Supplementary Information

Patients and Seizures

Over a period of eight months, we enrolled patients with epilepsy who were admitted to the long-term video-telemetry monitoring (LTM) unit at Children's Hospital Boston for assessment. The patients stayed at the LTM for 1-7 days for characterization of events or localization of the seizure onset zone in the setting of pre-surgical evaluation. Electroencephalography (EEG) recordings were performed using conventional scalp electrodes (10-20 system) at a sampling rate of 256 Hz or implanted intracranial electrodes at a sampling rate of 500 Hz (XLTEK, Oakville, ON, Canada). The wrist-worn biosensors were placed on both wrists such that the electrodes were in contact with the ventral side of the forearms. Electrodermal activity (EDA) and accelerometry (ACM) recordings were sampled at 20 Hz and synchronized with the video-EEG recordings by generating technical artifacts at the beginning and end of each session for offline realignment. Each recording session lasted approximately 24 hours and batteries were replaced on a daily basis.

Wrist-worn EDA and ACM Biosensor

The design and construction of the MIT custom-built wearable EDA and ACM biosensor has been previously described (Poh et al., 2010). Briefly, exosomatic EDA was quantified as skin conductance by applying direct current to the stratum corneum of the epidermis beneath measuring electrodes. The biosensor module also contains a three dimensional accelerometer for measurements of physical activity. A microcontroller digitizes the analog signals via a 12-bit A-D and the data is written to an onboard microSD card. To

achieve a wide dynamic range of skin conductance measurements, the analog conditioning circuitry utilizes non-linear feedback automatic bias control with low-power operational amplifiers.

We integrated the sensor module into a regular wristband made out of terrycloth, resulting in a comfortable and lightweight wearable sensor. Since all electronics and wiring are concealed within the wristband, the resulting device is inconspicuous, non-stigmatizing and allows for discrete monitoring of EDA. Furthermore, the electronic module can be easily detached when the user desires to wash the wristband. We used dry Ag/AgCl disc electrodes with contact areas of 1.0 cm^2 for our recordings as recommended in the literature (Fowles et al., 1981). These electrodes are disposable and can be snapped onto or removed from the wristband with ease. Although the electrodes are commonly placed on the palmar surface of the hand (e.g. medial and distal phalanges of the fingers and the thenar and hypothenar eminences), we used the ventral side of the distal forearms as recording sites. Placement of electrodes on the forearm are less susceptible to motion artifacts and highly correlated to palmar recordings (Poh et al., 2010). A 3.7 V lithium polymer battery with a capacity of 1100 mAh provides around 40 hours of operation. The battery can be recharged via a micro-USB cable.

EEG/ACM/EDA Analysis

Ictal video-EEG recordings were retrospectively reviewed by two board-certified clinical neurophysiologists who were blinded to the EDA data. Each GTC seizure was reviewed for EEG and clinical ictal onset and offset times, as well as EEG localization and seizure semiology on video recordings. EDA and ACM recordings were analyzed using custom

written software in MATLAB (The Mathworks, Inc.). Only recordings from one wrist-worn biosensor were utilized for seizure detection (default choice was the right forearm unless data was unavailable or corrupted).

Data Preprocessing/Reduction

A GTC seizure typically lasts for 1-2 minutes whereas the patients were monitored continuously throughout their stay (days) in the hospital. As such, there is a vast amount of non-seizure data (forming the majority class), which causes the data set to be highly imbalanced. Thus, pre-processing of the data is important to decrease the computational workload as well as reduce the degree of data imbalance during supervised learning.

The first step was to divide the data into non-movement and movement events. We combined information from all three axes of the accelerometer to calculate the magnitude of the net acceleration, a as:

$$a = \sqrt{a_x^2 + a_y^2 + a_z^2}$$

A sliding window of 10 seconds with 75% overlap was used to calculate the standard deviation, σ of the acceleration epoch $\left(a_1, a_2, \ldots, a_N\right)$:

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(a_i - \mu\right)^2}$$

where

$$\mu = \frac{1}{N}\sum_{i=1}^{N} a_i$$

Epochs with σ below 0.1 g were automatically discarded from further analysis and treated as non-motor, and hence non-seizure events. The remaining epochs were detrended using a smoothness priors approach with the smoothing parameter $\lambda = 300$.

(Tarvainen et al., 2002) and the discrete Fourier transform (DFT) was computed (200-point).

Generalized tonic-clonic (GTC) seizures are composed of two primary phases -- the tonic phase and the clonic phase. The tonic phase involves stiffening of the limbs and flexion or extension of the neck, back and extremities. During the clonic phase, muscles of the entire body start to contract and relax rapidly. These convulsions are manifest in the ACM signal as rhythmic activity typically above 2 Hz. Thus, each epoch was evaluated for important periods using an algorithm by Vlachos and colleagues (Vlachos et al., 2004). The underlying assumption is that the magnitudes of the coefficients of the DFT of a non-periodic time series are distributed according to an exponential distribution.

$$f(x) = \lambda e^{\lambda x}$$

Important periods will have powers that deviate from the power content of the majority of the periods and can be identified by locating outliers according to an exponential distribution. As a result, we seek for infrequent powers by setting the probability p to a very low value to derive a power threshold T_p.

$$p = P(x \geq T_p) = e^{-\lambda T_p}$$

Solving for the power threshold,

$$T_p = -\frac{\ln(p)}{\lambda}$$

For 99% confidence, we set $p = 0.01$ and λ is the reciprocal of the mean of the detrended acceleration signal power.

$$\frac{1}{\lambda} = \frac{1}{N}\sum_{i=1}^{N} a_i^2$$

Epochs with no frequency components that exceeded T_p were discarded and labeled as non-seizure events. Otherwise, $f_{dominant}$, the frequency component with the highest power beyond T_p was identified. If $f_{dominant} \geq 2$ Hz, the epoch was accepted for subsequent feature extraction.

Feature Extraction

A total of 19 features were computed to characterize each 10 s measurement epoch. These features were chosen to describe the time, frequency and phase space characteristics of the ACM signal (16 features) as well as the temporal traits of the EDA signal (3 features).

Time-domain Analysis (ACM)

To quantify the time-domain attributes of the ACM signal, we computed four different features. We calculated the mean, standard deviation, and root mean-squared of the net acceleration. In addition, we estimated the amount of force by accumulating the magnitude of accelerometer data a_{mag} from each axis throughout the 10 s epoch.

$$a_{mag} = \int_n^{n+\Delta} |a_x(t)| + |a_y(t)| + |a_z(t)| \, dt$$

Spectral Analysis (ACM)

The major energy band for daily activities falls between 0.3 and 3.5 Hz (Sun and Hill 1993) whereas during GTC seizures the power is typically concentrated at frequencies above 2 Hz (Fig. S3). To capture the spectral information of the net acceleration, we detrended the net acceleration using a smoothness priors approach with the smoothing

parameter $\lambda = 300$ (Tarvainen et al., 2002) and computed the power spectral density using Welch's method (8 segments of equal length, 50% overlap, Hamming window). The entire frequency spectrum was divided into eight non-overlapping bands and the total integrated power within each spectral band was included as a feature (8 features). The dominant frequency within each epoch (across the entire 0 to 10 Hz band) along with its maximum power were also computed as features (2 features). Thus, a total of 10 ACM spectral features were included for classification.

Non-linear Analysis: Recurrence Quantitative Analysis (ACM)

Recurrence plots provide a graphical method designed to locate hidden recurring patterns and compute non-linear dynamical measures (Eckmann et al., 1987). This technique allows signals to be represented in state (phase) space by constructing embedded vectors $\vec{\mathbf{x}}(k)$ using the method of time delays:

$$\vec{\mathbf{x}}(k) = \left[x(k), x(k+d), \ldots, x(k+(m-1)d)\right]^T$$

where m is the embedding dimension and d is the time delay.

Recurrence analysis was performed using the Recurrence Plot toolbox for Matlab (Marwan et al., 2007). The optimal parameter $m = 5$ was chosen as the embedding dimension where the amount of false nearest neighbors approached zero (Kennel et al., 1992). The delay $d = 1$ was calculated from the first minimum of the mutual information function (Fraser and Swinney 1986). The recurrence plot $\mathbf{R}(i,j)$ was then constructed by computing distances between all pairs of embedded vectors; a critical radius $\varepsilon = 1$ was established to create a binary plot showing, for a given moment in time, the times at which the state space trajectory visited roughly the same area in the state space.

$$\mathbf{R}(i,j) = \Theta\left(\varepsilon - \left\|\vec{\mathbf{x}}(i) - \vec{\mathbf{x}}(j)\right\|\right)$$

where $\Theta(x)$ is the Heaviside step function.

An example of a recurrence plot constructed from a seizure epoch is shown in Fig. S4. The short line segments parallel to the main diagonal suggest that the time series is deterministic. These small scale structures were quantified using recurrence quantification analysis (Marwan et al., 2007). Other examples of recurrences plots from various daily activities are shown in Fig. S4B-D. The first feature included was the Shannon entropy *ENTR* of the lengths of the diagonal lines, which reflects the complexity of the deterministic structure in the system.

$$ENTR = -\sum_{l=l_{\min}}^{N} p(l) \ln p(l)$$

where $p(l)$ is the probability that a diagonal line has exactly length l estimated from the histogram $P(l)$ of the lengths l of the diagonal lines.

$$p(l) = \frac{P(l)}{\sum_{l=l_{\min}}^{N} P(l)}$$

The second feature computed was laminarity *LAM*, the percentage of recurrence points which formed vertical lines. *LAM* is related with the amount of laminar phases in the system (intermittency).

$$LAM = \frac{\sum_{v=v_{\min}}^{N} vP(v)}{\sum_{v=1}^{N} vP(v)}$$

where $P(v)$ is the histogram of the lengths v of the vertical lines.

To summarize, 16 ACM features were computed including 4 time-domain (mean, standard deviation, root mean-squared and accumulated magnitude), 10-secondpectral

(dominant frequency, maximum power, and integrated power values from 8 non-overlapping frequency bands) and 2 non-linear features (entropy and laminarity).

EDA Analysis

Since GTC seizures are associated with an increase in EDA, we included three features extracted from each 10-second EDA epoch. First, the EDA recordings were low-pass filtered (Hamming window, length = 1025, 3 Hz) to reduce artifacts. We performed a linear least squares fit to the EDA segment and computed the slope as the first feature. The number of measurement points within the epoch that were greater than the previous point (i.e. $x(n) > x(n+1)$) was determined as the second feature. The third feature corresponded to the difference between the EDA measured at the start and end of the 10-second epoch.

Support Vector Machines

Support Vector Machines (SVMs) are state-of-the-art binary classification methods that usually exhibit good resistance to overfitting and have shown excellent performance in complicated pattern recognition problems (Brown et al., 2000, Guyon et al., 2002). An SVM can learn a decision boundary in the form of a hyperplane that separates two classes. This hyperplane is selected such that the classification margin, which is the geometric distance between the hyperplane and the boundary cases of each class (i.e. the support vectors), is maximized (Noble 2006, Vapnik 2000). Moreover, SVMs can map the original finite dimensional feature space into a much higher dimensional space through the use of a kernel function to improve the separability of the data. We chose the

Gaussian Radial Basis kernel function (RBF) as it provides non-linear mapping of the original feature vectors into a higher dimensional space.

An SVM is a good choice for the task of seizure detection because its unique learning mechanism allows it to perform well with moderately imbalanced data without any modifications (Akbani et al., 2004). Since an SVM only takes into account those instances that are close to the boundary for building its model, it is unaffected by negative instances far away from the boundary even if they are large in number. This is important given that the number of non-seizure instances far outnumber the seizure instances.

Model Selection and Testing Methodology

SVMs were implemented using LibSVM, a publicly available software library for support vector classification (Chang and Lin 2001). The problem of seizure detection was posed as a supervised learning task in which the goal was to classify each 10-second epoch as seizure or non-seizure based on extracted features from EDA and ACM recordings. If any epoch between the start and end of a labeled seizure was correctly classified as a seizure event, the seizure was considered detected (true positive). If multiple epochs within the seizure duration were detected, these were treated as a single true positive. False positives that occurred within 30 seconds apart from each other were treated as a single false alarm.

For evaluation, we employed a *double cross-validation* method, where the data were divided into a training set and a test set, and the training set was further subdivided into a learning set and a validation set (Cherkassky et al., 2007). The SVM classification model was trained on data from the learning and validation sets, and then tested on a

dataset that was not touched in training. Each feature in the training data was linearly scaled to the range [0, 1] to assure commensurability of the various features before applying SVM. The same scaling template was applied to the testing data before performing classification. The SVM model required two parameters to be optimized: the penalty ("soft margin") parameter of the error term *C* that specifies the trade-off between maximizing the classification margin and minimizing the training error, and the RBF kernel parameter γ that controls the curvature of the hyperplane. The optimal values of *C* and γ are typically selected based on the best cross-validation accuracy. However, in learning imbalanced data, the overall classification accuracy is not an appropriate measure of performance since a trivial classifier that predicts every instance as the majority class (non-seizure) would achieve very high accuracy but be of little use. As such, we used the F-measure to evaluate the performance of the SVM while searching for the optimal pair of *C* and γ.

$$\text{F} - \text{measure} = \frac{2 \cdot \text{true positive}}{2 \cdot \text{true positive} + \text{false negative} + \text{false positive}}$$

Non-Patient-Specific Seizure Detection

For practical use, the generic mode is important for a seizure detector to be of immediate use to every patient right "out of the box". As more examples of seizures are obtained over time, the algorithm can then improve and become more customized for each individual. We first examined the performance of a *non-patient-specific* or generic seizure detection algorithm that excludes all data from a test patient in the training phase. Since limited data was available, we employed a double *leave-one-patient-out* cross-validation strategy.

The total number of patients with recorded GTC seizures is denoted by M_{SZ}. In this approach, the entire recording session from a single patient P_{test} was set aside for the testing phase. The remaining data from the other $M_{SZ}-1$ patients was included in the training phase. The goal of the training phase was to optimize the values of *C* and γ for model selection. As it is not known beforehand which values of *C* and γ are best for a given problem, we performed a grid-search to evaluate various combinations of the parameters ($C = 2^{-5}, 2^{-3}, \ldots, 2^{11}; \gamma = 2^{-15}, 2^{-13}, \ldots, 2^{-1}$). To avoid selection of SVM parameters that would result in overfitting to the training set, the training data was subdivided into two parts, of which data from a single patient was considered "unknown" (validation set). Sequentially data from each "unknown" patient were tested using the classifier trained on the remaining $M_{SZ}-2$ patients (learning set). As such, we performed ($M_{SZ}-1$)-fold cross-validation for each possible combination of *C* and γ. The pair of *C* and γ that produced the highest average cross-validation F-measure was selected. Using the selected parameter values, the whole training set (data from all $M_{SZ}-1$ patients) was trained again to generate the final classifier and tested on the data from the single patient P_{test} that had been set aside. This entire process was repeated with a re-training of the SVM in each round, such that data from each patient were excluded once for testing, and the remainder used as training data (M_{SZ}-fold cross-validation). Overall performance was determined by taking the average of the performance of all tested patients.

Semi-Patient-Specific Seizure Detection

Since the GTC seizure manifestation in ACM and EDA signals may vary from patient to patient, it is reasonable that an adaptive approach which takes advantage of the

consistency of an individual patient's unique seizure signature as well as typical non-seizure activity patterns improves performance even further. To allow the SVM to learn from previous examples of seizures from the test patient if that patient had more than a single GTC seizure recording available, we also implemented a double *leave-one-seizure-out* cross-validation strategy to assess the performance of our seizure detection algorithm. As the detector was not trained solely on data from a particular test patient but included examples from all other patients, this approach was *semi-patient-specific*.

We divided the data from the M_{SZ} patients into N_{SZ} seizure recordings and N_{NS} non-seizure recordings ($N_{NS} = 2N_{SZ}$). In this approach, one seizure recording R_{testSZ} along with two non-seizure recordings R_{testNS} were set aside for testing. The remaining data from the other $N_{SZ} - 1$ seizure and $N_{NS} - 2$ non-seizure recordings were used for training the SVM. A grid-search was performed to evaluate various combinations of the parameters ($C = 2^{-5}, 2^{-3}, \ldots, 2^{11}; \gamma = 2^{-15}, 2^{-13}, \ldots, 2^{-1}$). To avoid overfitting to the training set, the training data was again separated into two parts, of which data from a single seizure recording together with two non-seizure recordings were considered "unknown". Sequentially data from each "unknown" seizure and non-seizure recordings were tested using the classifier trained on the remaining $N_{SZ} - 2$ seizure and $N_{NS} - 4$ non-seizure recordings. As such, we performed ($N_{SZ} - 1$)-fold cross-validation for each possible combination of C and γ. The pair of C and γ that produced the highest average cross-validation F-measure was selected. Using the selected parameter values, the whole training set (data from all $N_{SZ} - 1$ seizure and $N_{NS} - 2$ non-seizure recordings) was trained again to generate the final classifier and tested on the data from the single seizure recording R_{testSZ} and two non-seizure recordings R_{testNS} that had been withheld. This

entire process was repeated with a re-training of the SVM in each round, such that data from each seizure and non-seizure recording were excluded once for testing, and the remainder used as training data N_{SZ}-fold cross-validation). Overall performance was determined by taking the average of the performance of all tested recordings.

Generalization of False Alarm Rate to New Group of Patients

To evaluate the generalization of the seizure detection algorithm in terms of its false alarm rate, we utilized recordings from M_{NS} patients who did not experience GTC seizures for testing. Data from all the M_{SZ} patients with GTC seizures were used to train the SVM and tested on the recordings from all M_{NS} patients that had been withheld.

We utilized recordings from M_{NS} patients who did not experience GTC seizures for testing the generalization of our false alarm rate. We used data from all the M_{SZ} patients with GTC seizures to train the SVM classifier. A grid-search was performed to evaluate various combinations of the parameters ($C = 2^{-5}, 2^{-3}, \ldots, 2^{11}; \gamma = 2^{-15}, 2^{-13}, \ldots, 2^{-1}$). To avoid overfitting, the leave-one-seizure-out approach was employed such that N_{SZ}-fold cross-validation was performed for each possible combination of C and γ. The pair of C and γ that produced the highest average cross-validation F-measure was selected. Using the selected parameter values, the whole training set (entire data set from M_{SZ} patients with GTC seizures) was trained again to generate the final classifier and tested on the recordings from all M_{NS} patients that had been withheld.

Comparison With Previous Work

Previous work on non-EEG based wearable GTC seizure detectors have focused primarily on ACM as the input signal. Table S1 provides a comparison of the key features for previous ACM-based seizure detectors. Schulc and colleagues utilized a Wii Remote mounted on the forearm using a stocking to detect four GTC seizures from three patients based on a visually determined intensity and duration threshold of the ACM signal (Schulc, et al. 2011). However, the results are not directly comparable and have to be taken with caution because independent assessment of the algorithm was not performed on separate test data to determine whether the perfect sensitivity achieved was due to overfitting. Two commercial wrist-worn motion sensors, the SmartWatch and EpiLert have been evaluated in clinical studies (Kramer, Kipervasser, Shlitner and Kuzniecky 2011, Lockman, Fisher and Olson 2011). The SmartWatch produced a large number of false alarms whereas the EpiLert had a very low rate of false detections (8 in 1692 hours). One drawback of the EpiLert study is that patients expected to have a high incidence of seizure-like movements such as dystonic posturing, subtle behavioral automatism and pseudoseizures were excluded from testing, which could explain the low false alarm rate. These devices also employed proprietary algorithms that were not described. Our proposed seizure-detection methodology was tested on the largest number of patients and the results compare favorably with these studies.

References

Akbani R, Kwek S, Japkowicz N. (2004) Applying support vector machines to imbalanced datasets. *Proc Eur Conf Mach Learn* 2004:39-50.

Brown MPS, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, Ares M, Haussler D. (2000) Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Nat Acad Sci USA* 97:262.

Chang CC, Lin CJ. (2001) LIBSVM: a library for support vector machines. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Cherkassky V, Cherkassky VS, Mulier F. (2007) *Learning from data: Concepts, theory, and methods*. Wiley-IEEE Press, New York.

Eckmann JP, Kamphorst SO, Ruelle D. (1987) Recurrence plots of dynamical systems. *Europhys Lett* 4:973.

Fowles DC, Christie MJ, Edelberg R, Grings WW, Lykken DT, Venables PH. (1981) Publication recommendations for electrodermal measurements. *Psychophysiology* 18:232-239.

Fraser AM, Swinney HL. (1986) Independent coordinates for strange attractors from mutual information. *Phys Rev A* 33:1134.

Guyon I, Weston J, Barnhill S, Vapnik V. (2002) Gene selection for cancer classification using support vector machines. *Mach Learn* 46:389-422.

Kennel MB, Brown R, Abarbanel HDI. (1992) Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys Rev A* 45:3403.

Kramer U, Kipervasser S, Shlitner A, Kuzniecky R. (2011) A Novel Portable Seizure Detection Alarm System: Preliminary Results. *J Clin Neurophys* 28:36-38.

Lockman J, Fisher RS, Olson DM. (2011) Detection of seizure-like movements using a wrist accelerometer. *Epilepsy Behav* 20:638-641

Marwan N, Carmen Romano M, Thiel M, Kurths J. (2007) Recurrence plots for the analysis of complex systems. *Physics Reports* 438:237-329.

Noble WS. (2006) What is a support vector machine? *Nat Biotechnol* 24:1565-1568.

Poh MZ, Swenson NC, Picard RW. (2010) A Wearable Sensor for Unobtrusive, Long-term Assessment of Electrodermal Activity. *IEEE Trans Biomed Eng* 57:1243-1252.

Schulc E, Unterberger I, Saboor S, Hilbe J, Ertl M, Ammenwerth E, Trinka E, Them C. (2011) Measurement and quantification of generalized tonic-clonic seizures in epilepsy patients by means of accelerometry--An explorative study. *Epilepsy Res* 95:173-183.

Sun M, Hill J. (1993) A method for measuring mechanical work and work efficiency during human activities. *J Biomech* 26:229-241.

Tarvainen MP, Ranta-Aho PO, Karjalainen PA. (2002) An advanced detrending method with application to HRV analysis. *IEEE Trans Biomed Eng* 49:172-175.

Vapnik VN. (2000) *The nature of statistical learning theory*. Springer Verlag, New York.

Vlachos M, Meek C, Vagena Z, Gunopulos D. (2004) Identifying similarities, periodicities and bursts for online search queries. ACM, pp. 131-142.