Multi-center clinical assessment of improved wearable multimodal convulsive seizure detectors

Research Article

Multi-center clinical assessment of improved wearable multimodal convulsive seizure detectors

Francesco Onorati, PhD, §[1] Giulia Regalia, PhD, §*[1] Chiara Caborni,[1] Matteo Migliorini, PhD,[1] Daniel Bender,[1] Ming-Zher Poh, PhD,[2] Cherise Frazier, MD,[3] Eliana Kovitch Thropp, CPNP,[4] Elizabeth D. Mynatt, PhD, Exec Dir,[5] Jonathan Bidwell,[3,4,5] Roberto Mai, MD,[6] W. Curt LaFrance, Jr, MD, MPH,[7] Andrew S. Blum, MD, PhD, [8] Daniel Friedman, MD,[9] Tobias Loddenkemper, MD,[10] Fatemeh Mohammadpour-Touserkani, MD, [10] Claus Reinsberger, MD, PhD,[11] Simone Tognetti, PhD,[1] and Rosalind W. Picard, PhD [1,2]

1. Empatica Inc., Milan, Italy, and Cambridge, MA
2. MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA
3. Emory University Hospital Midtown, Atlanta, GA
4. Children's Healthcare of Atlanta (CHOA), Atlanta, GA
5. Georgia Institute of Technology, Atlanta, GA
6. "Claudio Munari" Epilepsy Surgery Center, Ospedale Niguarda, Milan, Italy
7. Division of Neuropsychiatry and Behavioral Neurology, Rhode Island Hospital, Brown University, Providence, RI
8. Department of Neurology, Rhode Island Hospital, Brown University, Providence, RI
9. Department of Neurology, NYU Langone Medical Center, New York, NY
10. Department of Neurology, Boston Children's Hospital, Boston, MA
11. Department of Neurology, Brigham and Women's Hospital, Boston, MA

§ Equally contributing authors

* Corresponding author: Giulia Regalia
Corresponding author's address: via Stendhal 36, 20144, Milan, Italy
Corresponding author's phone and fax: +39 236566106
Corresponding author's e-mail address: gr@empatica.com

Running title: Wearable multimodal motor seizure detectors

Summary

Objective:

New devices are needed for monitoring seizures, especially those associated with sudden unexpected death in epilepsy (SUDEP). They must be unobtrusive, automated and provide false alarm rates bearable in everyday life. This study quantifies the performance of new multimodal wrist-worn convulsive seizure detectors.

Methods:

Hand-annotated video-electroencephalography seizure events were collected from 69 patients at 6 clinical sites. Three different wristbands were used to record electrodermal activity (EDA) and accelerometer (ACM) signals, obtaining 5,928 hours of data, including 55 convulsive epileptic seizures (6 focal tonic-clonic seizures and 49 focal-to-bilateral-tonic-clonic seizures) from 22 patients. Recordings were analyzed off-line to train and test two new machine learning classifiers and a published EDA and ACM-based classifier. Moreover, wristband data were analyzed to estimate seizure-motion duration and autonomic responses.

Results:

The two novel classifiers consistently outperformed the previous detector. The most efficient (Classifier III) yielded sensitivity of 94.55%, and false alarm rate (FAR) of 0.2 events/day. No nocturnal seizures were missed. Most patients had less than 1 false alarm every 4 days with FAR below their seizure frequency. When increasing the sensitivity to 100% (no missed seizures) the FAR is up to 13 times lower than the previous detector. Furthermore, all detections occurred before the seizure ended, providing reasonable latency (median: 29.3 s, range: 14.8-151 s). Automatically estimated seizure durations were correlated with true durations, enabling reliable annotations. Finally, EDA measurements confirmed the presence of post-ictal autonomic dysfunction, exhibiting a significant rise in 73% of the convulsive seizures.

Significance:

The proposed multimodal wrist-worn convulsive seizure detectors provide seizure counts that are more accurate than previous automated detectors and typical patient self-reports, while maintaining a tolerable FAR for ambulatory monitoring. Furthermore, the multimodal system provides an objective description of motor behaviour and autonomic dysfunction, aimed at enriching seizure characterization, with potential utility for SUDEP warning.

Key Words:

Epilepsy, Convulsive seizures, Electrodermal activity, Machine learning

1. Introduction

Epilepsy is among the most common neurological disorders, with an estimated 65 million patients worldwide[1]. While rare, "sudden unexpected death in epilepsy" (SUDEP) is the most common cause of death in epilepsy[1]. SUDEP is more likely to occur in patients who have at least one (primary or secondarily) generalized tonic-clonic (GTC) seizure a year, and when a patient is unattended after a seizure[2-3]. While SUDEP's general cause remains unknown, SUDEP can occur after prolonged post-ictal generalized EEG suppression (PGES), and is associated with autonomic dysfunction such as terminal apnea preceding terminal asystole[4,5,6]. It is crucial to develop systems to detect seizures, to measure possible biomarkers of SUDEP, and to alert caregivers for assistance, as an early application of aid can be protective[6,7].

The gold standard for monitoring seizures is video-electroencephalography (v-EEG) in epilepsy monitoring units (EMUs), an impractical procedure for long-term use. Moreover, patients may experience seizures with different semiology, or may not experience any during admission. Today's clinical trials rely on seizure counts and symptoms observed by patients/caregivers, despite that self-reported counts are often inaccurate[8], especially during sleep[9].

Wearable automated seizure detectors may improve existing practice by providing continuous ambulatory monitoring, potentially more accurate seizure counts, and alerts for early intervention[10,11,12]. Existing automated seizure detectors[11] measure motion to detect seizures with a motor manifestation. Algorithms based on wrist acceleration (ACM)[13,14,15] or electromyogram (EMG)[16,17] have been commercialized into the SmartWatch[13,18], Epi-Care Free watch[15,19], Epilert[14], Brain Sentinel[16] and EDDI alarm[17]. Except for two studies[15,18], most algorithms have been tested on relatively small datasets, (number of seizures and recording hours), which prevents robust estimates of sensitivity and false alarm rates. Rarely are

objective characterizations about seizure events, beyond seizure counts, provided to the patient or clinician[11].

While only small studies have been performed to date, multimodal systems (e.g. combining ACM with EMG[20,21] or with electrodermal activity, EDA[22]) have shown increased sensitivity with reduced false alarms[10,12]. Moreover, physiological parameters may be useful to assess SUDEP risk; for example, the amplitude of EDA accompanying GTC seizures has been shown to correlate to the duration of PGES[23].

In this work we started from a pioneering study on secondarily GTC, i.e. focal motor to bilateral tonic-clonic seizures (FTCb), showing that combining EDA with ACM leads to more sensitive and specific detection than ACM alone[22]. The combination takes advantage of the detection, by a comfortably-worn wristband, of a wide range of motor seizures using ACM sensors - tonic-clonic, tonic, clonic, myoclonic, hypermotor[24] - and of the measurement of the sympathetic nervous system activity using EDA[25], including peri-ictal autonomic dysregulation[6,23,26]. The primary contribution of this work is two improved detection algorithms trained on a significantly larger dataset containing focal motor tonic-clonic (FTC) and FTCb seizures (hereafter referred to as CS, convulsive seizures). The secondary contribution is a new automated ability to quantify each seizure's autonomic dysfunction and motor duration to help objectively characterize seizures and possible biomarkers of SUDEP.

2. Methods

Patients

69 patients diagnosed with epilepsy (24 children, ages 4-18 years, median: 14 years, 9 females; 45 adults, ages: 19-60 years, median: 37 years, 28 females) were admitted for v-EEG monitoring at 6 clinical sites: Children's Hospital Boston (14 patients), New York University Langone Medical Center (18 patients), Rhode Island Hospital (5 patients), Emory University

Hospital (15 patients) and Children's Healthcare of Atlanta (CHOA), Egleston and Scottish Rite hospitals (12 patients) and Niguarda Hospital (5 patients). The study was approved by their Institutional Review Boards, and participants (or their caregivers) provided written informed consent. Two board-certified clinical neurophysiologists at each clinic examined the v-EEG recording and labeled the EEG seizure onset time and the duration of clinically relevant variations in the v-EEG signals. Seizure terminology was revised according to the International League Against Epilepsy (ILAE) seizure classification[27].

Wrist acceleration and electrodermal activity recordings

During v-EEG monitoring, patients wore one of three wristbands measuring ACM and EDA, synchronized with the v-EEG at the start of each monitoring period. Wristbands included the E3 and E4 (Empatica Inc.), and the iCalm (MIT Media Lab), all featuring comparable embedded 3-axis ACM sensors and EDA sensors placed on the ventral side of the forearm. If seizure semiology reported an asymmetric involvement of arms, the wristband was placed on the wrist where convulsions appeared earlier and/or were more evident; otherwise the device was usually worn on the non-dominant arm. Five patients wore devices on both wrists. The resulting dataset consisted of 246 days (5,928 hours, hours per patient: median=74.3, range=3.5-386.8) of ACM and EDA measurements.

Seizure detection: development of automated classifiers based on wristband data

The next step was to build an automated classifier that could detect whether or not an ACM and EDA measurement exhibited seizure patterns. The process is depicted in Fig. 1. Starting from the feature set introduced in prior work[22], a feature set derived from time-domain, frequency-domain and nonlinear analyses was constructed. The features were computed on 10-second epochs of 3-axis ACM, ACM magnitude, and EDA signals, with 75% overlap of epochs[22]. Three different feature sets were extracted: (I), the set employed in Poh's study (19 features, 16 ACM and 3 EDA features)[22]; (II), a larger set (46 features, 40 ACM and 6 EDA

features); and (III), a reduced set (25 of the 46 features, 22 ACM and 3 EDA features). Feature set III was built to maximize classifier performance and minimize computational cost for future real-time implementation. On each feature set, a supervised machine learning classifier was built in order to classify each epoch as seizure or non-seizure. All signals were analyzed off-line using MATLAB (MathWorks Inc.). To simulate on-line detection, all "future-time" datapoints were hidden.

Seizure detection: performance assessment

A double cross-validation approach was adopted to test the three classifiers[22]. We split the dataset into three non-overlapping parts, each part containing epochs from 1/3 of the patients experiencing seizures and 1/3 of the patients without seizures. Two parts were used for training and tuning a nonlinear support vector machine using a leave-one-seizure-patient-out cross-validation. The held-out third part was used as a testing set. This procedure was repeated three times, i.e. holding out a different third part in each tournament. Thus performances could be evaluated on the whole dataset (5,928 hours), using no data for both training/tuning and testing at the same time.

For performance evaluation, we considered non-overlapping segments labeled as seizure, defined as intervals between the clinical onset and the clinical offset according to v-EEG labelling, and non-seizure segments, defined as intervals not including seizure events. To facilitate comparisons and performance computation, non-seizure segments were split into sub-segments with a duration equal to the mean duration of seizure segments, in order to deal with non-seizure events of approximately the same length as seizure events. This procedure allowed for better estimates of true negatives, a complicated task for systems trained to detect only the event of interest[28]. The number of seizure and non-seizure segments containing at least one alarm were accumulated for each of the three held-out third parts of data, obtaining the total number of true positives and false positives across all 5,928 hours (247 days). Table

S1 details the results from each held-out third of the data (i.e. each tournament) and the cumulative performances we report in Results. Note that our reported Results are more conservative than if the performances of the three cross-validation runs were simply averaged.

Sensitivity (Sens) was obtained by dividing the total number of true positives (accumulated over the held-out thirds) by the total number of seizures. False Positive Rate (FPR) was obtained by dividing the total number of false positives by the total number of non-seizure segments (Equivalently: 100% - Specificity). The False Alarm Rate (FAR) was computed as the total number of false alarms divided by the total of 247 days. The resulting (FPR, Sens) and (FAR, Sens) pairs corresponding to different values of the classifier decision threshold were used to build Receiver Operating Characteristic (ROC) curves (Fig. 2). The area under the curve (AUC) was computed on (FPR, Sens) ROC curves[29]. The optimal decision threshold was selected in order to provide the highest Sens with the lowest FAR, i.e. the point closest to the upper left corner in the (FAR, Sens) ROC. To statistically compare the classifiers, 95% confidence intervals (CI) for the Sens, the difference in Sens (ΔSens), the FPR, the difference in FPR (ΔFPR)[30] and the AUC[31] were used.

Additional performance metrics included: the number of seconds between the seizure clinical onset and the classifier detection time (seizure detection latency); the number of detected seizures with respect to the total number of alarms (precision); the weighted mean between sensitivity and precision[32] (F-score), and the ratio between FAR and seizure rate (SR), both measured per day. The number of false alarms triggered during resting/sleeping periods was determined applying a rest detection algorithm to ACM measurements[33].

Seizure characterization: estimating motion duration and post-ictal EDA response

To estimate the seizure motor duration, an on-line algorithm was implemented to designate a neighborhood of "where the standard deviation of the ACM was above 0.05 g". Pearson's correlation was then performed between these estimated durations and the durations assessed

by v-EEG labeling.

To quantify each seizure's peri-ictal EDA response (EDR), peri-ictal EDA recordings were segmented from 60 minutes before v-EEG seizure onset to 120 minutes afterward. A "significant EDR" was identified when EDA increased more than twice the standard deviation of the pre-ictal baseline[23,34]. The EDR was considered ended when the EDA fell below 10% of the EDA peak value. Significant EDRs were analyzed in terms of the peak amplitude with respect to the baseline, their response latency (i.e., the difference between the starting time of the EDR and the v-EEG onset), the response duration (i.e., the difference between the starting time and the ending time of the EDR), and the natural logarithm of the area under the curve of the rising phase from the starting time to the peak of the EDR, and of the total response from the starting time to the ending time of the EDR, called respectively LogAUCrise and LogAUCtot. These features were computed only for significant EDRs. Comparisons of pre-ictal vs post-ictal measurements were performed through a two-sample Kolmogorov-Smirnov test. To account for multiple comparisons, the resulting *p*-values were adjusted through the false discovery rate (FDR) procedure[23].

3. Results

Seizure data collected with the ACM and EDA wristbands

Of 69 patients, 22 experienced at least one CS during their admission, providing a total of 55 recorded CSs, including 6 FTC and 49 FTCb. None of the captured seizures were nonepileptic. 32 CSs (12 patients) were recorded with Empatica E4, 9 CSs (4 patients) with Empatica E3, and 14 CSs (6 patients) with the MIT iCalm. A more detailed summary of the recordings is in Table 1. 135 seizures other than FTC and FTCb were recorded and are not considered in this work. Individual ACM magnitude and EDA signals during the peri-ictal period are shown in Figs. S1 and S2, respectively. Since patients were not confined to bed during the monitoring period, recorded data contain activities in the clinical environment that

involve convulsive-like movements of the wrist such as brushing teeth, eating, and washing. Additionally, some patients at Emory Healthcare, while being monitored during admissions, engaged in dancing (non-seizure related).

Performance comparison of convulsive seizure detectors

A trade-off exists between Sens and FAR, which can be described as follows: consider the case of a detector that outputs "there is a seizure" at every moment. This detector will never miss a single seizure and will obtain Sens=100%; however, getting an alert every moment would be insufferable. Dually, if a detector outputs "there is no seizure" at every moment, it will have FAR=0, but it will also miss all the seizures (Sens=0%). Therefore, we compared classifiers by means of ROC curves, which quantify the trade-off between Sens and FAR, to maximize the Sens while minimizing the FAR. Overall, Classifier II and Classifier III have ROC curves that lie above Classifier I's ROC curve (Fig. 2), thus outperforming the previously published classifier. Comparisons between the AUC values of Classifier I and Classifier II (ΔAUC=0.0691, p=0.015) and between the AUC values of Classifier I and Classifier III (ΔAUC=0.0728, p=0.012) demonstrate statistically higher AUC values for the two novel classifiers (Table 2). At high levels of Sens, Classifier II and Classifier III achieved a FAR almost one order of magnitude lower than Classifier I (Fig. 2). Note that all three classifiers were able to detect all CSs (Sens=100%) but with a much higher FAR for Classifier I: 16.7 compared to 1.26 for Classifier II (13 times higher) and 5.9 for Classifier III (3 times higher).

At their optimal thresholds, marked by black squares (Fig. 2), Classifier I detected 46 out of 55 CSs (Sens=83.64%), including 3 (50%) FTC and 43 (87.7%) FTCb seizures; Classifier II detected 51 out of 55 CSs (Sens=92.73%), including 3 (50%) FTC and 48 (97.9%) FTCb seizures, and Classifier III detected 52 out of 55 CSs (Sens=94.55%), including 3 (50%) FTC and 49 (100%) FTCb seizures (Table 2). Fig. 3A shows the positive detections per patient.

Sens values at the optimal thresholds were statistically different between Classifier II and Classifier I (ΔSens=9.09%, $CI_{\Delta Sens}$=[0.41%-19.31%]) and between Classifier III and Classifier I (ΔSens=10.91%, $CI_{\Delta Sens}$=[1.78%-21.74%]). All classifiers detected the seizures before the v-EEG offset (Fig. 3B) with comparable latencies (Table 2), i.e., median 31.2 s, range 14.9-116 s (Classifier I); median 29.3 s, range 13.8-153 s (Classifier II); median 29.3 s, range 14.8-151 s (Classifier III).

At each optimal threshold, 71 false alarms were generated by Classifier I (overall FAR=0.29), 51 by Classifier II (FAR=0.21) and 50 by Classifier III (FAR=0.20) over the 69 patients. FPR values at the optimal threshold (Table 2) were statistically different between Classifier II and Classifier I (ΔFPR=0.008%, $CI_{\Delta FPR}$=[0.001%-0.016%]) and between Classifier III and Classifier I (ΔFPR=0.009%, $CI_{\Delta FPR}$=[0.002%-0.017%]). Fig. 3C shows histograms of individual patients' FAR values for each classifier. Most patients had fewer than 1 false alarm every 4 days (FAR<0.25): 41 of 69 patients (60%) for Classifier I, rising to 49 patients (71%) for Classifier II and 47 patients (69%) for Classifier III. In the worst case, some patients had up to 2 false alarms/day. The overall FAR/SR was lower for Classifier II and III compared to Classifier I (Table 2) with FAR/SR < 1 for 20 out of 22 seizure patients (90%) for both Classifier II and Classifier III. Classifier I showed a FAR/SR < 1 for 14 seizure patients (64%) (Fig. 3D). For Classifier I, 4 of 71 false alarms were generated during rest, while Classifiers II and III triggered no false alarms during rest.

Seizure characterization based on wrist ACM and EDA signals

The automated estimation of seizure intervals reflected expert-labeled seizure duration. Correlation between the estimated duration of motor convulsions and the v-EEG-based duration was statistically significant (r=0.73, p<0.0001, Fig. 4A) for the detected seizures by Classifier III. Furthermore, 40 out of 55 CSs (73%) exhibited a significant EDR upon seizure onset, including 3 of 6 FTC (50%) and 37 of 49 FTCb (76%). According to the FDR

procedure, the autonomic dysregulation following CSs with a significant EDR lasted 13 minutes, as shown in Fig. 4B. Features of EDA profiles in the post-ictal period are summarized in Table S2.

4. Discussion

The present work introduces two novel automated machine learning classifiers for detecting convulsive epileptic seizures, by combining motor activity using ACM sensors with sympathetic activity measured as EDA. Both signals exhibit marked changes upon the onset of most CSs[22,23,34]. Both classifiers can operate within a nonstigmatizing wrist-worn device, the location preferred by most patients[35], providing wearability for EMU and home use.

4.1 Seizure detection

There have been previous attempts to build an automated motor seizure detector that combines EDA and ACM. The main limitation was a relatively low sample number of seizures (16 FTCb seizures from 7 pediatric patients[22] and 21 predominantly motor seizures from 4 patients[36]) recorded with one type of device at a single clinical site[22,36]. As a significant advance, we developed two new automated detectors and tested them on a much larger and more diverse pool of EMU data (55 CSs from 22 patients, adult and pediatric) collected at six clinical sites, with different clinical teams, and recorded with three different devices. This diversity, while requiring greater effort, boosts generalizability and overcomes the limitations of most studies[11].

The results presented in the current study contribute to advancing the state of the art. The new Classifier II and Classifier III significantly and consistently outperformed Classifier I, trained using Poh's study feature set[22]. While Classifier I missed 9 seizures at its optimal threshold, Classifiers II and III missed 4 and 3 seizures, respectively. When tuning the three classifiers to the decision threshold at the same Sens level, Classifiers II and III yielded FAR's one order

of magnitude lower than Classifier I. As the training phases were performed on the same dataset, this direct comparison between Classifier I and the two novel classifiers is meaningful because none of the classifiers had an easier task than the others. Moreover, at their optimal decision thresholds, the two new classifiers were able to detect all nocturnal seizures, while Classifier I missed three (seizures 10, 24 and 25, Figs. S1 and S2). Also, the two new classifiers did not trigger any false alarms during quiescent periods. Many seizures and most SUDEPs are sleep-related[37]; thus, accurate performance at night is vital.

All three classifiers provided detection latencies acceptable for most patients[38]. However, all failed to detect three FTC seizures from our youngest pediatric patient (age 4). Visual inspection of wristband signals reveals mild motor activity and no significant ictal EDR (seizures 32, 33 and 34, Figs. S1-S2). To detect these seizures, it would be necessary to lower the decision threshold. In return, the FAR values of Classifier II and III would increase to values which may be too disruptive for some patients and families (i.e. FAR>1)[38], even if they would be considerably lower (13 and 3 times, respectively) than the Classifier I's FAR.

The two new methods in this study perform better than other published wrist-worn CSs detectors using ACM alone. At its best decision threshold, Classifier III yields Sens=94.55% and FAR=0.19. A Sens≈95% is acceptable for most patients[38] and is higher than the sensitivities reported for other devices, including SmartWatch (Sens=31%, 16/51 GTC/FTCb seizures[18]; Sens=88%, 7/8 GTC/FTCb seizures[13]; and Sens=92%, 12/13 GTC/FTCb seizures[39]), Epi-Care Free (Sens=56%, 9/16 FTCb seizures[19], Sens=90%, 35/39 FTCb seizures[15]), and Epilert (Sens=90%, 20/22 tonic/tonic-clonic seizures[14]). The Epi-Care Free and Epilert showed respectively FAR=0.2[15] and FAR=0.11[14], both comparable to Classifier III's FAR=0.20. On the other hand, our two new methods achieved better sensitivity with similar FAR, on a larger dataset, over more clinical sites and using three different devices. One study using Smartwatch reported more than 204 false alarms[13] and another 81 false

alarms[39]; however, their total recording hours were not reported, making their FAR indeterminate.

Another variable to evaluate the impact of a wearable seizure detector for patients is the ratio of the number of false alarms vs. the number of seizures per patient[12]. Our novel classifiers achieved a ratio lower than 1 for *most* of the patients and overall, the ratio of the total number of false alarms to the total number of seizures was ~1, which is acceptable for most patients and caregivers[38].

It was not possible to perform a direct comparison on our dataset with respect to the ACM+EDA classifier proposed by another group[36]. These authors achieved a sensitivity of 90.5% on 21 motor seizures and they reported that "while in the aforementioned study (n.b., Poh's study[22]) only one false alarm per day was encountered, we encountered a high number of false alarms"[36]. Comparing our two new classifiers at Sens=90%, Classifier II and Classifier III reached a FAR of approximately 0.12 and 0.16 respectively, which is a substantial improvement.

4.2 Seizure characterization

EDA and ACM data offer the opportunity to objectively characterize seizure physiology, beyond capabilities provided by systems based on motion alone (ACM and/or EMG). ACM analysis permitted reliably estimating seizure durations, except in one case in which convulsions were preceded by a long non-motor lead in (seizure 52 on Fig. S1, an outlier on Fig. 3A). Moreover, our data confirmed previous findings showing considerable autonomic activation in the early post-ictal phase reflected by a significant EDR, comparable to values reported in Poh's study[23]. However, Poh's study reported that all FTCb seizures (12/12) exhibited an EDR significantly higher than baseline for more than 50 minutes, while we observed such a response for 73% of CSs (50% of FTC and 76% of FTCb) lasting 13 minutes

on average. This discrepancy could be explained by the fact that we analyzed a larger, more heterogeneous population, whereas Poh's population focused on pharmacologically refractory pediatric patients undergoing a workup toward epilepsy surgery[23]. A possible explanation is that more severe cases of epilepsy (e.g., with earlier age of onset, refractory to multiple anti-epileptic drugs, and needing presurgical evaluation) are associated with higher and longer EDRs. The measure of the autonomic impact of each seizure has previously been found to correlate with the duration of PGES after FTCb and GTC seizures[23,34], which has been proposed as a biomarker for SUDEP[5]. While no device has been shown to reduce the risk of SUDEP, which would require very large studies, incorporating EDA analysis in a home seizure detector may help to identify, characterize and alert caregivers to potentially dangerous seizures.

4.3 Limits and future work

The main limitation of our study and of all studies in this space is that patients were not in their home settings. While patients could get out of bed, shake dice, dance, play gesture-controlled video games, brush their teeth, etc., in everyday life patients may be more likely to engage in sports and physical labor that may lead to higher FARs. Real-time performance assessment outside of EMUs is essential to ascertain that the system will perform well for most patients[12]. To this aim, the main challenge will be to collect accurate ground truth data in real-life, which would likely require a multimodal system[32]. Even if ambulatory-EEG devices are reliable[40], they are uncomfortable, encumbering, and less preferred to wristbands by patients[35]. Self-reports are not an accurate standalone alternative[8,9].

New technologies offer the opportunity for applying machine learning as a tool within precision medicine, tuning the classifier to provide optimal tailored performance for each patient, taking into account the patient's unique seizure features and cost of false alarms

compared to true detections, which can depend upon seizure frequency[38]. Thus, future systems may be personalized to provide the best performance for each patient based on longitudinal real-life data collection.

Another important limitation of this study is that analyses were done retrospectively, differently from above-mentioned studies with real-time analysis[13,15,18,19,39]. Based on the robust cross-validation approach on a large number of CSs and on our simulated real-time processing, our expectation is that performance of an algorithm embedded in a real-time system will not change significantly (as verified by preliminary tests underway). Future work will train a classifier using feature set III on all the data, and apply it on a separate test set in a prospective validation study, with real-time analysis and seizure detection.

Future work should also involve evaluation of the algorithm for other types of motor seizures, e.g. hypermotor or clonic. The EDA might also help detect and characterize seizures with subtle or no motor movement. A recent review suggests that the EDA signal is one of the most promising alternatives for a widespread variety of epileptic episodes[24]. A study with an Empatica E3 wristband reported that 97% of 34 predominantly non-motor seizures could be detected with a hierarchical classifier based on EDA[36]. However, using mainly EDA dramatically decreased specificity; thus, information from other physiological signals is necessary to detect non-convulsive seizures, such as heart rate and arterial oxygenation[41]. Finally, a seizure detector incorporating EDA could be suitable for other important applications, such as identifying triggering factors reflected in autonomic activations (e.g. stress or deep sleep) or using EDA biofeedback for training patients to prevent epileptic seizures[42].

Key points

- Two multimodal automated convulsive seizure detectors were developed using accelerometry and electrodermal activity data, recorded with wrist-worn devices.
- A more diverse pool of data than prior clinical studies (55 seizures, 22 adult and pediatric patients, 6 sites, 3 devices) was used to test the algorithms.
- Direct comparison with the best state-of-the-art system using accelerometry and electrodermal activity showed significantly higher sensitivity (≈95 %).
- Most patients had fewer than 1 false alarm every 4 days and 90% of patients had a rate of false alarms lower than their seizure rate. No false alarms occurred during resting periods.
- In addition to seizure detection, the algorithm allowed reliable annotation of motor convulsion lengths and revealed post-ictal autonomic dysfunction in 73% of cases.

Acknowledgements

We are grateful to Francesca Coughlin at Boston Childrens Hospital, Sarah Barnard at NYU Langone Medical Center, Kevin Taylor at Emory and Children's Healthcare of Atlanta Hospital (CHOA), Anita Curran at Lifespan and Rhode Island Hospital, and Katrina Sambusida at Ospedale Niguarda Ca' Granda, for their generous support in helping collect and label v-EEG and wristband data. Some of this work was supported by grants from: Epilepsy Foundation, Norman Prince Neurosciences Institute, Brown Institute for Brain Sciences, Epilepsy Research Foundation, American Epilepsy Society, Patient-Centered Outcomes Research Institute (PCORI), Pediatric Epilepsy Research Foundation, Citizens United for Research in Epilepsy (CURE) Foundation, HHV-6 Foundation, Lundbeck, Eisai Ltd, Upsher-Smith Inc., Acorda Therapeutics Inc., and Pfizer Inc.

Disclosure of Conflicts of Interest

R.W.P., S.T., D.B., F.O., G.R., M.M. and C.C. are employees of Empatica Inc., which manufactured two of the devices used in this work and developed the two new algorithms

tested in this work. T.L. is part of pending patent applications to detect and predict seizures and to diagnose epilepsy with devices different from the ones used in this work. T.L., C.R., W.C.L. and A.S.B. have received sensors from Empatica Inc. and Affectiva Inc. to perform the reported research. The remaining authors have no conflict of interest. We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

References

1. Institute of Medicine (US) Committee on the Public Health Dimensions of the Epilepsies. Epilepsy Across the Spectrum: Promoting Health and Understanding. Washington (DC): National Academies Press (US); 2012.

2. Cheshire W, Tatum W. Sudden Unexpected Death in Epilepsy. Hospital Physician Board Review Manual: Epilepsy. Turner White Communications, 2014: 2(part 5).

3. Devinsky O, Nashef L. SUDEP: The death of nihilism. Neurology 2015;85:1534–1535.

4. Ryvlin P, Nashef L, Lhatoo SD, et al. Incidence and mechanisms of cardiorespiratory arrests in epilepsy monitoring units (MORTEMUS): a retrospective study. Lancet Neurol. 2013;12:966–77.

5. Lhatoo SD, Nei M, Raghavan M, et al. Nonseizure SUDEP: Sudden unexpected death in epilepsy without preceding epileptic seizures. Epilepsia 2016;57:1161–8.

6. Dlouhy BJ, Gehlbach BK, Kreple CJ, et al. Breathing Inhibited When Seizures Spread to the Amygdala and upon Amygdala Stimulation. J Neurosci 2015;35:10281–9.

7. Seyal M, Bateman LM, Li C-S. Impact of peri-ictal interventions on respiratory dysfunction, post-ictal EEG suppression, and post-ictal immobility. Epilepsia 2013;54:377–82.

8. Fisher RS, Blum DE, DiVentura B, et al. Seizure diaries for clinical research and practice: limitations and future prospects. Epilepsy Behav 2012;24:304–10.

9. Hoppe C, Poepel A, Elger CE. Epilepsy: Accuracy of patient seizure counts. Arch Neurol 2007;64:1595–9.

10. Van de Vel A, Cuppens K, Bonroy B, et al. Non-EEG seizure-detection systems and potential SUDEP prevention: state of the art. Seizure 2013;22:345–55.

11. Jory C, Shankar R, Coker D, et al. Safe and sound? A systematic literature review of seizure detection methods for personal use. Seizure 2016;36:4–15.

12. Van Andel J, Thijs RD, de Weerd A, et al. Non-EEG based ambulatory seizure detection designed for home use: What is available and how will it influence epilepsy care? Epilepsy Behav 2016;57:82–9.

13. Lockman J, Fisher RS, Olson DM. Detection of seizure-like movements using a wrist accelerometer. Epilepsy Behav 2011;20:638–41.

14. Kramer U, Kipervasser S, Shlitner A, et al. A novel portable seizure detection alarm system: preliminary results. J Clin Neurophysiol Off Publ Am Electroencephalogr Soc

2011;28:36–8.

15. Beniczky S, Polster T, Kjaer TW, et al. Detection of generalized tonic–clonic seizures by a wireless wrist accelerometer: A prospective, multicenter study. Epilepsia 2013;54:e58–61.

16. Szabó CÁ, Morgan LC, Karkar KM, et al. Electromyography-based seizure detector: Preliminary results comparing a generalized tonic–clonic seizure detection algorithm to video-EEG recordings. Epilepsia 2015;56:1432–7.

17. Conradsen I, Beniczky S, Hoppe K, et al. Automated algorithm for generalized tonic-clonic epileptic seizure onset detection based on sEMG zero-crossing rate. IEEE Trans Biomed Eng 2012;59:579–85.

18. Patterson AL, Mudigoudar B, Fulton S, et al. SmartWatch by SmartMonitor: Assessment of Seizure Detection Efficacy for Various Seizure Types in Children, a Large Prospective Single-Center Study. Pediatr Neurol 2015;53:309–11.

19. Van de Vel A, Verhaert K, Ceulemans B. Critical evaluation of four different seizure detection systems tested on one patient with focal and generalized tonic and clonic seizures. Epilepsy Behav 2014;37:91–4.

20. Milosevic M, Van de Vel A, Bonroy B, et al. Automated Detection of Tonic-Clonic Seizures using 3D Accelerometry and Surface Electromyography in Pediatric Patients. IEEE J Biomed Health Inform 2016; 20:1333-41.

21. Conradsen I, Beniczky S, Wolf P, et al. Automatic multi-modal intelligent seizure acquisition (MISA) system for detection of motor seizures from electromyographic data and motion data. Comput Methods Programs Biomed 2012;107:97-110.

22. Poh M-Z, Loddenkemper T, Reinsberger C, et al. Convulsive seizure detection using a wrist-worn electrodermal activity and accelerometry biosensor. Epilepsia 2012; 53:e93–7.

23. Poh M-Z, Loddenkemper T, Reinsberger C, et al. Autonomic changes with seizures correlate with post-ictal EEG suppression. Neurology 2012;78:1868–76.

24. Ulate-Campos A, Coughlin F, Gaínza-Lein M, et al. Automated seizure detection systems and their effectiveness for each type of seizure. Seizure 2016;40:88–101.

25. Boucsein W. Electrodermal Activity. Springer Science & Business Media, 2012.

26. Moseley BD. Seizure-Related Autonomic Changes in Children: J Clin Neurophysiol 2015;32:5–9.

27. Scheffer IE, Berkovic S, Capovilla G, et al. ILAE classification of the epilepsies: position paper of the ILAE commission for classification and terminology. Epilepsia 2017;58:512-521.

28. Biswas B. Study design and Analysis Issues for Diagnostic Monitoring Devices. JSM 2015 Proceedings 2015, 261–8. Abstract.

29. Weng CG, Poon J. A New Evaluation Measure for Imbalanced Datasets. Proceedings of the 7th Australasian Data Mining Conference 2008, 87:27–32. Abstract.

30. Altman DG, Machin D, Bryant TN, et al. Statistics with confidence. 2nd ed. Bristol: British Medical Journal; 2000.

31. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983;148:839–43.

32. Bidwell J, Khuwatsamrit T, Askew B, et al. Seizure reporting technologies for epilepsy treatment: A review of clinical information needs and supporting technologies. Seizure 2015;32:109–17.

33. Tracy DJ, Xu Z, Choi L, et al. Separating Bedtime Rest from Activity Using Waist or Wrist-Worn Accelerometers in Youth. PLoS ONE 2014;9:e92512.

34. Sarkis RA, Thome-Souza S, Poh M-Z, et al. Autonomic changes following generalized tonic clonic seizures: An analysis of adult and pediatric patients with epilepsy. Epilepsy Res 2015;115:113–8.

35. Hoppe C, Feldmann M, Blachut B, et al. Novel techniques for automated seizure registration: Patients' wants and needs. Epilepsy Behav 2015;52:1–7.

36. Heldberg BE, Kautz T, Leutheuser H, et al. Using wearable sensors for semiology-independent seizure detection - towards ambulatory monitoring of epilepsy. Conference Proceedings Annul Internation Conference IEEE Eng Med Biol 2015;2015:5593–6. Abstract.

37. Lamberts RJ, Thijs RD, Laffan A, et al. Sudden unexpected death in epilepsy: people with nocturnal seizures may be at highest risk. Epilepsia 2012; 53:253–7.

38. Van de Vel A, Smets K, Wouters K, et al. Automated non-EEG based seizure detection: Do users have a say? Epilepsy Behav 2016;62:121–8.

39. Velez M, Fisher RS, Bartlett V, et al. Tracking generalized tonic-clonic seizures with a wrist accelerometer linked to an online database. Seizure 2016;39:13–8.

40. Faulkner HJ, Arima H, Mohamed A. The utility of prolonged outpatient ambulatory EEG. Seizure 2012;21:491–5.

41. Cogan D, Birjandtalab J, Nourani M, et al. Multi-Biosignal Analysis for Epileptic Seizure Monitoring. Int J Neural Syst 2017;27:1650031.

42. Micoulaud-Franchi J-A, Kotwas I, Lanteaume L, et al. Skin conductance biofeedback training in adults with drug-resistant temporal lobe epilepsy and stress-triggered seizures: a proof-of-concept study. Epilepsy Behav 2014;41:244–50.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Figure S1. ACM magnitude signals of the 55 convulsive seizures (CSs) recorded, identified by seizure ID and patient (PT) ID. The signals are zoomed in a short neighborhood of the seizure (from 1 minute before the onset to 3 minutes after the end of the epileptic event). The pink line marks the seizure onsets. Asterisks (*) identify CSs that occurred during the night. Focal motor seizures are marked with "FTC" (all other events are focal motor to bilateral tonic-clonic seizures, i.e. FTCb).

Figure S2. EDA signals during the 55 individual convulsive seizures (CSs) recorded, identified by seizure ID and patient (PT) ID. EDA recordings are zoomed-in around the seizure onset (from 5 minutes before the onset to 100 minutes after the end of the epileptic event) and are expressed in normalized units. The pink line marks seizure onset. Asterisks (*) identify CSs that occurred during the night. Focal motor seizures are marked with "FTC" (all other events are focal motor to bilateral tonic-clonic seizures, i.e. FTCb)

Table S1. Cross-validation results.

Results related to the three feature sets under comparison in this work highlighting performances obtained at each tournament of the cross-validation analysis (black font). Each tournament corresponds to training/tuning a classifier on 2/3 of data and testing it on the left-out third. Average and cumulative performances along the three tournament reported in the main text are shown (blue font). Note that the paper reports the more conservative test, which accumulates the errors instead of averaging them.

Table S2. Summary of post-ictal EDA profiles.

Characteristics of post-ictal EDRs, reported as median (25^{th}, 75^{th} percentiles / min-max), for CSs exhibiting a statistically significant EDR (N=40).

Figures

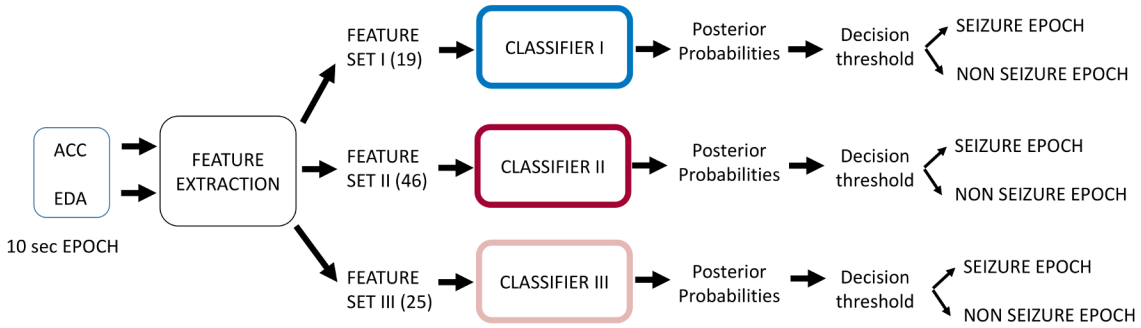

Figure 1. Overview of the workflow used for the development of convulsive seizure detectors tested in the present work. EDA and ACM signals were segmented in sliding epochs of 10 s (75% overlap). Three different feature sets were computed on each epoch: one made of 19 features, originally used in Poh et al 2012[22] (feature set for Classifier I), one of 46 features (feature set for Classifier II) and one of 25 features (feature set for Classifier III, a subset of the 46 features). Classifiers were constructed and validated using a cross-validation approach. For each epoch, a posterior probability estimate was provided as output by the classifier. Each epoch was classified as a seizure or non-seizure epoch by applying a decision threshold to the posterior probability estimates.

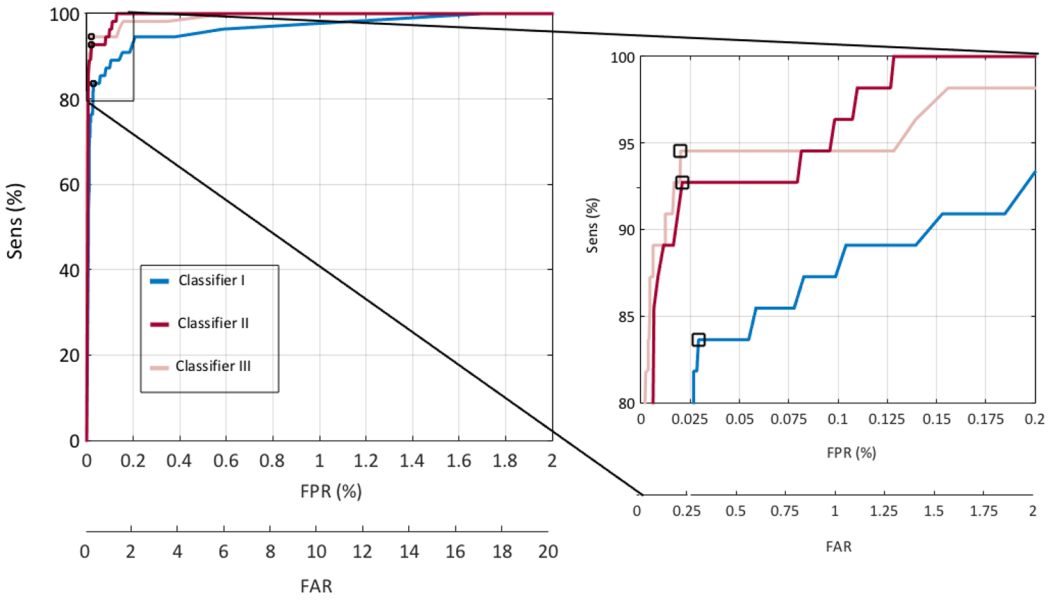

Figure 2. Receiver operating characteristic (ROC) curves of the three classifiers under comparison obtained with a double cross-validation approach. The three classifiers differ in the feature set they use: Classifier I uses 19 features originally proposed by Poh et al 2012[22], while Classifier II and Classifier III employ new sets of 46 and 25 features, respectively. The x-axis shows the false positive rate (FPR) and the false alarm rate (FAR, i.e. the number of false alarms in 24 hours), while the y-axis shows the sensitivity (Sens, i.e. the percentage of detected seizures). A Zoom at the top-left corner of the ROC is provided to better view the performances at higher Sens levels. In particular, at Sens≅85%, FAR=0.6 for Classifier I, FAR=0.06 for Classifier II and FAR=0.04 for Classifier III. At Sens≅90%, FAR=1.5 for Classifier I, FAR=0.16 for Classifier II and FAR=0.155 for Classifier III. Finally, at Sens≅95%, FAR=2 for Classifier I, FAR=0.8 for Classifier II and FAR=0.2 for Classifier III. The three classifiers are able to detect all the CSs (Sens=100%) at the cost of a much higher FAR for Classifier I: 16.7 compared to 1.26 for Classifier II and 5.9 for Classifier III. Black squares superimposed to each curve mark performance at the optimal decision threshold selected with a cost function maximizing Sens and minimizing FAR.

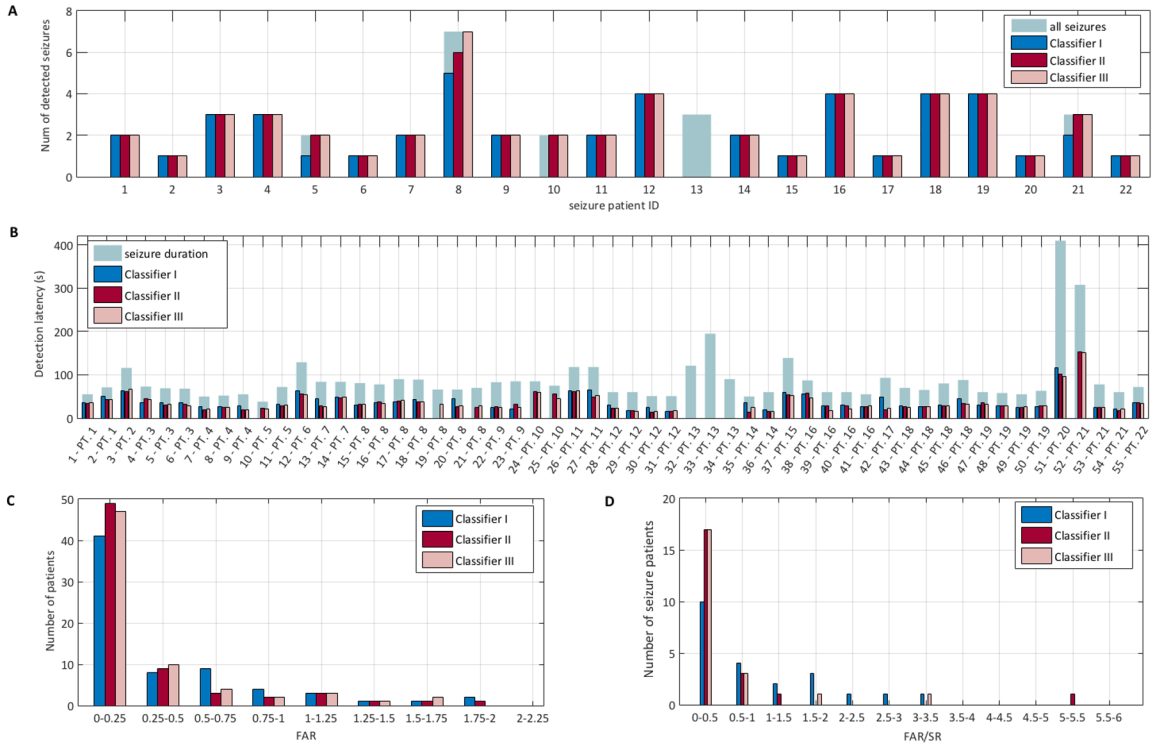

Figure 3. A. Number of detected convulsive seizures (CSs) per seizure patient (N=22) using the three different classifiers. **B.** Latencies of detection (seconds relative to the start of the seizure determined using v-EEG) for each seizure with the 3 classifiers. Each seizure is identified by seizure number (N=55) and patient (PT) ID. The absence of colored bars indicates undetected CSs. **C.** Histograms of false alarm rates (FAR, i.e. number of false alarms per 24 hours) per patient (N=69) using the three classifiers. **D**. Histograms of FAR/SR, i.e. number of false alarms divided by the number of seizures per seizure patient (N=22), using the three classifiers.

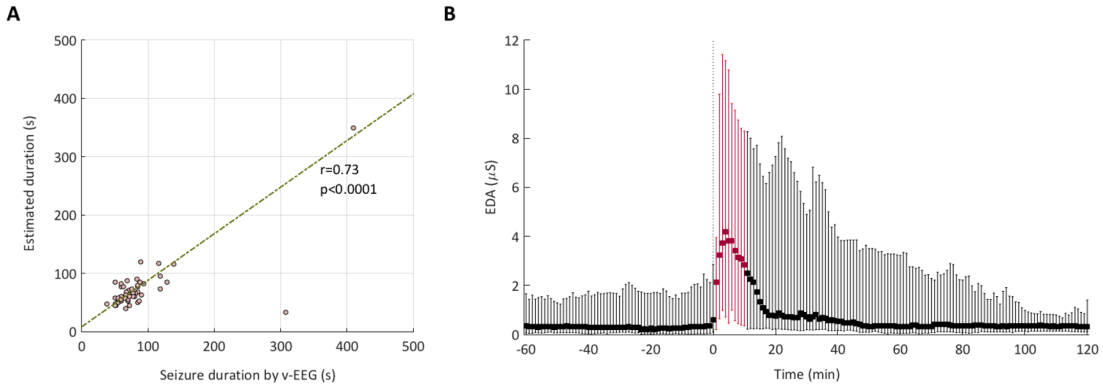

Figure 4. Seizure characterization. **A.** Correlation (Pearson's correlation coefficient, r) between seizure duration assessed by v-EEG labeling and estimated duration based on ictal ACM analysis, performed on CS detected by Classifier III at its optimal threshold (N=52). The green dotted line represents the linear regression line. **B.** High-resolution profiles of autonomic alterations computed every minute during a peri-ictal period of 3 hours (1 hour before the onset, 2 hours afterward), aligned to the EEG seizure onset. The square box associated with each epoch represents the median EDA level across CSs, while the bars span the inter-quartile range (N=55). Each 60-second post-ictal measurement epoch was sequentially compared with the baseline level taken as the average of the entire 60-minute pre-ictal period. Epochs in red indicate statistically significant epochs after accounting for multiple comparisons using the false discovery rate controlling procedure ($p<0.05$, two-sample Kolmogorov-Smirnov test).

Tables

Table 1: Summary of recorded convulsive seizures

Number of patients (number with seizures)	*69 (22)*
Total number of CS	55
Number of CS per patient (range)	1-7
Median seizure duration (range)	72 (38-410) seconds
Number of FTCb seizures (number of patients)	49 (20)
Number of FTC seizures (number of patients)	6 (2)
Number of seizures occurring during sleep	19 (35%)

Characteristics of convulsive seizures (CSs) recorded with wristband sensors. FTCb= Focal motor to bilateral tonic-clonic; FTC= focal motor tonic-clonic (unilateral).

Table 2: Seizure detector performance comparison

	Classifier I	Classifier II	Classifier III
AUC	0.86 CI=[0.80 0.93]	0.93 CI=[0.89 0.98]	0.94 CI=[0.89 0.98]
Sens	83.64 % CI=[71.75-91.14] %	92.73 % CI=[82.74-97.14] %	94.55 % CI=[85.15-98.13] %
FPR	0.029 % CI=[0.023 0.037] %	0.021 % CI=[0.016 0.028] %	0.02 % CI=[0.015 0.027] %
FAR (false alarms per day)	0.29	0.21	0.20
Detection latency (s)	31.2 range=[14.9-116] N=47	29.3 range=[13.8-153] N=51	29.3 range=[14.8-151] N=52
Precision	39 %	50 %	51 %
F-score	0.53	0.65	0.67
FAR/SR	1.3	0.93	0.91

Performance metrics for the three classifiers under comparison. All metrics apart from the area under the ROC curve (AUC) refer to performances at each classifier's optimal decision threshold. CI=confidence interval at 95% confidence. Range=Min/max range. N=number of detected seizures. FAR=false alarm rate. SR=convulsive seizure rate. Detection latency is reported as median and range values.