files 2 and 3 in the training set?

After performing this experiment, we determined that the fraction of seizure onsets detected remained at 100% no matter which files were used for training and which files were left out for testing. Neither the mean detection delay nor average energy savings changed significantly when different data sets were used for training.

Figure 11 shows the resulting false alarm rate when the channels selected using the original training subsets were used to train detectors using data from a different training subset. In all cases, the false alarm rate remains fairly low. Thus, for this patient, the detection performance appears to be unaffected when any of the channel subsets as selected by our algorithm is used to build a detector.

8. RESULTS FOR ALL PATIENTS

The original SVM-based detector detected 99% of the seizure onsets with an average false alarm rate of 0.35 events per hour and a mean detection delay of 7.8 seconds. The subset detection algorithm detected 97% of all the seizure onsets while lowering the average false alarm rate to 0.19 false events per hour. However, the mean detection delay rose to 11.2 s. Figures 12, 13, and 14 show the detection performance for all the patients. Note that for 12 of the 16 patients, the subset method detected all of the seizures detected by the original method.

From Figure 13, we see that the decrease in the false alarm rate that was observed for Patient 38 is also observed in the other patients. Again, it is likely that the eliminated channels exhibited seizure-like activity during non-seizure periods. By removing them, the algorithm reduced the false alarm rate.

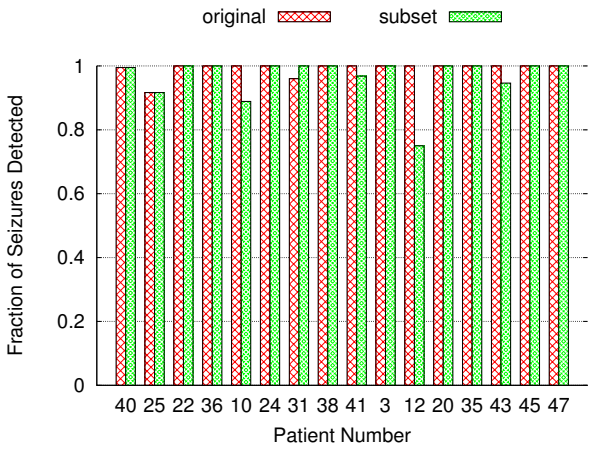


Figure 12: The average fraction of seizures detected decreased slightly from 99% to 97%. For any one patient, this amounted to missing at most one seizure for one training subset.

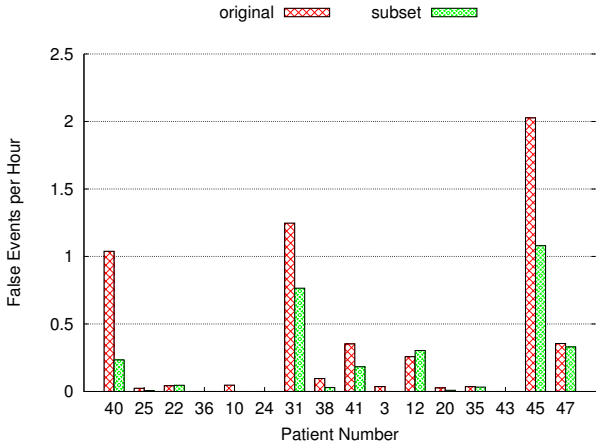


Figure 13: Using the subset detector lowered the average rate of false events per hour from 0.35 to 0.19.

Figure 15 shows the average number of channels that was selected by the wrapper algorithm for each patient. The large variation in the mean number of channels across patients can be attributed to the fact that each patient experiences seizures in a different way. Moreover, different seizures will originate from different parts of the brain. Overall, the average number of channels in the reduced channel detectors was 4.6 channels.

Figure 16 illustrates the average energy savings achieved by the subset detector for each patient for the off-head architecture. The algorithm achieved an average savings of 69%.

Our approach effectively reduce overall energy consumption. Using a reduced channel detector will reduce the amount of data we need to collect, process, and/or transmit, but what is the impact on battery lifetime? The impact depends on the battery being used. A number of different types of batteries are available commercially including lithium-ion (Li-Ion), lithium-polymer (Li-Poly), alkaline, and nickel-metal hydride (NiMh). In our prototype, two AAA alkaline batteries are used to power the system. Table 4 shows the energy densities of various common batteries. For the

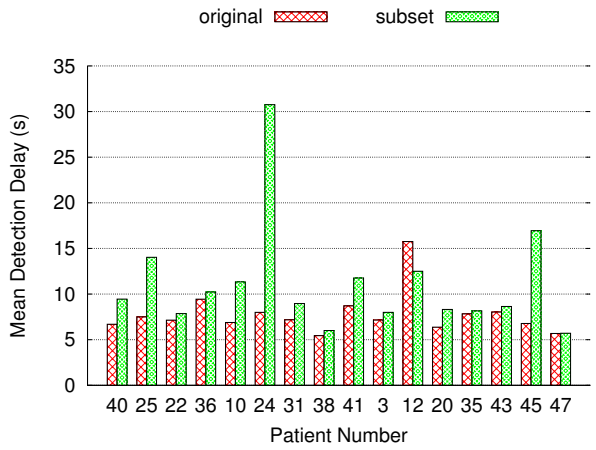


Figure 14: The average detection delay increased from 7.8 s to 11.2 s for most patients. For some patients, the increase in delay may be unacceptable. In those cases, an objective function that minimizes delay would be more appropriate.

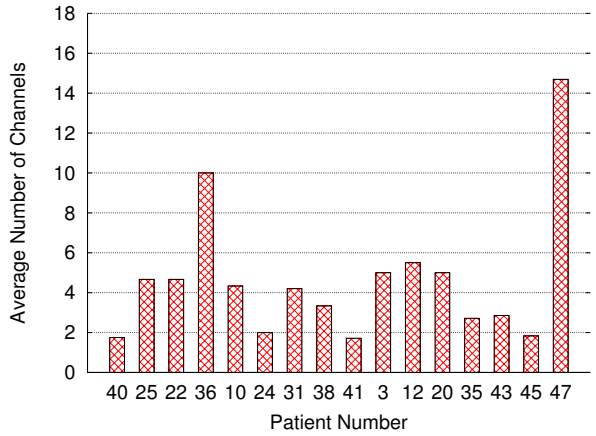


Figure 15: The average number of channels that the wrapper algorithm selected was 4.6.

ambulatory seizure onset detection application, the battery is worn on or near the head, thus, using the lithium-ion coin cell which has one-third the weight of the other batteries is preferred. Figure 17 shows the lifetime of the system while the number of channels for the detector is varied. A curve is shown for each of the three battery types. We have assumed that the batteries do not self-discharge and can maintain the same voltage level regardless of the remaining battery capacity. Note that the battery is not the only contributor to the weight, the current single channel prototype weighs approximately 30 to 40 g. However, using off-the-shelf components, a single channel system including the amplifiers could be reduced to 10 g, which is the weight of the EZ430-RF2500. Thus, reducing the number of channels will also reduce overall device weight.

9. FURTHER IMPROVEMENTS

We have shown that the wrapper approach can be used to find subsets of channels with which to build reduced channel detectors that have high specificity, high sensitivity, and reasonable detection latency compared to the original 18 channel detector. We can further reduce the energy consumption of the subset detector by

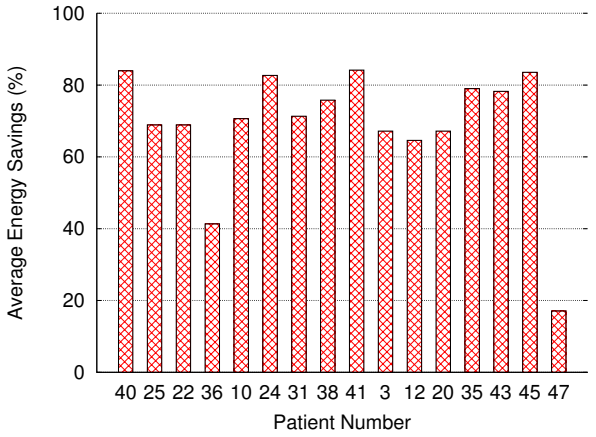


Figure 16: The average energy savings achieved by the subset detection algorithm was 69%.

Table 4: Specifications for some common batteries. The Li-Ion batteries are rechargeable batteries, while the AAA battery is a non-rechargeable alkaline battery. For long-term seizure onset detection where the device is placed on or near the head, the use of a smaller lightweight battery is preferred.

	Li-Ion (phone)	AAA	Li-Ion (coin)
Weight (g)	24	23	7.8
Capacity (mAh)	850	900	160
Nominal Voltage (V)	3.7	3.0	3.6
Ideal Capacity (kJ)	11.3	9.7	2.1

adding a low-power *screening detector* in front of the detection algorithm as suggested by [26]. We hypothesize that non-seizure onset segments are easier to detect and thus, a less complex, lower power classifier can be trained to detect those segments. If we can construct such a detector with a low false negative rate, we can combine this simpler detector with the original subset detector as follows.

Initially, the screening detector is used to label each window. Any segment that the screener labels as non-seizure onset is labeled negative. Any segment in which the screening detector believes a seizure is present is forwarded on to the original subset detector. If the subset detector labels the window positive, it continues to analyze all future windows until a negative window is found.

Combining the detectors in this way ensures that it will have no false positives relative to the subset detector. As long as the power consumption of the screener and the number of false alarms issued by the screener is low, the combination of the detectors can decrease energy consumption while maintaining low false positive and false negative rates.

9.1 Constructing the screener

One way to construct a screener is to build a detector that uses fewer channels than the subset detector. Conveniently, we can use the backward elimination algorithm again. To determine which channels to keep, we use the following objective function: (a) Find the screening detector that has the lowest number of false negatives, but detects at least as many seizures as the subset detector. (b) Retain the channels used by that detector. Channels selected by this approach are used to construct the final SVM-based screening detector.

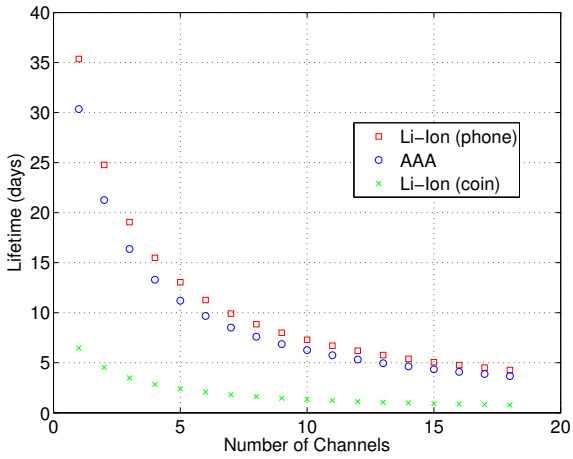


Figure 17: With all 18 channels active, the heavier Li-Ion battery has an ideal battery lifetime of about 4.3 days. The AAA battery has an ideal battery lifetime of 3.7 days and the Li-Ion coin has an ideal battery lifetime of 0.78 days. By reducing the channels to 4, the battery lifetime increases to 15 days, 13 days and 2.8 days respectively.

Table 5: Detection performance is unchanged when a screening algorithm is added.

Average fraction of seizures detected	96%
Mean detection latency (s)	12.1
False events per hour	0.17
Average energy savings	85%

Note that using a screener will not reduce the number of channels worn by the patient. However, because screening can reduce the number of channels that are actively processed, the energy consumption of the device will be reduced and thus, the lifetime of the battery will be extended. Note that in some cases, the algorithm could not find a set of channels that satisfied our criteria. For patients where the subset algorithm used only a single channel, we did not run the screening algorithm.

9.2 Screening results

Table 5 shows the results of adding a screening detector to the subset algorithm when possible. Notice that no significant change was observed in the fraction of seizure onsets detected, the average false alarm rate, or the mean detection latency. However, we were able to improve the energy savings for some patients. By using the screening method, we were able to achieve an average energy savings of 85% over the original 18 channel detector. Figure 18 shows the additional energy savings achieved for each patient. Figure 19 shows the average number of channels used by the screener for each patient. The mean number of channels used by the screener for patients who could be screened was 1.17.

10. DISCUSSION

One concern with the results we have presented is that the seizures and spatial distribution of the seizures for a patient will change over time. Thus, the channels selected based on previously collected data will be irrelevant as time progresses. Unfortunately, published information about how seizures change over years or decades is not available. Clinicians generally believe that adults

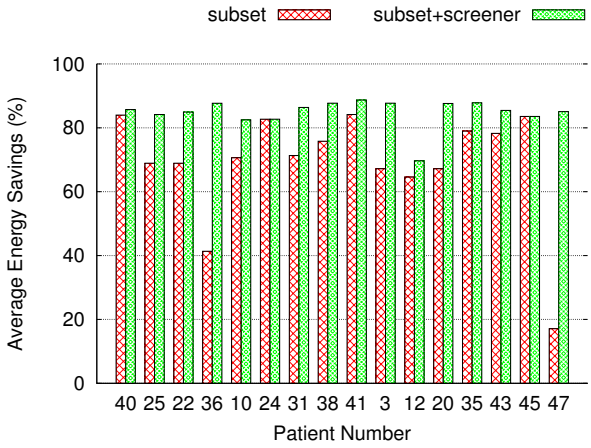


Figure 18: The additional energy savings achieved by the combination of the screening detector with the subset detection algorithm was approximately 16%.

with epilepsy tend to have stable EEGs. In our own clinical studies, we have successfully used a patient-specific detector trained using data collected over a year earlier to detect a seizure in real-time.

Another concern is that using scalp electrodes is simply too cumbersome for long-term monitoring use. Unfortunately, other methods of seizure onset detection that do not use EEG electrodes are more error-prone. For example, one could try to use electromyography (EMG), the electrical signals generated by muscle movement, to detect a seizure. There are several problems with such an approach. First, not all patients experience strong muscle movement during seizures. Second, deciding where to place the EMG electrodes will be patient-specific since patients experience seizures in different ways. Finally, muscle movements that indicate clinical onset of seizure often lag the initial electrographic seizure onset. Thus, even if it were possible to detect a seizure as a result of a change in the EMG signal, the detection latency would likely be larger.

11. RELATED WORK

Reducing the power consumption of mobile devices and systems has been well studied by many researchers [14, 8, 16]. For the most part, a reduction in energy consumption is achieved through some form of duty cycling. Deciding when and how frequently to turn a device off often depends on the application. Typically, the decision to turn devices off is based either on user input or on other environmental changes. In many previous systems, power savings is achieved by trading off the performance or fidelity of certain applications (*e.g.*, a low-resolution image is displayed or a task takes longer to complete). In this paper, we use software techniques to select channels in order to reduce energy consumption of seizure onset detection. However, unlike previous work, for many patients, we are able to achieve energy savings with only a slight degradation in the detection performance.

A few researchers have explored the use of feature selection to reduce the complexity or energy consumption of a detector or classifier. In [30], the authors use a genetic algorithm to construct a decision tree where the fitness function takes into account the cost of performing a test and cost of making classification errors. The work is evaluated using medical data sets. Energy consumption is not considered since the medical tests are not intended for real-time use.

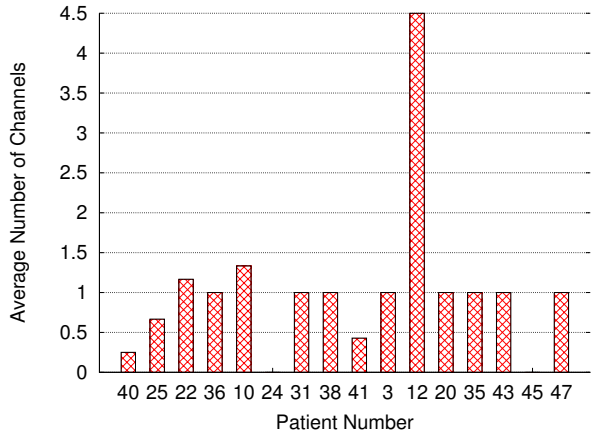


Figure 19: The average number of channels used in the screener was 1.17. For some patients, a screener was not needed since the subset algorithm used only a single channel. For other patients, the screeners tended to degrade detection accuracy on the training data. Thus, we did not use a screener for those patients.

More recently, Benbasat *et al*. take into account the energy cost of sampling various sensors for performing wearable gait monitoring [3]. They describe a framework for automatically generating power-efficient decision tree classifiers for wearable gait monitoring nodes. Similar to AdaBoost, decision tree learning algorithms perform embedded feature selection. By modifying the learning algorithm to take into account energy cost, a tree that is "energy-aware" can be built. Once a decision tree is constructed, the system dynamically consults only the sensors that are necessary to determine the system state. However, because sensors are activated sequentially over time, the detection latency of such systems will be greater than those that examine all the data simultaneously. We have not yet tested the use of a decision tree algorithm for our application. However, decision tree classifiers tend to have lower specificity and sensitivity than SVM-based classifiers. Nevertheless, it may be worthwhile to compare these methods to ours.

The authors of [20] also explore ways to scale the data or feature set used by an ECG-based application for use on a mobile device. The authors show how to use their methods to trade off classification accuracy for bandwidth. In particular, they use a feature selection method based on mutual information to "scale" the data set such that computation is appropriate for a mobile device. Other researchers [10, 15] have applied feature selection techniques such as Recursive Feature Elimination for reducing channels in EEG-based applications.

The idea of using a lower energy, but lower fidelity device most of the time is similar to our concept of screening. In [25], the authors use a low-power radio to receive incoming wake-up messages. This allows the rest of the device to remain off when the device is not being used and results in dramatic energy savings. The use of multiple radios to reduce energy consumption is further explored by various researchers [21, 1] who use Bluetooth or cell phone technologies to implement the wake-up mechanism. The wake-up mechanism is similar to screening in that a higher-power device is activated only when an incoming message is detected on the low-power channel.

Similar to wake-on-wireless systems, a number of researchers have built sensor-based hardware platforms designed to detect rare and infrequent events [7, 17]. In these systems, the CPU stays in a very low-power mode and is activated only when a sensor detects some signal energy.

Previously, we investigated the use of a screener detector to reduce energy consumption for the seizure detection problem [26]. In this paper, we use feature subset selection algorithms for improving patient comfort and reducing energy consumption. Secondly, we introduce an energy model based on a real hardware implementation to determine the change in energy consumption.

12. CONCLUSION

Medical researchers have developed good algorithms for extracting clinically useful information from physiological sensors. In order to deploy these algorithms for wearable medical monitoring, *e.g.*, for seizure onset detection, the sensors and monitoring devices must be comfortable enough for prolonged wear. To make a wearable seizure onset monitor more comfortable, both the device weight and number of scalp electrodes need to be minimized.

In this paper, we described a machine learning-based technique to automatically construct a reduced channel detector. Starting from an existing technique for constructing 18 channel patient-specific detectors, we showed how to use machine learning to automatically construct reduced channel detectors. Reducing the number of channels improves wearability since fewer electrodes are less intrusive and are easier to mount. Moreover, reducing the number of channels also reduces the energy used by the monitoring device since less data will need to be processed.

Using data from 16 different patients, we automatically constructed patient-specific reduced channel seizure detection algorithms and compared their detection performances and energy consumptions to those of the original 18 channel detector. On average, only 4.6 channels were needed. The detectors were able to detect 97% of all seizures, a degradation of only 2%. For three quarters of the patients, there was no degradation in the detection rate. While there was an increase in the average detection delay from 7.8 s to 11.2 s, the average rate of false alarms per hour decreased from 0.35 to 0.19.

We also described a prototype implementation of a single channel EEG monitoring device built using off-the-shelf components. Using this implementation, we derived an energy consumption model which was used to evaluate the energy consumed by our reduced channel detectors. On average, using fewer channels allowed our system to use 69% less energy than the 18 channel detector. This translates to an increase in battery lifetime of 3.3×. Finally, we showed how to achieve additional energy savings by using a screening detector to rule out segments of non-seizure data at low-cost. While screening does not reduce the number of channels that need to be worn, it can further reduce energy consumption by 16% while still detecting 96% of the seizures.

Finally, it is important to note that, while the focus of this paper is on selecting channels for seizure onset detection, the technique of selecting a subset of channels can be easily applied to other sensor-based domains where data from multiple sensors is used to perform detection. The effectiveness of the algorithm relies on the objective function chosen. Using an appropriate objective function to select a subset of sensors can be beneficial if resources such as energy or bandwidth are limited.

13. ACKNOWLEDGEMENTS

We would like to thank our shepherd Suman Banerjee and the anonymous reviewers for providing us with comments that helped to improve the paper. In addition, we would also like to express

our gratitude to Texas Instruments and Cathy Wicks for donating some of the hardware used to build our prototype. Finally, we thank Jenna Wiens and Dorothy Curtis for their invaluable comments on the manuscript.

14. REFERENCES

[1] Y. Agarwal, R. Chandra, A. Wolman, P. Bahl, K. Chin, and R. Gupta. Wireless Wakeups Revisited: Energy Management For VoIP over Wi-Fi Smartphones. In *MobiSys '07*, pages 179–191, 2007.

[2] A.-T. Avestruz, W. Santa, D. Carlson, R. Jensen, S. Stanslaski, A. Helfenstine, and T. Denison. A 5 μW/Channel Spectral Analysis IC for Chronic Bidirectional Brain-Machine Interfaces. *IEEE J. Solid-State Circuits*, 43(12):3006–3024, Dec 2008.

[3] A. Y. Benbasat and J. Paradiso. A Framework for the Automated Generation of Power-Efficient Classifiers for Embedded Sensor Nodes. In *SenSys '07: Proceedings of the 5th International Conference on Embedded Networked Sensor Systems*, pages 219–232, 2007.

[4] CC2500: Low-Cost Low-Power 2.4 GHz RF Transceiver (Rev. B), September 2007. Texas Instruments Datasheet.

[5] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[6] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines (and other kernel-based learning methods)*. Cambridge University Press, 2000.

[7] P. Dutta, M. Grimmer, A. Arora, S. Bibyk, and D. Culler. Design of a Wireless Sensor Network Platform for Detecting Rare, Random, and Ephemeral Events. In *IPSN '05: Proceedings of the 4th International Symposium on Information Processing in Sensor Networks*, pages 497–502, 2005.

[8] J. Flinn and M. Satyanarayanan. Energy-Aware Adaptation for Mobile Applications. In *SOSP '99*, pages 48–63, October 1999.

[9] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

[10] E. Glassman and J. V. Guttag. Reducing the Number of Channels for an Ambulatory Patient-Specific EEG-based Epileptic Seizure Detector by Applying Recursive Feature Elimination. In *IEEE EMBS 2006*, pages 2175–2178, August 2006.

[11] I. Guyon and A. Elisseeff. An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, 3:1157–1182, March 2003.

[12] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning*, 46(1–3):389–422, January 2002.

[13] R. Kohavi and G. H. John. Wrappers for Feature Subset Selection. *Artificial Intelligence*, 97(1-2):273–324, 1997.

[14] R. Kravets and P. Krishnan. Power Management Techniques for Mobile Communication. In *MobiCom '98*, pages 157–168, 1998.

[15] T. N. Lal, M. Schröder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, and B. Schölkopf. Support Vector Channel Selection in BCI. *IEEE Transactions on Biomedical Engineering*, 51(6):1003–1010, June 2004.

[16] J. R. Lorch and A. J. Smith. Operating System Modifications for Task-based Speed and Voltage Scheduling. In *MobiSys '03*, pages 215–229, May 2003.

[17] M. Malinowski, M. Moskwa, M. Feldmeier, M. Laibowitz, and J. A. Paradiso. CargoNet: A Low-cost Micropower Sensor Node Exploiting Quasi-Passive Wakeup for Adaptive Asychronous Monitoring of Exceptional Events. In *SenSys '07: Proceedings of the 5th International Conference on Embedded Networked Sensor Systems*, pages 145–159, 2007.

[18] R. Matthews, N. J. McDonald, P. Hervieux, P. J. Turner, and M. A. Steindorf. A Wearable Physiological Sensor Suite for Unobtrusive Monitoring of Physiological and Cognitive State. In *Proceedings of the 29th Annual International Conference of the IEEE*, pages 5276–5281, August 2007.

[19] MSP430xG461x Mixed Signal Microcontroller, October 2007. Texas Instruments Datasheet.

[20] Y.-T. Peng and D. Sow. Data Scaling in Remote Health Monitoring Systems. In *IEEE International Symposium on Circuits and Systems*, 2008.

[21] T. Pering, Y. Agarwal, R. Gupta, and R. Want. CoolSpots: Reducing the Power Consumption of Wireless Mobile Devices with Multiple Radio Interfaces. In *MobiSys '06*, pages 220–232, 2006.

[22] H. Qu and J. Gotman. A Patient-Specific Algorithm for the Detection of Seizure Onset in Long-Term EEG Monitoring: Possible Use as a Warning Device. *IEEE Transactions on Biomedical Engineering*, 44(2), February 1997.

[23] A. J. Rowan and E. Tolunsky. *Primer of EEG: With A Mini-Atlas*. Butterworth-Heinemann, 2003.

[24] S. U. Schuele and H. O. Lüders. Intractable epilepsy: management and therapeutic alternatives. *The Lancet Neurology*, 7(6):514–524, June 2008.

[25] E. Shih, P. Bahl, and M. J. Sinclair. Wake on Wireless: An Event Driven Energy Saving Strategy for Battery Operated Devices. In *MobiCom '02*, pages 160–171, 2002.

[26] 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.

[27] A. Shoeb, B. Bourgeois, S. Treves, S. Schachter, and J. Guttag. Impact of Patient-Specificity on Seizure Onset Detection Performance. In *IEEE EMBS 2007*, pages 4110–4114, September 2007.

[28] A. Shoeb, H. Edwards, J. Connolly, B. Bourgeois, S. T. Treves, and J. Guttag. Patient-Specific Seizure Onset Detection. *Epilepsy and Behavior*, 5(4):483–498, August 2004.

[29] A. Shoeb, S. Schachter, D. Schomer, B. Bourgeois, S. Treves, and J. Guttag. Detecting Seizure Onset in the Ambulatory Setting: Demonstrating Feasibility. In *IEEE EMBS 2005*, pages 3546–3550, September 2005.

[30] P. D. Turney. Cost-Sensitive Classification: Empirical Evaluation of a Hybrid Genetic Decision Tree Induction Algorithm. *Journal of Artificial Intelligence Research*, 2:369–409, 1995.

[31] M. van den Broek and E. Beghi. Accidents in Patients with Epilepsy: Types, Circumstances, and Complications: A European Cohort Study. *Epilepsia*, 45(6):667–672, 2004.

[32] N. Verma, A. Shoeb, J. Guttag, and A. Chandrakasan. A Micro-power EEG Acquisition SoC with Integrated Seizure Detection Processor for Continuous Patient Monitoring (To appear). In *Proceedings of the 2009 Symposium on VLSI Circuits*, June 2009.

[33] E. Waterhouse. New Horizons in Ambulatory Electroencephalography. *Engineering in Medicine and Biology Magazine, IEEE*, 22(3):74–80, May-June 2003.