




Multicenter clinical assessment of improved wearable multimodal convulsive seizure detectors

*†¹ Francesco Onorati, *†¹ Giulia Regalia, *† Chiara Caborni, *† Matteo Migliorini, *† Daniel Bender, ‡ Ming-Zher Poh, § Cherise Frazier, ¶ Eliana Kovitch Thropp, # Elizabeth D. Mynatt, §¶# Jonathan Bidwell, ** Roberto Mai, †† W. Curt LaFrance Jr , ‡‡ Andrew S. Blum, §§ Daniel Friedman, ¶¶ Tobias Loddenkemper, ¶¶ Fatemeh Mohammadpour-Touserkani, ## Claus Reinsberger, *† Simone Tognetti, and *†‡ Rosalind W. Picard

Epilepsia, 58(11):1870–1879, 2017
doi: 10.1111/epi.13899



Francesco Onorati is a Principal Senior Data Scientist and Bioengineer at Empatica, Milan, Italy.



Giulia Regalia is a Senior Data Scientist and Bioengineer at Empatica, Milan, Italy.

SUMMARY

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

Methods: Hand-annotated video-electroencephalographic seizure events were collected from 69 patients at six clinical sites. Three different wristbands were used to record electrodermal activity (EDA) and accelerometer (ACM) signals, obtaining 5,928 h of data, including 55 convulsive epileptic seizures (six focal tonic-clonic seizures and 49 focal to bilateral tonic-clonic seizures) from 22 patients. Recordings were analyzed offline to train and test two new machine learning classifiers and a published classifier based on EDA and ACM. 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 an FAR of 0.2 events/day. No nocturnal seizures were missed. Most patients had <1 false alarm every 4 days, with an FAR below their seizure frequency. When increasing the sensitivity to 100% (no missed seizures), the FAR is up to 13 times lower than with 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 postictal 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 behavior and autonomic dysfunction, aimed at enriching seizure characterization, with potential utility for SUDEP warning.

KEY WORDS: Epilepsy, Convulsive seizures, Electrodermal activity, Machine learning.

Epilepsy is among the most common neurological disorders, with an estimated 65 million patients worldwide.¹ Although 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

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, six sites, three 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 ($\approx 95\%$)
- Most patients had <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 postictal autonomic dysfunction in 73% of cases

(GTC) seizure a year, and when a patient is unattended after a seizure.^{2,3} Although SUDEP's general cause remains unknown, SUDEP can occur after prolonged postictal generalized electroencephalographic suppression (PGES), and is associated with autonomic dysfunction such as terminal apnea preceding terminal asystole.^{4–6} 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, although 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.^{10–12} Existing automated seizure detectors¹¹ measure motion to detect seizures with a motor manifestation. Algorithms based on wrist acceleration (accelerometer [ACM])^{13–15} 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 (regarding number of seizures and recording hours), which prevents robust estimates of sensitivity and false alarm rates (FARs). Rarely are objective characterizations about seizure events, beyond seizure counts, provided to the patient or clinician.¹¹

Although 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 seizures, that is, focal motor to bilateral tonic-clonic (FTCb) seizures, 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, hypermotor²⁴—and of the measurement of the sympathetic nervous system activity using EDA,²⁵ including periictal 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 convulsive seizures [CSs]). 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.

METHODS

Patients

Sixty-nine patients diagnosed with epilepsy (24 children, age = 4–18 years, median = 14 years, nine females; 45 adults, age = 19–60 years, median = 37 years, 28 females) were admitted for v-EEG monitoring at six clinical sites: Children's Hospital Boston (14 patients), New York University Langone Medical Center (18 patients), Rhode Island Hospital (five patients), Emory University Hospital

Accepted August 23, 2017; Early View publication October 4, 2017.

*Empatica, Milan, Italy; †Empatica, Cambridge, Massachusetts, U.S.A.; ‡MIT Media Lab, Massachusetts Institute of Technology, Cambridge, Massachusetts, U.S.A.; §Emory University Hospital Midtown, Atlanta, Georgia, U.S.A.; ¶Children's Healthcare of Atlanta, Atlanta, Georgia, U.S.A.; #Georgia Institute of Technology, Atlanta, Georgia, U.S.A.; **Claudio Munari Epilepsy Surgery Center, Niguarda Hospital, Milan, Italy; ††Division of Neuropsychiatry and Behavioral Neurology, Rhode Island Hospital, Brown University, Providence, Rhode Island, U.S.A.; ‡‡Department of Neurology, Rhode Island Hospital, Brown University, Providence, Rhode Island, U.S.A.; §§Department of Neurology, New York University Langone Medical Center, New York, New York, U.S.A.; ¶¶Department of Neurology, Boston Children's Hospital, Boston, Massachusetts, U.S.A.; and ##Department of Neurology, Brigham and Women's Hospital, Boston, Massachusetts, U.S.A.

Address correspondence to Giulia Regalia, via Stendhal 36, 20144 Milan, Italy. E-mail: gr@empatica.com

¹Equally contributing authors.

and Children's Healthcare of Atlanta (15 patients), Egleston and Scottish Rite hospitals (12 patients), and Niguarda Hospital (five 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 seizure classification.²⁷

Wrist acceleration and EDA 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), and the iCalm (MIT Media Lab), all featuring comparable embedded three-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 nondominant arm. Five patients wore devices on both wrists. The resulting dataset consisted of 247 days (5,928 h; per patient median = 74.3 h, range = 3.5–386.8 h) 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 ACM and EDA measurements 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 non-linear analyses was constructed. The features were computed on 10-s epochs of three-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 study²² (19 features; 16 ACM and three EDA features), (II) a larger set (46 features, 40 ACM and six EDA features), and (III) a reduced set (25 of the 46 features, 22 ACM and three 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 to classify each epoch as seizure or nonseizure. All signals were analyzed offline using MATLAB (MathWorks). To simulate online detection, all "future-time" data points 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 nonoverlapping parts, each part containing epochs from one-third of the patients experiencing seizures and one-third 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, that is, holding out a different third part in each round. Thus, performances could be evaluated on the whole dataset (5,928 h), using no data for both training/tuning and testing at the same time.

For performance evaluation, we considered nonoverlapping segments labeled as seizure, defined as intervals between the clinical onset and the clinical offset according to v-EEG labeling, and nonseizure segments, defined as intervals not including seizure events. To facilitate comparisons and performance computation, nonseizure segments were split into subsegments with a duration equal to the mean duration of seizure segments, to deal with nonseizure events of approximately the same length as seizure events. This procedure allowed for better estimates of true

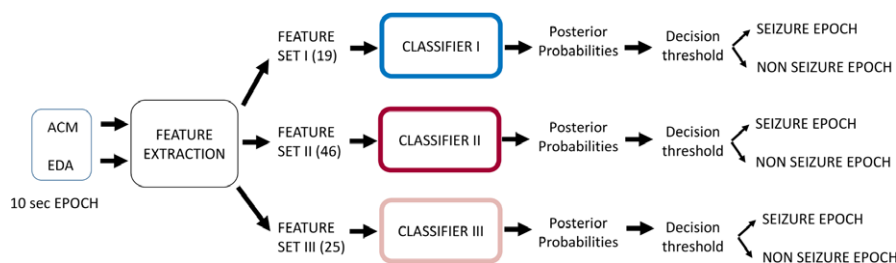


Figure 1.

Overview of the workflow used for the development of convulsive seizure detectors tested in the present work. Electrodermal activity (EDA) and accelerometer (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.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 nonseizure epoch by applying a decision threshold to the posterior probability estimates.

Epilepsia © ILAE

negatives, a complicated task for systems trained to detect only the event of interest.²⁸ The numbers 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 h (247 days). Table S1 details the results from each held-out third of the data (i.e., each round), and the cumulative performances are reported in Results. Note that our reported results are more conservative than if the performances of the three cross-validation rounds 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 nonseizure segments (equivalently, $100\% - \text{specificity}$). The 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 to provide the highest Sens with the lowest FAR, that is, the point closest to the upper left corner in the FAR/Sens ROC. To statistically compare the classifiers,

95% confidence intervals (CIs) for the Sens, the difference in Sens (ΔSens), the FPR, the difference in FPR (ΔFPR),³⁰ and the AUC³¹ 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 precision³² (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 by applying a rest detection algorithm to ACM measurements.³³

Seizure characterization: estimating motion duration and postictal EDA response

To estimate the seizure motor duration, an online algorithm was implemented to designate a neighborhood where the standard deviation of the ACM was >0.05 g-unit (unit of gravitational acceleration). Pearson correlation was then performed between these estimated durations, and the durations assessed by v-EEG labeling.

To quantify each seizure's periictal EDA response (EDR), periictal EDA recordings were segmented from 60 min before v-EEG seizure onset to 120 min afterward. A "significant EDR" was identified when EDA

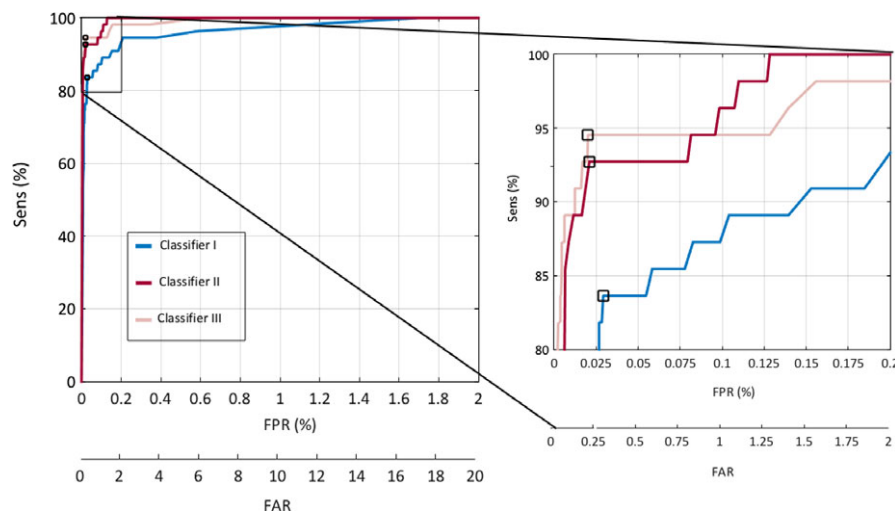


Figure 2.

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.,²² whereas 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 h), whereas 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 $\text{Sens} \cong 85\%$, $\text{FAR} = 0.6$ for Classifier I, $\text{FAR} = 0.06$ for Classifier II, and $\text{FAR} = 0.04$ for Classifier III. At $\text{Sens} \cong 90\%$, $\text{FAR} = 1.5$ for Classifier I, $\text{FAR} = 0.16$ for Classifier II, and $\text{FAR} = 0.155$ for Classifier III. Finally, at $\text{Sens} \cong 95\%$, $\text{FAR} = 2$ for Classifier I, $\text{FAR} = 0.8$ for Classifier II, and $\text{FAR} = 0.2$ for Classifier III. The three classifiers are able to detect all the convulsive seizures ($\text{Sens} = 100\%$) 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. Squares superimposed on each curve mark performance at the optimal decision threshold selected with a cost function maximizing Sens and minimizing FAR.

Epilepsia © ILAE

increased more than twice the standard deviation of the preictal 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 ν -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 AUC 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 preictal versus postictal 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

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 six FTC and 49 FTCb seizures. None of the captured seizures was nonepileptic. Thirty-two CSs (12 patients) were recorded with Empatica E4, nine CSs (four patients) with Empatica E3, and 14 CSs (six patients) with the MIT iCalm. A more detailed summary of the recordings is given in Table 1. One hundred thirty-five seizures other than FTC and FTCb were recorded and are not considered in this work. Individual ACM magnitude and EDA signals during the periictal period are shown in Figs. S1 and S2, respectively. Because 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 (not seizure related).

Table 1. Summary of recorded CSs

| | |
|---|-------------|
| Number of patients (number with seizures) | 69 (22) |
| Total number of CSs | 55 |
| Number of CSs per patient, range | 1–7 |
| Median seizure duration, s (range) | 72 (38–410) |
| 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 CSs recorded with wristband sensors. CSs, convulsive seizures; FTC, focal motor tonic–clonic (unilateral); FTCb, focal motor to bilateral tonic–clonic. | |

Performance comparison of CS detectors

A tradeoff 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 tradeoff 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 (Δ AUC = 0.0691, $p = 0.015$) and between the AUC values of Classifier I and Classifier III (Δ AUC = 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 an 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 (three times higher).

At their optimal thresholds, marked by squares (Fig. 2), Classifier I detected 46 of 55 CSs (Sens = 83.64%), including three (50%) FTC and 43 (87.7%) FTCb seizures; Classifier II detected 51 of 55 CSs (Sens = 92.73%), including three (50%) FTC and 48 (97.9%) FTCb seizures; and Classifier III detected 52 of 55 CSs (Sens = 94.55%), including three (50%) FTC and 49 (100%) FTCb seizures (Table 2). Figure 3A shows the positive detections per patient. Sens values at the optimal thresholds were statistically different between Classifier II and Classifier I (Δ Sens = 9.09%, CI_{Δ Sens} = 0.41–19.31%) and between Classifier III and Classifier I (Δ Sens = 10.91%, CI_{Δ Sens} = 1.78–21.74%). All classifiers detected the seizures before the ν -EEG offset (Fig. 3B) with comparable latencies (Table 2), that is, median = 31.2 s, range = 14.9–116 s (Classifier I); median = 29.3 s, range = 13.8–153 s (Classifier II); and 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 (Δ FPR = 0.008%, CI_{Δ FPR} = 0.001–0.016%) and between Classifier III and Classifier I (Δ FPR = 0.009%, CI_{Δ FPR} = 0.002–0.017%). Figure 3C shows histograms of individual patients’ FAR values for each classifier. Most patients had fewer than one 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

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 | 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 |

Detection latency is reported as median and minimum/maximum range values. Performance metrics are shown for the three classifiers under comparison. All metrics apart from the AUC refer to performances at each classifier's optimal decision threshold.

AUC, area under the receiver operating characteristic curve; CI, 95% confidence interval; FAR, false alarm rate (false alarms per day); FPR, false positive rate; n, number of detected seizures; Sens, sensitivity; SR, convulsive seizure rate.

47 patients (69%) for Classifier III. In the worst case, some patients had up to two false alarms per day. The overall FAR/SR was lower for Classifier II and III compared to Classifier I (Table 2) with FAR/SR < 1 for 20 of 22 seizure patients (90%) for both Classifier II and Classifier III. Classifier I showed an FAR/SR < 1 for 14 seizure patients (64%; Fig. 3D). For Classifier I, four of 71 false alarms were generated during rest, whereas 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 of 55 CSs (73%) exhibited a significant EDR upon seizure onset, including three of six FTC (50%) and 37 of 49 FTCb (76%) seizures. According to the FDR procedure, the autonomic dysregulation following CSs with a significant EDR lasted 13 min, as shown in Fig. 4B. Features of EDA profiles in the postictal period are summarized in Table S2.

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.

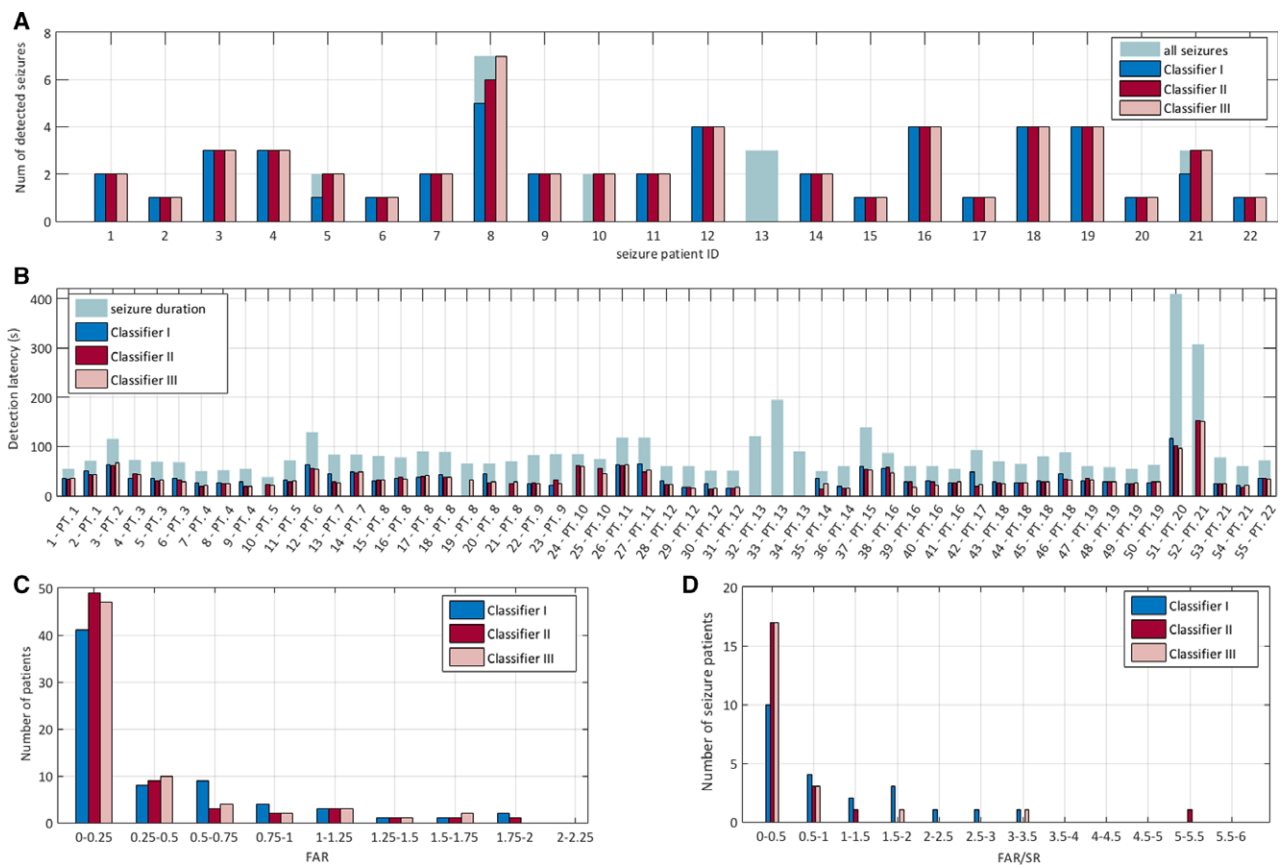
Seizure detection

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 seven pediatric patients²² and 21 predominantly motor seizures from four patients³⁶) 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 study²² feature set. Whereas Classifier I missed nine seizures at its optimal threshold, Classifiers II and III missed four and three seizures, respectively. When tuning the three classifiers to the decision threshold at the same Sens level, Classifiers II and III yielded FARs 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, whereas 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 years). Visual inspection of wristband signals reveals mild motor activity and no significant ictal EDR (seizures 32, 33, and 34; Figs. S1 and 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 that may be too disruptive for some patients and families

**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 video-electroencephalography) for each seizure with the three 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 h) per patient ($n = 69$) using the three classifiers. (D) Histograms of FAR/seizure rate (SR), that is, number of false alarms divided by the number of seizures per seizure patient ($n = 22$), using the three classifiers.

Epilepsia © ILAE

(i.e., $FAR > 1$),³⁸ even if they would be considerably lower (13 and three times, respectively) than Classifier I's FAR.

The two new methods in this study perform better than other published wrist-worn CS detectors using ACM alone. At its best decision threshold, Classifier III yields Sens = 94.55% and FAR = 0.19. A Sens \approx 95% is acceptable for most patients³⁸ and is higher than the sensitivities reported for other devices, including SmartWatch (Sens = 31%, 16/51 GTC/FTCb seizures¹⁸; Sens = 88%, seven/eight GTC/FTCb seizures¹³; and Sens = 92%, 12/13 GTC/FTCb seizures³⁹), Epi-Care Free (Sens = 56%, nine/16 FTcb seizures,¹⁹ Sens = 90%, 35/39 FTcb seizures¹⁵), and Epilert (Sens = 90%, 20/22 tonic/tonic-clonic seizures¹⁴). Epi-Care Free and Epilert showed, respectively, FAR = 0.2¹⁵ and FAR = 0.11,¹⁴ both comparable to Classifier III's FAR = 0.20. Conversely, 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 >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 versus the number of seizures per patient.¹² Our novel classifiers achieved a ratio < 1 for most of the patients, and the overall 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 study²²) 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 an FAR of approximately

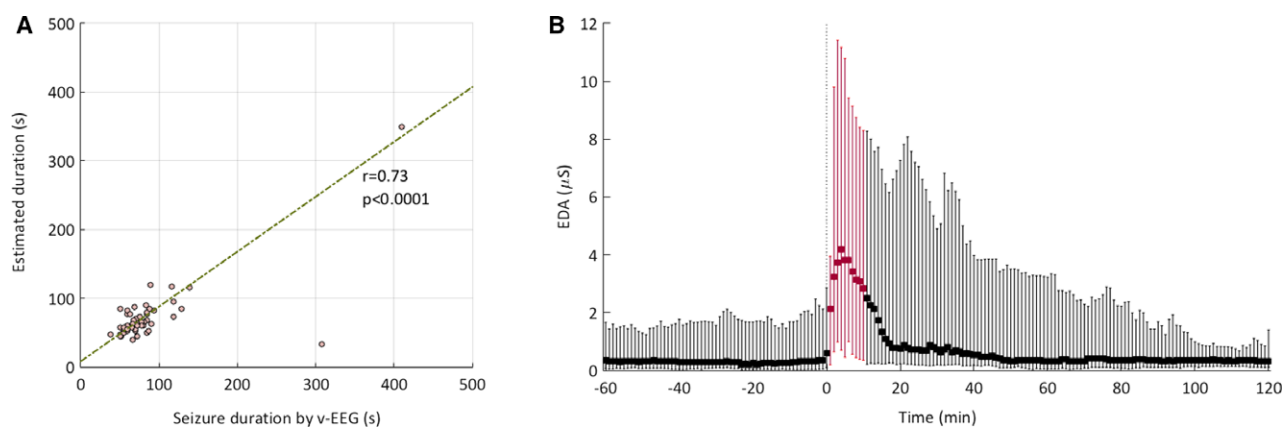


Figure 4.

Seizure characterization. **(A)** Correlation (Pearson correlation coefficient, r) between seizure duration assessed by video-electroencephalography (v-EEG) labeling and estimated duration based on ictal accelerometer analysis, performed on convulsive seizures (CSs) 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 periictal period of 3 h (1 h before the onset, 2 h afterward), aligned to the EEG seizure onset. The square associated with each epoch represents the median electrodermal activity level across CSs; the bars span the interquartile range ($n = 55$). Each 60-s postictal measurement epoch was sequentially compared with the baseline level taken as the average of the entire 60-min preictal 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).

Epilepsia © ILAE

0.12 and 0.16, respectively, which is a substantial improvement.

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 nonmotor lead-in (seizure 52 in Fig. S1, an outlier in Fig. 3A). Moreover, our data confirmed previous findings showing considerable autonomic activation in the early postictal 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 of 12) exhibited an EDR significantly higher than baseline for >50 min, whereas we observed such a response for 73% of CSs (50% of FTC and 76% of FTCb) lasting 13 min on average. This discrepancy could be explained our analyzing a larger, more heterogeneous population, whereas Poh's²³ population focused on pharmacologically refractory pediatric patients undergoing a workup toward epilepsy surgery. A possible explanation is that more severe cases of epilepsy (e.g., with earlier age of onset, refractory to multiple antiepileptic 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 Although 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.

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. Although patients could get out of bed, shake dice, dance, play gesture-controlled video games, brush their teeth, and so forth, 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 abovementioned 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, for example, 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 nonmotor 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 nonconvulsive 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

ACKNOWLEDGMENTS

The authors would like to thank Francesca Coughlin at Boston Childrens Hospital, Sarah Barnard at New York University Langone Medical Center, Kevin Taylor at Emory and Children's Healthcare of Atlanta Hospital, Anita Curran at Lifespan and Rhode Island Hospital, and Katrina Sambusida at Niguarda Hospital, for their generous support in helping collect and label v-EEG and wristband data. Some of this work was supported by grants from the Epilepsy Foundation, Norman Prince Neurosciences Institute, Brown Institute for Brain Sciences, Epilepsy Research Foundation, American Epilepsy Society, Patient-Centered Outcomes Research Institute, Pediatric Epilepsy Research Foundation, Citizens United for Research in Epilepsy Foundation, HHV-6 Foundation, Lundbeck, Eisai, Upsher-Smith, Acorda Therapeutics, and Pfizer.

CONFLICT OF INTEREST

Rosalind W. Picard, Simone Tognetti, Daniel Bender, Francesco Onorati, Giulia Regalia, Matteo Migliorini, and Chiara Caborni are employees of Empatica, which manufactured two of the devices used in this work and developed the two new algorithms tested in this work. Tobias Loddenkemper 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. Tobias Loddenkemper, Claus Reinsberger, W. Curt LaFrance Jr, and Andrew S. Blum have received sensors from Empatica and Affectiva to perform the reported research. The remaining authors have no conflict of interest. Authors confirm that they have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

REFERENCES

- Institute of Medicine (U.S.) Committee on the Public Health Dimensions of the Epilepsies. *Epilepsy across the spectrum: promoting health and understanding*. Washington, DC: National Academies Press; 2012.
- Cheshire W, Tatum W. *Sudden unexpected death in epilepsy. Hospital physician board review manual: epilepsy*. Vol 2 (part 5) Wayne, PA: Turner White Communications; 2014.
- Devinsky O, Nashef L. SUDEP: the death of nihilism. *Neurology* 2015;85:1534–1535.
- 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;12:966–977.
- Lhatoo SD, Nei M, Raghavan M, et al. Nonseizure SUDEP: sudden unexpected death in epilepsy without preceding epileptic seizures. *Epilepsia* 2016;57:1161–1168.
- Dlouhy BJ, Gehlbach BK, Kreple CJ, et al. Breathing inhibited when seizures spread to the amygdala and upon amygdala stimulation. *J Neurosci* 2015;35:10281–10289.
- 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;54:377–382.
- Fisher RS, Blum DE, DiVentura B, et al. Seizure diaries for clinical research and practice: limitations and future prospects. *Epilepsy Behav* 2012;24:304–310.
- Hoppe C, Poepel A, Elger CE. Epilepsy: accuracy of patient seizure counts. *Arch Neurol* 2007;64:1595–1599.
- 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;22:345–355.
- 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.
- 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;57:82–89.
- Lockman J, Fisher RS, Olson DM. Detection of seizure-like movements using a wrist accelerometer. *Epilepsy Behav* 2011;20:638–641.
- Kramer U, Kipervasser S, Shlitner A, et al. A novel portable seizure detection alarm system: preliminary results. *J Clin Neurophysiol* 2011;28:36–38.
- Beniczky S, Polster T, Kjaer TW, et al. Detection of generalized tonic-clonic seizures by a wireless wrist accelerometer: a prospective, multi-center study. *Epilepsia* 2013;54:e58–e61.
- 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;56:1432–1437.
- 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;59:579–585.
- 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;53:309–311.
- 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;37:91–94.
- 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;20:1333–1341.
- 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.
- 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–e97.
- Poh M-Z, Loddenkemper T, Reinsberger C, et al. Autonomic changes with seizures correlate with post-ictal EEG suppression. *Neurology* 2012;78:1868–1876.
- 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*. New York: 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. In *JSM Proceedings, Statistical Computing Section*. Alexandria, VA: American Statistical Association; 2015:261–268.
29. Weng CG, Poon J. A new evaluation measure for imbalanced datasets. In *Proceedings of the 7th Australasian Data Mining Conference*. Vol. 87. Darlinghurst, Australia: Australian Computer Society; 2008:27–32.
30. Altman DG, Machin D, Bryant TN, et al. *Statistics with confidence*. 2nd Ed. Bristol, UK: 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;148:839–843.
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;32:109–117.
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;115:113–118.
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. *Conf Proc IEEE Eng Med Biol Soc* 2015;2015:5593–5596.
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;53:253–257.
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;62:121–128.
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;39:13–18.
40. Faulkner HJ, Arima H, Mohamed A. The utility of prolonged outpatient ambulatory EEG. *Seizure* 2012;21:491–495.
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;41:244–250.

SUPPORTING INFORMATION

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

Figure S1. Accelerometer 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 min before the onset to 3 min after the end of the epileptic event). The pink line marks the seizure onsets. *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. Electrodermal activity (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 min before the onset to 100 min after the end of the epileptic event) and are expressed in normalized units. The pink line marks seizure onset. *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 relate to the three feature sets under comparison in this work; performances obtained at each round of the cross-validation analysis are highlighted (black font). Each round corresponds to training/tuning a classifier on two-thirds of the data and testing it on the left-out one-third. Average and cumulative performances along the three rounds reported in the main text are shown (blue font). Note that the article reports the more conservative test, which accumulates the errors instead of averaging them.

Table S2. Summary of postictal electrodermal activity profiles. Characteristics of postictal electrodermal activity responses (EDRs) are reported as median (25th, 75th percentiles/min–max) for convulsive seizures exhibiting a statistically significant EDR (n = 40).