Received: 13 December 2023 | Revised: 19 March 2024 | Accepted: 25 March 2024

DOI: 10.1111/epi.17974

Epilepsia®

RESEARCH ARTICLE

Reliable detection of generalized convulsive seizures using an off-the-shelf digital watch: A multisite phase 2 study

Yash Shashank Vakilna[1,2] | **Xiaojin Li**[1,2] | **Jaison S. Hampson**[1,2] | **Yan Huang**[1,2] | **John C. Mosher**[1,2] | **Yuri Dabaghian**[1,2] | **Xi Luo**[3] | **Blanca Talavera**[1,2] | **Sandipan Pati**[1,2] | **Masel Todd**[4] | **Ryan Hays**[5] | **Charles Akos Szabo**[6] | **Guo-Qiang Zhang**[1,2] | **Samden D. Lhatoo**[1,2]

[1]Department of Neurology, University of Texas Health Science Center at Houston, Houston, Texas, USA

[2]Texas Institute of Restorative Neurotechnologies, University of Texas Health Science Center at Houston, Houston, Texas, USA

[3]Department of Biostatistics and Data Science, University of Texas Health Science Center at Houston, Houston, Texas, USA

[4]Department of Neurology, University of Texas Medical Branch, Galveston, Texas, USA

[5]Department of Neurology, University of Texas Southwestern Medical Center, Dallas, Texas, USA

[6]Department of Neurology, University of Texas Health Science Center at San Antonio, San Antonio, Texas, USA

Correspondence
Yash Shashank Vakilna, Department of Neurology, Texas Institute of Restorative Neurotechnologies (TIRN), McGovern Medical School, UTHealth | University of Texas Health Science Center at Houston, 1133 John Freeman Blvd, Houston, TX 77030, USA.
Email: yashvakilna@gmail.com

Funding information
Foundation for the National Institutes of Health, Grant/Award Number: R01NS126690

Abstract

Objective: The aim of this study was to develop a machine learning algorithm using an off-the-shelf digital watch, the Samsung watch (SM-R800), and evaluate its effectiveness for the detection of generalized convulsive seizures (GCS) in persons with epilepsy.

Methods: This multisite epilepsy monitoring unit (EMU) phase 2 study included 36 adult patients. Each patient wore a Samsung watch that contained accelerometer, gyroscope, and photoplethysmographic sensors. Sixty-eight time and frequency domain features were extracted from the sensor data and were used to train a random forest algorithm. A testing framework was developed that would better reflect the EMU setting, consisting of (1) leave-one-patient-out cross-validation (LOPO CV) on GCS patients, (2) false alarm rate (FAR) testing on non-seizure patients, and (3) "fixed-and-frozen" prospective testing on a prospective patient cohort. Balanced accuracy, precision, sensitivity, and FAR were used to quantify the performance of the algorithm. Seizure onsets and offsets were determined by using video-electroencephalographic (EEG) monitoring. Feature importance was calculated as the mean decrease in Gini impurity during the LOPO CV testing.

Results: LOPO CV results showed balanced accuracy of .93 (95% confidence interval [CI] = .8–.98), precision of .68 (95% CI = .46–.85), sensitivity of .87 (95% CI = .62–.96), and FAR of .21/24 h (interquartile range [IQR] = 0–.90). Testing the algorithm on patients without seizure resulted in an FAR of .28/24 h (IQR = 0–.61). During the "fixed-and-frozen" prospective testing, two patients had three GCS, which were detected by the algorithm, while generating an FAR of .25/24 h (IQR = 0–.89). Feature importance showed that heart rate-based features outperformed accelerometer/gyroscope-based features.

Significance: Commercially available wearable digital watches that reliably detect GCS, with minimum false alarm rates, may overcome usage adoption and

other limitations of custom-built devices. Contingent on the outcomes of a prospective phase 3 study, such devices have the potential to provide non-EEG-based seizure surveillance and forecasting in the clinical setting.

KEYWORDS

machine learning, seizure detection, seizure forecasting, wearable devices

1 | INTRODUCTION

Generalized convulsive seizures (GCS) carry a significant risk of physical injuries and mortality, including sudden unexpected death in epilepsy. Persons with epilepsy (PWE) often require postseizure aid and on occasion hospitalization. In recent times, custom wearable devices, some US Food and Drug Administration (FDA)-approved, have been used for GCS detection. These, using a variety of biometric monitoring technologies, allow objective detection, quantification, and increasingly, forecasting of seizures. As a result, the International League Against Epilepsy along with the International Federation of Clinical Neurophysiology published guidelines for the development and usage of wearable devices for the detection of GCS.[1,2] Studies have utilized specialized wristwatches,[1,3,4,5,6,7,8,9] undermattress devices,[10] and surface electromyogram (EMG) detectors,[11-13] as well as standard scalp[14] and implanted subscalp electroencephalographic (EEG)[15] devices. Physiological data acquisition device capabilities have included accelerometry, heart rate (HR), HR variability, gyroscopy, electrodermal activity, EEG, and EMG activity. Some devices have Bluetooth capability to alert caregivers to seizure occurrences, thus providing seizure alarm systems that aim to reduce morbidity and mortality. Although these are a welcome addition to patient safety mitigation efforts, potential barriers to widespread usage include device efficacy, ease of use, expense, insurance company approval, patient/caregiver adoption of such technology, and fear of stigma. Although sophisticated off-the-shelf (OTS) wearable devices capable of continuous physiological monitoring are already readily available, with 21% of US adults using them regularly,[16] they are not currently configured for seizure detection. Seizure detection capabilities using such devices may provide opportunities for readily adoptable, non-EEG-based seizure detection, automating seizure diaries, and long-term forecasting, without the increased cost of bespoke device technology.[17]

We designed a prospective multicenter, phase 2 study[1] with the primary goal of systematically evaluating the utility of an OTS device, the Samsung (SM-R800) watch, for robust detection of GCS. An OTS acquisition-based GCS detection algorithm was compared against the gold standard epilepsy monitoring unit (EMU) video-EEG-recorded seizure data. Previous studies have used machine learning to detect seizures using classical machine learning approaches such as support vector machine,[4,12,18,19,20] k-nearest neighbor,[5,21] random forest-based algorithms,[5,21] and deep learning algorithms,[3] and limit evaluation to leave-one-patient-out cross-validation (LOPO CV),[22] which does not capture real clinical scenarios. LOPO CV provides an overly optimistic estimate of algorithm performance due to (1) small sample size of testing datasets, (2) lack of independence in training and testing datasets, and (3) strong auto correlation in time series data.[23] Therefore, we used a framework with three different evaluation methods: (1) LOPO CV on GCS patients, (2) testing of false alarm rate (FAR) on patients without seizures, and (3) prospective "fixed-and-frozen"[6] testing. Prospective "fixed-and-frozen" testing is an out-of-sample testing method wherein the hyperparameters of the machine learning algorithm are fixed while evaluating the performance of the classifier, as per the

Key points

- OTS digital watches performed on a par with custom-built wearable devices in detecting GCS.
- Using a machine learning technique, our GCS detection algorithm had a sensitivity of .87 and false alarm rate of .21 per 24 h.
- Evaluation using a multitiered framework ensured reliable performance in our GCS detection algorithm.
- Features based on heart rate extracted from PPG were more important than features extracted from accelerometer or gyroscope data in detecting GCS.
- OTS devices could facilitate the development of automated seizure tracking tools that characterize seizure cycles and possibly forecast seizures.

guidelines of the FDA. A secondary aim was to quantify the contribution of various sensor-based features.

The downstream impact of successful seizure detection by such OTS wearable devices is the ability to carry out long-term (months to years) seizure surveillance and effectuate personalized seizure forecasting in PWE.

2 | MATERIALS AND METHODS

2.1 | Study design

This institutional review board-approved prospective multisite EMU study was conducted at the University of Texas Health Science Center at Houston, the University of Texas Health Science Center at San Antonio, the University of Texas Southwestern Medical Center, and the University of Texas Medical Branch at Galveston (the University of Texas Epilepsy Research Consortium). Each consenting patient who met the inclusion criteria (adult patients who are evaluated or treated for epilepsy) was asked to wear the Samsung (SM-R800) Watch. This OTS device has a three-axis accelerometer sensor (Acc; sampling frequency = 25 samples per second [sps], range = −8 to 8 g-force), three-axis gyroscope sensor (Gyro; sampling frequency = 25 sps, range = −500 to +500 °/s), and photoplethysmographic (PPG) HR (sampling frequency = 1 sps, range = 0–230 beats per minute) sensor. The overall workflow of the GCS detection algorithm is shown in Figure S1A, and an example of the GCS FAR waveform recorded by the device is shown in Figure S1B. Seizure onsets and offsets were determined using video-EEG monitoring (10–20 EEG system or stereo-EEG recording system) by a board-certified epileptologist. The periods during which the OTS device did not establish contact with patients were detected by observing that their HRs registered as zero. Subsequently, these time intervals were excluded from the analysis. The device clocks of video-EEG acquisition (Nihon Kohden EEG 1260 system) and Samsung watch were periodically synchronized using the network time protocol of the same server.

2.2 | Feature extraction

HR was computed from the PPG signal using Samsung's proprietary algorithm. Acc and Gyro signals were high pass filtered (passband frequency = .5 Hz) to remove baseline shifts. The continuous signal was segmented into 7-min segments with 50% overlap between successive windows, to capture the transition from the time course of seizure onset and offset in the same window.

A total of 110 time-domain and frequency-domain features were extracted from the built-in sensors. The significance of the difference in means was quantified using a two-sided Wilcoxon rank-sum test with a significance level of $\alpha = .05$ with Bonferroni corrections (Figure S3), and the mean effect size was computed using Cohen d for two-sample input (Figure S4). Only features that had a significant difference in means between segments containing GCS and segments without GCS were selected to train the machine learning algorithm, reducing the number of features to 68. These features were categorized into three groups: (1) Acc/Gyro time-domain features, (2) Acc/Gyro frequency-domain features, and (3) HR features. A further description of features is provided in Appendix S1. All features were computed using custom scripts implemented in MATLAB R2021a.

To mitigate the imbalance in the number of seizure and nonseizure epochs, a resampled dataset was created by randomly sampling a subset of nonseizure feature sets. The ratio of the number of seizure samples to nonseizure samples was maintained as 1:10. This resampled feature set was used to train the random forest model.

2.3 | Machine learning-based GCS detection algorithm

Random forest is an ensemble learning method that involves training multiple decision trees during the training phase and aggregating the output of the trees to get a classification probability. Decision making using multiple decision trees leads to a reduction in sample variance without increasing the bias when compared to a single decision tree.[24] The algorithm comprised training multiple decision trees on bootstrapped subsamples and aggregating results to get a classification probability per class. We chose random forest because of (1) adaptability, as the algorithm allows determination of feature importance, which can be harnessed to further fine-tune the algorithm on additional training data; (2) scalability, because it has a high degree of parallelizability, which makes handling large datasets efficient; and (3) robustness, because the process of training using multiple trees decreases its sensitivity to noise.[25] In this study, the random forest classifiers with 1000 decision trees were trained on the computed features, and the Gini index was used to calculate root nodes. Algorithm training and testing were implemented using the Sklearn Python library. A full description of the parameters used for training is provided in Appendix S1.

The variable importance measure in a random forest algorithm is estimated based on its contribution to the reduction in Gini impurity averaged over the decision trees. Feature importance is computed for every classifier that is tested during the LOPO CV step and averaged.

2.4 | Postprocessing algorithm

Seizure detection probabilities were estimated for every 3.5-min segment and subjected to a postprocessing algorithm to obtain hourly alarms. The postprocessing algorithm divides the detection probability vector into windows of 1 h with no overlap. If any of the values in the detection probability vector for an hourly window exceeded a predefined threshold, the algorithm labeled the window as containing GCS. As a result, a single label indicating the presence of GCS was generated every hour. The threshold was chosen at .3 to maximize GCS detection while keeping FAR at an acceptable level across all the evaluation methods.

2.5 | Evaluation framework

We divided the EMU patients into three sets, namely, GCS patients, patients who failed to seize during their stay, and a prospective validation patient cohort (patients admitted to the EMU after a specific date). For each group, we used different evaluation criteria to achieve the most precise way to describe the true performance of the algorithm in these populations.

1. GCS patients were subject to LOPO CV;
2. Nonseizure patients were subject to FAR testing (independent testing of FAR); and
3. Patients in the prospective validation cohort were subject to "fixed-and-frozen" prospective testing.

2.6 | LOPO CV on GCS patients

LOPO CV is useful in quantifying algorithm performance while fine-tuning hyperparameters and selecting the best set of features that distinguish GCS waveforms from nonseizure waveforms. LOPO CV was performed by removing one patient from the training set followed by training the classifier on the resampled feature set of the entire recording of the remaining eight GCS patients. The trained algorithm was then tested on the entire recording of the said patient, which the algorithm had not encountered previously. The entire procedure was repeated nine times for nine patients with GCS.

2.7 | FAR testing on nonseizure patients

One of the biggest limitations that prevents widespread adoption of machine learning-based seizure detection algorithms is frequent false alarms.[26] Therefore, FAR testing was performed on patients without seizures. The algorithm was trained on a resampled feature set computed on all nine patients with GCS and tested on the entire recording of the patients without any seizures. False detections were identified using the same threshold that was used to detect GCS.

2.8 | "Fixed-and-frozen" prospective testing on prospective patient cohort

As per the FDA guidelines on detection algorithms,[6] the detection algorithm should be "fixed-and-frozen" and should be tested on newly encountered datasets. The algorithm was trained on GCS patients, and no change was made in hyperparameters after a specific date. The trained algorithm was tested on eight patients who were admitted after a fixed date, two of whom had three GCS during the subsequent recording.

The output of the automatic GCS detection algorithm was compared to GCS labels created by the epileptologist after reviewing video-EEG. An hourly GCS label was classified as true detection if the GCS occurred during that hour. The performance of the algorithm was evaluated using the confusion matrix and computed by pooling detection labels across all patients. The confusion matrix was summarized using accuracy, balanced accuracy, precision, specificity, and sensitivity,[27,28] and 95% Clopper–Pearson confidence intervals (CIs) were calculated for each of the metrics. FAR, defined as the number of false detections in 24 h, was calculated for individual patients, and mean FAR was calculated by averaging across patients, and the interquartile range (IQR) was calculated as 25th and 75th percentile from the distribution of FAR across patients. A full description of the evaluation metrics that were used for testing can be found in Appendix S1.

3 | RESULTS

3.1 | Dataset

This study enrolled 36 adult patients with epilepsy (69% female, mean age = 33.2 ± 12.7 years), who satisfied our inclusion criteria and had at least 24 h of recorded data. The patient demographics are summarized in Table 1 and Figure S7. Eleven patients had GCS. A total of 862 h of data were recorded for patients with GCS and 1410 hours for patients without seizures. A total of 681 hours of data from nine patients were used to optimize the algorithm and evaluation using LOPO CV, 1116 h of data from 19 patients were used for FAR testing, and 680 h of data from

TABLE 1 Patient demographics.

Subject ID	Gender	Age, years	Duration, days	Age at onset, years	Seizure type	Etiology	Number of AEDs	Evaluation strategy
Sub01	F	24	2.06	12	Focal onset	Genetic	1	LOPO CV
Sub02	M	37	5.19	8	Generalized onset	Genetic	5	LOPO CV
Sub03	M	21	4.2	7	Focal onset	Structural	3	LOPO CV
Sub04	F	45	1.29	42	Focal onset	Structural	3	LOPO CV
Sub05	F	46	4.19	1	Focal onset	Structural	3	LOPO CV
Sub06	F	22	3.37	15	Focal onset	Unknown	1	LOPO CV
Sub07	F	36	2.52	4	Focal onset	Unknown	3	LOPO CV
Sub08	F	22	3.15	20	Unknown	Unknown	0	LOPO CV
Sub09	F	23	2.38	13	Generalized onset	Genetic	1	LOPO CV
Sub10	F	32	1.94	30	Unclassified	Genetic	2	FAR testing
Sub11	M	55	1.58	28	Focal onset	Unknown	2	FAR testing
Sub12	F	58	1.11	56	Unknown	Unknown	1	FAR testing
Sub13	F	34	2.05	31	Unknown	Unknown	3	FAR testing
Sub14	F	30	4.24	1	Focal onset	Unknown	4	FAR testing
Sub15	M	45	1.51	39	Focal onset	Unknown	2	FAR testing
Sub16	F	34	2.63	16	Focal onset	Genetic	3	FAR testing
Sub17	F	30	1.05	12	Focal onset	Unknown	3	FAR testing
Sub18	M	21	3.2	3	Focal onset	Unknown	2	FAR testing
Sub19	F	34	2.1	21	Focal onset	Unknown	3	FAR testing
Sub20	F	37	11.4	3	Focal onset	Unknown	4	FAR testing
Sub21	F	19	1.47	17	Unclassified	Unknown	0	FAR testing
Sub22	F	63	2.2	56	Unclassified	Unknown	1	FAR testing
Sub23	F	68	1.23	64	Focal onset	Unknown	2	FAR testing
Sub24	M	40	1.2	22	Focal onset	Unknown	3	FAR testing
Sub25	F	24	1.92	7	Generalized onset	Genetic	2	FAR testing
Sub26	F	21	1.9	2	Focal onset	Genetic	3	FAR testing
Sub27	F	35	2.37	19	Generalized onset	Genetic	2	FAR testing
Sub28	F	37	1.43	32	Focal onset	Unknown	1	FAR testing
Sub29	M	32	5.89	17	Focal onset	Unknown	2	Fixed-and-frozen testing
Sub30	F	22	2.49	2	Unclassified	Unknown	1	Fixed-and-frozen testing
Sub31	M	28	2.15	18	Focal onset	Unknown	3	Fixed-and-frozen testing

TABLE 1 (Continued)

Subject ID	Gender	Age, years	Duration, days	Age at onset, years	Seizure type	Etiology	Number of AEDs	Evaluation strategy
Sub32	M	19	1.72	14	Generalized onset	Genetic	3	Fixed-and-frozen testing
Sub33	F	21	3.46	1	Focal onset	Unknown	1	Fixed-and-frozen testing
Sub34	M	32	1.12	27	Focal onset	Genetic	2	Fixed-and-frozen testing
Sub35	M	29	1.28	9	Focal onset	Genetic	2	Fixed-and-frozen testing
Sub36	F	19	1.69	1	Focal onset	Unknown	3	Fixed-and-frozen testing

Abbreviations: AED, antiepileptic drug; F, female; FAR, false alarm rate; LOPO CV, leave-one-patient-out cross-validation; M, male.

eight patients were used for "fixed-and-frozen" prospective validation.

3.2 | Algorithm performance using LOPO CV on GCS patients

Results are visualized in Figure 1A and summarized in Figure 1D. The algorithm detected GCS better than chance in eight of nine patients and performed with a sensitivity of 1 for seven of nine patients. The sensitivity produced by the algorithm was .87 (95% CI = .62–.96). Eight false alarms were triggered by the algorithm in GCS patients, out of which three false alarms were triggered in two patients (Sub2, Sub8) having focal seizures, resulting in an FAR of .21 (IQR = 0–.90) per 24 h. Evaluation metrics are summarized in Figure 1B,C.

3.3 | Algorithm performance for FAR testing in nonseizure patients

In total, 13 false alarms were triggered over 1116 h of recording, resulting in an FAR of .28 (IQR = 0–.61) per 24 h (Figure 2A,B). Ten of 19 patients (with an average of 45.6 h of recording) had no false alarms, as shown in Figure 2C. False alarms were triggered by patients performing rhythmic activities such as brushing their teeth and exercising.

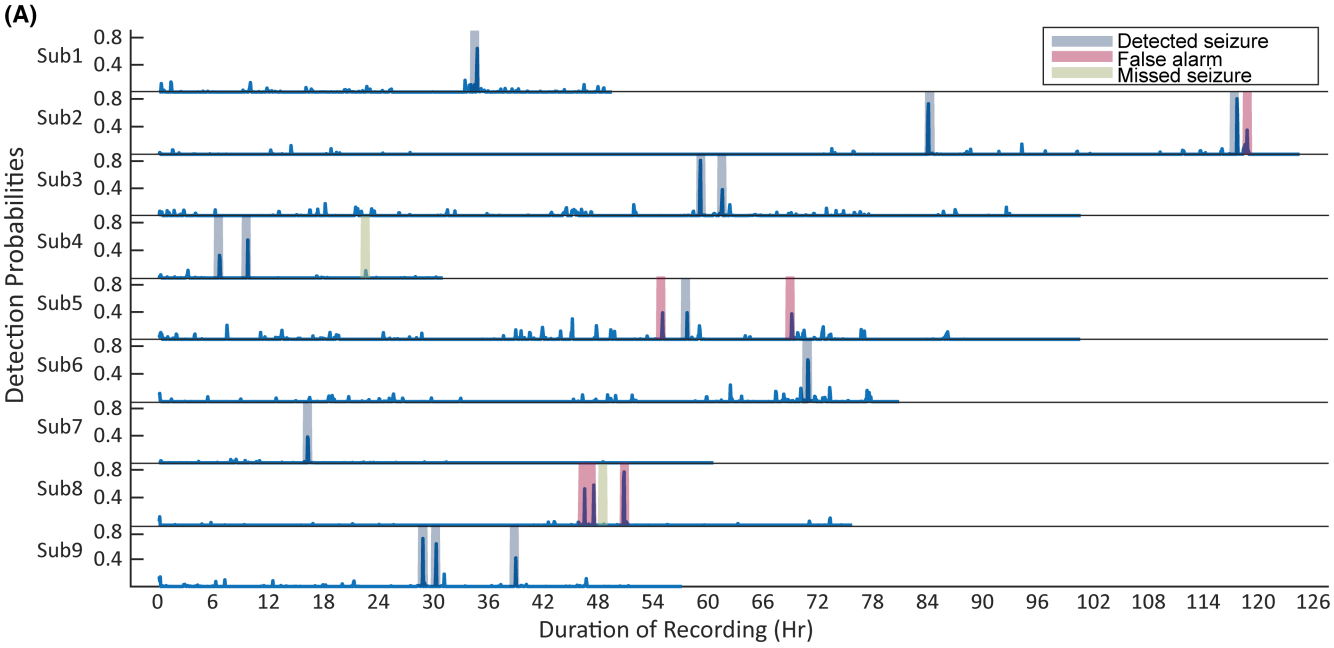
3.4 | Algorithm performance using "fixed-and-frozen" prospective testing on prospective patient cohort

The hyperparameters of the automatic detection algorithm were fixed after the LOPO CV testing stage, and the algorithm was tested on eight patients in the prospective validation cohort (admitted to the EMU after April 1, 2022). Detection probabilities were estimated on these patients, and the time stamps of suspected GCS events were verified with video-EEG monitoring epileptologist annotations. Of eight suspected time points, three were correctly identified as GCS, and five were correctly identified as false alarms, resulting in an FAR of .25 (IQR = 0–.89) per 24 h. The evaluation metrics were calculated and are listed in Figure 3B,C.

3.5 | Quantification of feature importance

The means of the features that were used to train the random forest algorithm were significantly different between GCS and non-GCS epochs (Figure S3). As seen in Figure 4,

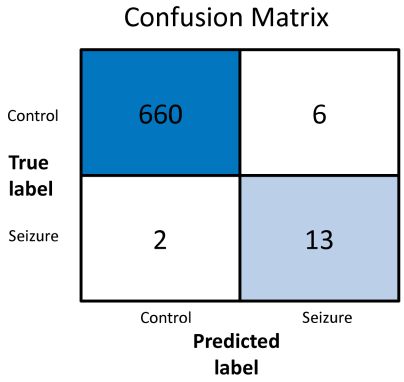
(A)



(B)

Measure	Value
Accuracy [95% CI]	0.99 (0.98-0.99)
Precision [95% CI]	0.68 (0.46-0.85)
Sensitivity [95% CI]	0.87 (0.62-0.96)
Specificity [95% CI]	0.99 (0.98-1.00)
Balanced accuracy [95% CI]	0.93 (0.8-0.98)
False alarm rate (/24 Hr) [IQR]	0.21 (0-0.9)

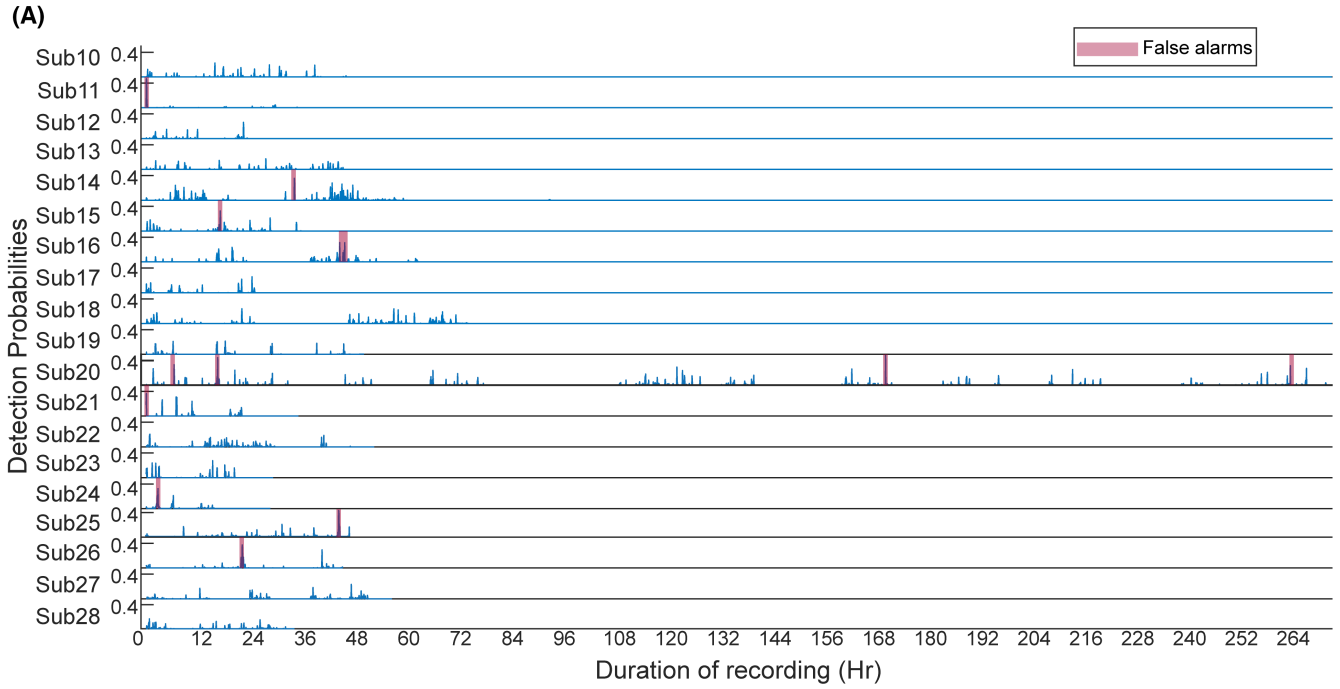
(C)



(D)

Subject ID	Duration (days)	Reported seizure counts	GCS detected	False alarms	Sensitivity	False alarm rate (/24 Hr)
1	2.06	1	1	0	1.00	0.00
2	5.19	2	2	1	1.00	0.19
3	4.20	2	2	0	1.00	0.00
4	1.29	3	2	0	0.67	0.00
5	4.19	1	1	2	1.00	0.48
6	3.37	1	1	0	1.00	0.00
7	2.52	1	1	0	1.00	0.00
8	3.15	1	0	3	0.00	0.95
9	2.38	3	3	0	1.00	0.00
Cumulative	**28.36**	**15**	**13**	**6**	**0.87**	**0.21 (IQR: 0-0.9)**

FIGURE 1 Leave-one-patient-out cross-validation (LOPO CV) results. (A) Subjectwise output of the classifier when tested with LOPO CV on patients with generalized convulsive seizures (GCS). (B) Evaluation metrics with their 95% confidence intervals (CIs) and (C) confusion matrix computed on hourly labels quantifying the performance of the algorithm. (D) Table summarizing the output of the classifier. IQR, interquartile range.



(B)

Subject ID	Duration (days)	False alarms	False alarm rate (/24 Hr)
10	1.94	0	0.00
11	1.58	1	0.63
12	1.11	0	0.00
13	2.05	0	0.00
14	4.24	1	0.23
15	1.51	1	0.66
16	2.63	2	0.76
17	1.05	0	0.00
18	3.20	0	0.00
19	2.10	0	0.00
20	11.40	4	0.35
21	1.47	1	0.68
22	2.20	0	0.00
23	1.23	0	0.00
24	1.20	1	0.84
25	1.92	1	0.52
26	1.90	1	0.53
27	2.37	0	0.00
28	1.43	0	0.00
Cumulative	**46.52**	**13**	**0.28 (IQR: 0-0.61)**

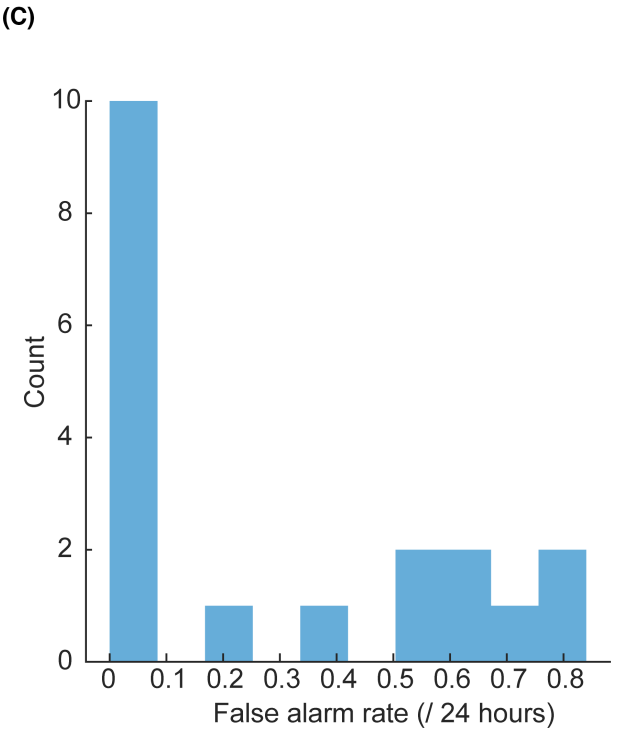
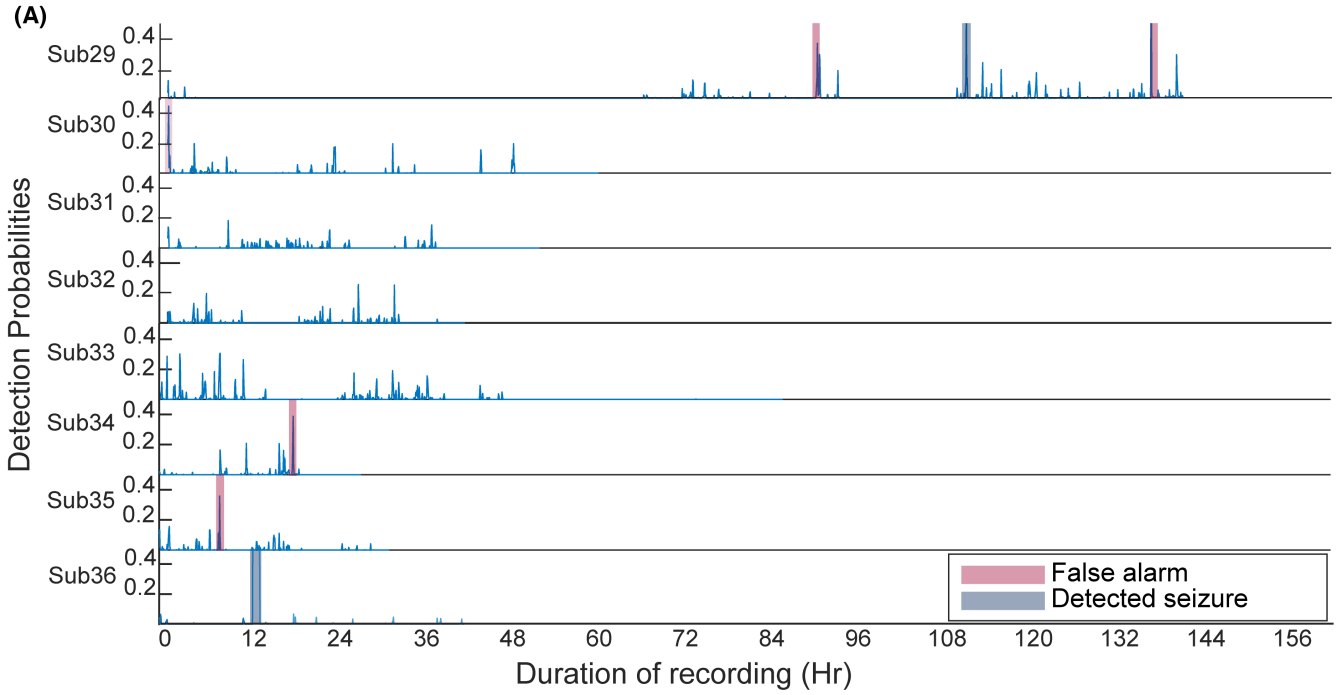


FIGURE 2 Results of quantification of false alarm rate (FAR) in patients without seizures. (A) Subjectwise output of the classifier when tested on patients without seizures. (B) Table summarizing the output of the classifier. (C) Histogram of FAR. IQR, interquartile range.

the feature importance of all features, as computed during LOPO CV, was greater than zero, indicating that every feature is useful in detecting GCS. Acc mean power and power spectral density bands (2.5–4.5 Hz, 4.5–6.5 Hz, 6.5–8.5 Hz, 8.5–10.5 Hz, 10.5–12.5 Hz) showed higher feature importance. Importance of HR variability (HRV) features was observed to be higher than that of Acc or Gyro features, indicating that the HRV features had a more significant impact on the detection of GCS.

4 | DISCUSSION

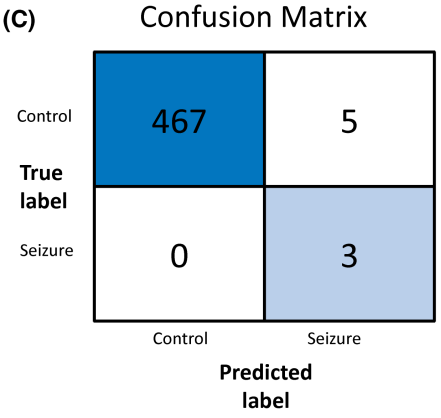
To the best of our knowledge, this represents one of the first prospective, multisite investigational studies utilizing a commercially available off-the-shelf digital watch, aimed at the development and rigorous evaluation of a machine learning-based seizure detection algorithm. Notably, our study employed a novel testing framework designed to address the inherent challenges associated



(B)

Measure	Value
Accuracy [95% CI]	0.99 (0.98-1.00)
Precision [95% CI]	0.38 (0.09-0.76)
Sensitivity [95% CI]	1.00 (0.29-1.00)
Specificity [95% CI]	0.99 (0.98-1.00)
Balanced accuracy [95% CI]	0.99 (0.64-1.00)
False alarm rate (/24 Hr) [IQR]	0.25 (0-0.89)

(C)



(D)

Subject ID	Duration (days)	Reported seizure counts	GCS detected	False alarms	Sensitivity	False alarm rate (/24 Hr)
29	5.89	1	1	2	1	0.34
30	2.49	0	0	1	-	0.40
31	2.15	0	0	0	-	0.00
32	1.72	0	0	0	-	0.00
33	3.46	0	0	0	-	0.00
34	1.12	0	0	1	-	0.89
35	1.28	0	0	1	-	0.78
36	1.69	2	2	0	1	0.00
Cumulative	**19.80**	**3**	**3**	**5**	**1**	**0.25 (IQR: 0-0.89)**

FIGURE 3 Results of "fixed-and-frozen" prospective testing. (A) Subjectwise output of the classifier when performing "fixed-and-frozen" prospective testing. (B) Evaluation metrics with their 95% confidence intervals (CIs) and (C) confusion matrix computed on hourly labels quantifying the performance of algorithm. (D) Table summarizing the output of the classifier. GCS, generalized convulsive seizures; IQR, interquartile range.

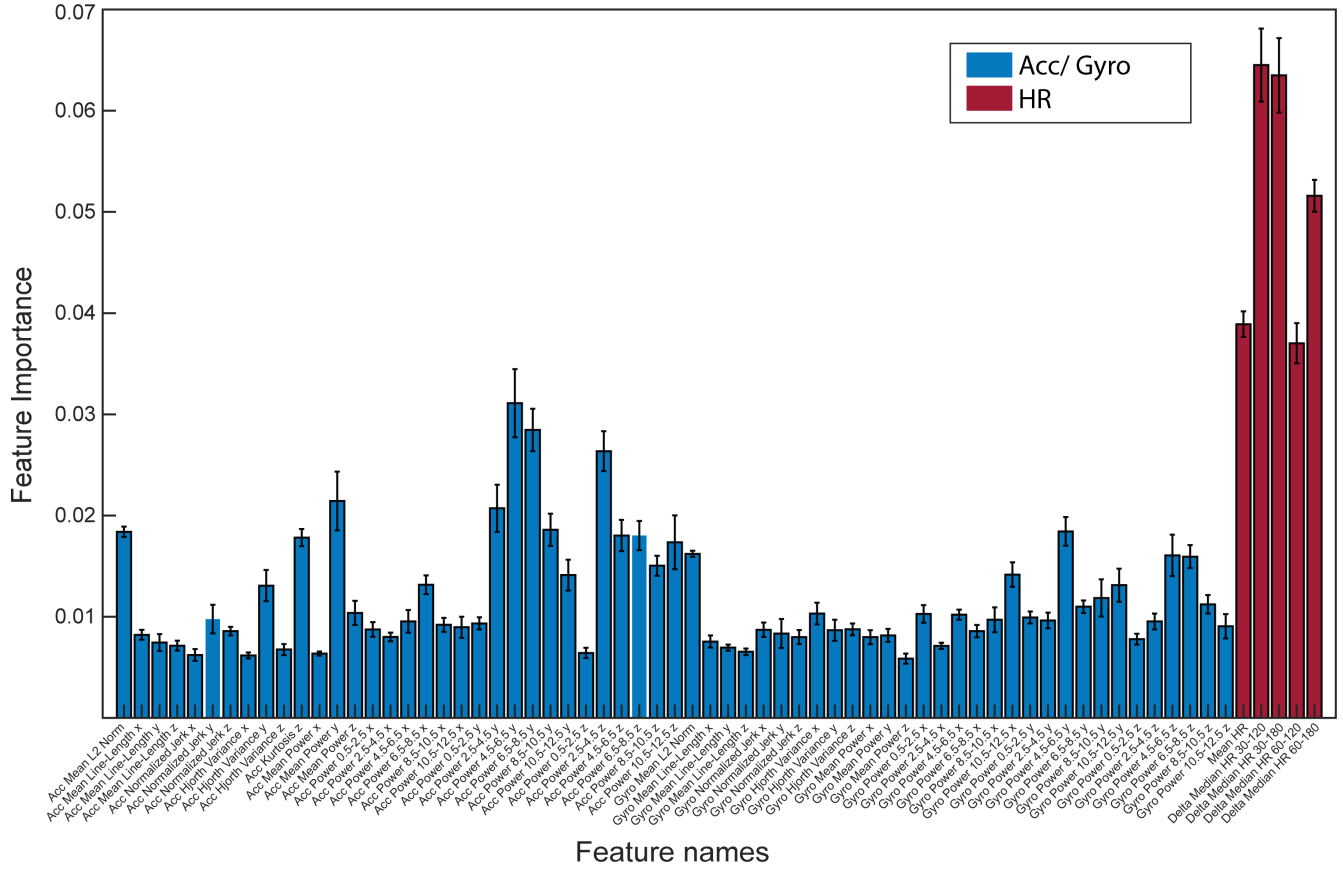


FIGURE 4 Feature importance calculated as mean reduction in Gini impurity averaged over the decision trees for all the classifiers during leave-one-patient-out cross-validation testing. The blue bars represent accelerometer sensor (Acc) and gyroscope sensor (Gyro) features, whereas the red bars represent heart rate (HR) features.

with estimating algorithmic performance on large-scale asymmetric datasets.

As is typical for individuals experiencing seizures, the duration of the actual seizure event is brief and is demarcated by prolonged interictal intervals, thereby giving rise to a highly imbalanced dataset. The new testing frameworks for GCS detection algorithms are vital in a real-world setting due to this problem of extreme class imbalance in long-term datasets. As described by Li et al.,[22] most seizure detection studies using existing evaluation methods report near-perfect algorithm performance but rarely result in routine clinical usage. In the context of wearable-based seizure detection, this is exemplified by overreliance on cross-validation algorithm testing (Table 2). Exceptions are studies[6,29] that focus on prospective testing of proprietary devices that use specialized hardware (but do not always disclose detection algorithm details). Although LOPO CV provides a nearly unbiased estimator of the generalized error,[23] it can provide an overly optimistic estimate of the model's performance due to (1) the small sample size of the testing set, because only one patient is used as the testing set for every iteration of the cross-validation; (2) lack of independence between training and testing set, as both are derived as subsets from the same superset of patient data; and (3) strong autocorrelation inherent in time series data.[23] Moreover, during the LOPO CV, the training and the testing dataset change in every iteration, which leads to the model being retrained during every iteration. However, this is not compatible with practical clinical applications. Once the model is determined and deployed, it rarely changes and does not need to be retrained when applying the existing model to new patient data. In this study, we developed a novel testing framework that leverages the advantages of LOPO, avoids its limitations, and adopts appropriate evaluation strategies for different application scenarios. The prospective testing is comprised of out-of-sample testing, which addresses the lack of independence, minimizes the impact of autocorrelation in the time series data, and tests the model in a practical clinical scenario. Moreover, FAR was estimated using the entire pool of patients who did not have seizures, thus providing a more reliable estimate of FAR. The need for such a multistage testing framework is now recognized, with studies such as Böttcher et al. [30] using a similar testing framework.

While evaluating the GCS detection algorithm using this testing framework, we were able to detect GCS with

TABLE 2 Characteristics of studies that used wearable devices to detect seizures.

Study	Device type	Device	Sensors	Seizure type	Evaluation paradigm	Sensitivity	False alarm rate
Current paper	Off-the-shelf	Samsung watch	HR, Acc, Gyro	Generalized tonic–clonic	1. LOPO CV 2. FAR testing on patients w/o seizures 3. Fixed-and-frozen testing	88% – 100%	.21/24 h .28/24 h .24/24 h
Mittlesteadt et al. (2020)[34]	Off-the-shelf	Fitbit Charge 2	HR	Epileptic seizure	LOPO CV	.58 (AUC)	Not reported
Kusmakar et al. (2019)[20]	Off-the-shelf	Apple iPod touch phone	Acc	Psychogenic nonepileptic seizures, generalized tonic–clonic, complex partial seizures	LOPO CV	95%	.72/24 h
Naganur et al. (2022)[35]	Off-the-shelf	Apple iPod touch phone	Acc	Tonic–clonic, psychogenic nonepileptic seizures	5-fold cross-validation	72.70%	2.43/24 h
Gubbi et al. (2016)[36]	Off-the-shelf	Apple iPod touch phone	Acc	Psychogenic nonepileptic seizures, epileptic seizures	LOPO CV	100%	85.77% (specificity)
Böttcher et al. (2021)[30]	Custom built	Empatica E4	Acc, EDA, PPG	Focal to bilateral tonic–clonic, generalized tonic–clonic	1. LOPO CV 2. FAR testing on patients w/o seizures 3. Fixed-and-frozen testing	100% – 91%	.46/24 h 0 .37/24 h
Tang et al. (2021)[3]	Custom built	Empatica E4	Acc, EDA, PPG	Focal to bilateral tonic–clonic, generalized tonic–clonic	LOPO CV, 10-fold cross-validation	80%	13.63/24 h
Onorati et al. (2021)[6]	Custom built	Empatica E4, Embrace	Acc, EDA, PPG	Generalized tonic–clonic, focal-to-bilateral tonic–clonic	Fixed-and-frozen testing	96%	.83/24 h
Johansson et al. (2019)[21]	Custom built	Shimmer 3	Acc	Tonic–clonic	Fixed-and-frozen testing	100%	1.2/24 h
De Cooman et al. (2018)[18]	Custom built	Individual sensors	Acc, EMG, ECG	Tonic–clonic	Hourly segment based	95.50%	.96/24 h
Meritam et al. (2018)[37]	Custom built	EpiCare Free	Acc	Generalized tonic–clonic	Fixed-and-frozen testing	90%	.1/24 h
Kusmakar et al. (2019)[20]	Custom built	Custom built	Acc	Generalized tonic–clonic	LOPO CV	100%	1.09/24 h
Vandecasteele et al. (2017)[4]	Custom built	Empatica E4	PPG	Complex partial frontotemporal	LOPO CV	32%	26.88/24 h
Onorati et al. (2021)[6]	Custom built	Empatica E3/E4, iCalm	Acc, EDA	Generalized tonic–clonic, focal-to-bilateral tonic–clonic	1/3 training, 1/3 tuning, 1/3 testing	94.55%	.19/24 h
van Andel et al. (2017)[29]	Custom built	Shimmer	Acc, chest leads	Generalized tonic–clonic, hypermotor	Fixed-and-frozen testing	70.90%	6.9/24 h
Velez et al. (2016)[38]	Custom built	SmartMonitor	Acc	Generalized tonic–clonic	Fixed-and-frozen testing	92.30%	Not specified
Heldberg et al. (2015)[5]	Custom built	Empatica E3	Acc, EDA	Motor, nonmotor	LOPO CV	85.30%	96.8% (specificity)

TABLE 2 (Continued)

Study	Device type	Device	Sensors	Seizure type	Evaluation paradigm	Sensitivity	False alarm rate
Patterson et al. (2015)[39]	Custom built	SmartWatch	Acc	Generalized tonic–clonic	Fixed-and-frozen testing	31%	Not specified
Milošević et al. (2014)[12]	Custom built	Wired Acc	Acc, sEMG	Generalized tonic–clonic	LOPO CV	100%	9.32/24 h
Beniczky et al. (2013)[9]	Custom built	EpiCare Free	Acc	Complex-partial, hypermotor, PNES, focal tonic	Fixed-and-frozen testing	90%	.19/24 h
Poh et al. (2012)[19]	Custom built	Custom	Acc, EDA	Generalized tonic–clonic	LOPO CV, leave-one-seizure-out cross-validation	94%	.74/24 h
Kramer et al. (2011)[40]	Custom built	Individual sensors	Acc	Motor	Not specified	90.90%	.11/ 24 h

Abbreviations: Acc, accelerometer sensor; AUC, area under the curve; ECG, electrocardiogram; EDA, electrodermal activity; EMG, electromyogram; FAR, false alarm rate; Gyro, gyroscope sensor; HR, heart rate; LOPO CV, leave-one-patient-out cross-validation; PNES, psychogenic nonepileptic seizures; PPG, photoplethysmogram; sEMG, surface EMG.

a sensitivity of .87 (13/15; 95% CI = .62–.99) and FAR of .21 (IQR = 0–.90) per 24 h during LOPO CV. When tested on 1116 h of recording for patients without a seizure, the algorithm demonstrated an FAR of .28 (IQR = 0–.61) per 24 h. While performing "fixed-and-frozen" prospective testing on eight patients that were not encountered previously, two patients underwent three GCS, which were detected by the algorithm, while generating an FAR of .25 (IQR = 0–.89) per 24 h.

The distribution of feature importance, as computed during LOPO CV, showed that all the selected features contribute to the correct identification of GCS. Among Acc/Gyro features, power spectral density features (2.5–12.5 Hz) showed higher feature importance. This is expected, as clonic periods of GCS would result in characteristic repetitive shaking that contributes to power spectral density features. Notably, HR features were assigned higher feature importance, which is expected, as HRV has been known to increase during GCS.[31] Furthermore, statistical analysis of the mean effect size showed HRV features had larger discriminability in identifying seizure and nonseizure segments as quantified by Cohen *d* effect size (Figure S4).

Most studies use custom-built devices for seizure detection such as EpiCare[32] and Shimmer armbands (Table 2).[21] The FDA-approved Empatica digital watch, which also measures electrodermal activity, requires minimal preprocessing to obtain a continuous sensor data stream and is a popular device choice.[19] However, such devices have exponentially higher price tags and monthly subscription fees compared to OTS devices such as the Samsung watch (~$200). Bespoke device adoption and longevity of use is a significant issue, and device-caused irritation is a frequent complaint.[33] OTS watches are manufactured on a large scale. Users have a large variety of customization options from parent companies as well as smaller accessory manufacturers. Intrusiveness and fear of stigma are also major considerations among PWE.[17] Because OTS devices have become ubiquitous accessories and are used by individuals from diverse backgrounds, with approximately 21% of adults in the United States reporting consistent utilization of such technology,[16] wearing such a device does not carry the same concerns, and hence provides major advantages for the users. Additionally, such devices have a range of other non-health-related functionalities (apps and tools) that such custom-made devices cannot provide, and thus obviate the need for additional, bespoke seizure monitoring wearables. However, OTS watches undergo periodic software updates that could potentially affect the sensor recordings, and therefore impact the performance of the seizure detection algorithm, further emphasizing the importance of developing highly robust algorithms, which are relatively insensitive to noise.

Wearable device testing for seizure detection is not new (Table 2), although the approaches used vary in practical value, and OTS devices are relatively underrepresented. Mittlesteadt et al.[34] developed a machine learning algorithm using an OTS activity tracker, Fitbit Charge 2, for seizure detection. They used a three-layer multilayer perceptron as their machine-learning model and used LOPO CV to test their algorithm on 12 patients with 53 total seizures. They reported that an area under the receiver operating characteristic of .53, indicating that the algorithm performs seizure detection at an above-chance level. However, they noted that low specificity and large interpatient variability made usage infeasible in the routine clinical setting. Another study by Kusmakar et al.[20] used a wrist-worn smartphone (Apple iPod touch) to detect GCS, psychogenic nonepileptic seizures, and complex partial seizures. They used a support vector data description algorithm as their machine learning model and trained it using time-domain features and Poincaré-based features derived from Acc. They trained and tested their algorithm on 21 GCS from 12 patients and obtained 95% sensitivity and an FAR of .72 per 24 h. However, their evaluation method consisted solely of LOPO CV. Although our model showed lower LOPO CV sensitivity, it is tested using a framework that better simulates clinical scenarios and reliably achieves lower FAR. Moreover, wrist-worn smartphones are more cumbersome as well as more expensive, therefore counteracting the primary advantage of using OTS detection devices. A recently published meta-analysis[35] evaluated the performance of a GCS detection algorithm using state-of-the-art wearable devices, which were predominantly custom-built, the only exception being the aforementioned Apple iPod touch. The sensitivity of the detection algorithm was .91 (95% CI = .85–.96) and subtotal FAR was .96/24 h (95% CI = .25–1.66), thus lending strong evidence to the claim that OTS devices perform on a par with custom-built wearable devices.

5 | LIMITATIONS

The scope of this study was limited to accurate GCS offline count detection and not seizure onset detection. Due to the 1-h time resolution of the seizure labels, the current algorithm is not suitable for alerting PWE or their caregivers in the event of GCS. The intended application of such a study is to create a tool that serves as an automated seizure diary that can be used by both clinicians and PWE to track seizures and characterize seizure cycles. Future work will aim to decrease the latency of seizure detection and enable the incorporation of this algorithm as a real-time system, which could be useful in alerting clinicians and caregivers when GCS occur. Additional work on non-GCS seizure detection is currently underway. Another limitation of this study was that patients were confined to the EMU, although this is the ideal setting for the acquisition of seizure data and algorithm validation. Moreover, the hardware and the software version of the device were fixed during the entire period of the study. The current study will be expanded on a larger cohort of patients to capture within-subject variability and the effects of variability in hardware and software versions.

6 | CONCLUSIONS

Although patients regularly performed some of their daily routine activity, false alarms were likely reduced due to the controlled nature of their environment compared to an ambulatory setting. Despite these limitations, this study demonstrates the feasibility of GCS detection using an OTS digital watch. Moreover, this study also provides a framework for testing seizure detection devices in the EMU setting, which could function as a preliminary test before deploying the device in an ambulatory setting.

AUTHOR CONTRIBUTIONS

Conception and design: Jaison S. Hampson, John C. Mosher, Sandipan Pati, Todd Masel, Ryan Hays, Charles Akos Szabo, Guo-Qiang Zhang, and Samden D. Lhatoo. *Acquisition and data analysis:* Yash Shashank Vakilna, Xiaojin Li, Jaison S. Hampson, Yan Huang, John C. Mosher, Yuri Dabaghian, Xi Luo, and Blanca Talavera. *Drafting the manuscript:* all authors.

ACKNOWLEDGMENTS

We thank Alex Gao and Samsung Research America for their help with device, platform, and data support.

FUNDING INFORMATION

This work was supported by the National Institutes of Health (grant R01NS126690).

CONFLICT OF INTEREST STATEMENT

This research received no direct financial support. However, Samsung Research America provided Samsung devices and infrastructure to access data that were used in the study. S.D.L. declares these potential conflicts of interest: associate editor, *Epilepsy Research* (journal); benchmark steward, steward in Area III: Improve treatment options for controlling seizures and epilepsy-related conditions while limiting side effects; benchmark steward, SUDEP Section, National Institute of Neurological Disorders and Stroke. J.C.M. declares these potential conflicts of interest: senior member of the Institute of Electrical and Electronics Engineers (IEEE) Medicine and Biology

Society, Signal Processing Society, and Antennas and Propagation Society; member of Human Brain Mapping; frequent reviewer for the National Institutes of Health study sections; reviewer present and past for *NeuroImage*, *IEEE Transactions on Biomedical Engineering*, *IEEE Transactions on Signal Processing*, *Physics in Medicine and Biology*, *Journal of Clinical Neurophysiology*, and the National Science Foundation; formerly on the editorial review board of *Human Brain Mapping*. None of the other authors has any conflict of interest to disclose.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

The University of Texas Health Institutional Review Board issued approval HSC-MS-23-0356. 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.

PATIENT CONSENT STATEMENT

All the patients that were enrolled in the study provided consent.

ORCID

Yash Shashank Vakilna https://orcid.org/0000-0001-5084-9625
Blanca Talavera https://orcid.org/0000-0001-7692-0186
Sandipan Pati https://orcid.org/0000-0002-9578-2820
Charles Akos Szabo https://orcid.org/0000-0001-6731-3245
Samden D. Lhatoo https://orcid.org/0000-0002-0260-3855

REFERENCES

1. Beniczky S, Wiebe S, Jeppesen J, Tatum WO, Brazdil M, Wang Y, et al. Automated seizure detection using wearable devices: a clinical practice guideline of against epilepsy and the international federation. Clin Neurophysiol. 2021;132(5):1173–84. https://doi.org/10.1016/J.CLINPH.2020.12.009
2. Beniczky S, Ryvlin P. Standards for testing and clinical validation of seizure detection devices. Epilepsia. 2018;59:9–13. https://doi.org/10.1111/epi.14049
3. Tang J, El Atrache R, Yu S, Asif U, Jackson M, Roy S, et al. Seizure detection using wearable sensors and machine learning: setting a benchmark. Epilepsia. 2021;62(8):1807–19. https://doi.org/10.1111/epi.16967
4. Vandecasteele K, de Cooman T, Gu Y, Cleeren E, Claes K, Paesschen WV, et al. Automated epileptic seizure detection based on wearable ECG and PPG in a hospital environment. Sensor. 2017;17(10):1–12. https://doi.org/10.3390/s17102338
5. Heldberg BE, Kautz T, Leutheuser H, Hopfengartner R, Kasper BS, Eskofier BM. Proceedings of the annual international conference of the IEEE engineering in medicine and biology society, EMBS. Vol 2015-Novem. Milan: Institute of Electrical and Electronics Engineers Inc; 2015. p. 5593–6. https://doi.org/10.1109/EMBC.2015.7319660
6. Onorati F, Regalia G, Caborni C, LaFrance WC Jr, Blum AS, Bidwell J, et al. Prospective study of a multimodal convulsive seizure detection wearable system on pediatric and adult patients in the epilepsy monitoring unit. Front Neurol. 2021;12:1444. https://doi.org/10.3389/fneur.2021.724904
7. Onorati F, Regalia G, Caborni C, Migliorini M, Bender D, Poh MZ, et al. Multicenter clinical assessment of improved wearable multimodal convulsive seizure detectors. Epilepsia. 2017;58(11):1870–9. https://doi.org/10.1111/epi.13899
8. Lockman J, Fisher RS, Olson DM. Detection of seizure-like movements using a wrist accelerometer. Epilepsy Behav. 2011;20(4):638–41. https://doi.org/10.1016/j.yebeh.2011.01.019
9. Beniczky S, Polster T, Kjaer TW, Hjalgrim H. Detection of generalized tonic-clonic seizures by a wireless wrist accelerometer: a prospective, multicenter study. Epilepsia. 2013;54(4):e58–e61. https://doi.org/10.1111/EPI.12120
10. Poppel KV, Fulton SP, McGregor A, Ellis M, Patters A, Wheless J. Prospective study of the Emfit movement monitor. J Child Neurol. 2013;28(11):1434–6. https://doi.org/10.1177/0883073812471858
11. Halford JJ, Sperling MR, Nair DR, Dlugos DJ, Tatum WO, Harvey J, et al. Detection of generalized tonic–clonic seizures using surface electromyographic monitoring. Epilepsia. 2017;58(11):1861–9. https://doi.org/10.1111/epi.13897
12. Milošević M, Van De Vel A, Bonroy B, Ceulemans B, Lagae L, Vanrumste B, et al. Detection of epileptic convulsions from accelerometry signals through machine learning approach. MLSP. Honolulu:IEEE Computer Society; 2014. https://doi.org/10.1109/MLSP.2014.6958863
13. Beniczky S, Conradsen I, Henning O, Fabricius M, Wolf P. Automated real-time detection of tonic-clonic seizures using a wearable EMG device. Neurology. 2018;90(5):e428–e434. https://doi.org/10.1212/WNL.0000000000004893
14. Japaridze G, Loeckx D, Buckinx T, Armand Larsen S, Proost R, Jansen K, et al. Automated detection of absence seizures using a wearable electroencephalographic device: a phase 3 validation study and feasibility of automated behavioral testing. Epilepsia. 2022;64(S4):1–7. https://doi.org/10.1111/EPI.17200
15. Ahnood A, Duy Truong N, Fleiss B, Nikpour A, Kavehei O. Novel electrode architecture for subgaleal electroencephalography: a feasibility study. Sensors Diagnost. 2022;1(2):245–61. https://doi.org/10.1039/D1SD00020A
16. Vogels EA. About one-in-five Americans use a smart watch or fitness tracker. Pew Res Cent. 2020;1–6. https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/ Accessed February 7, 2023.
17. Bruno E, Viana PF, Sperling MR, Richardson MP. Seizure detection at home: do devices on the market match the needs of people living with epilepsy and their caregivers? Epilepsia. 2020;61(S1):S11–S24. https://doi.org/10.1111/EPI.16521
18. De Cooman T, Varon C, Van De Vel A, Ceulemans B, Lagae L, Van Huffel S. Comparison and combination of electrocardiogram, electromyogram and accelerometry for tonic-clonic seizure detection in children. 2018 IEEE EMBS Int Conf Biomed Heal Informatics, BHI 2018; Las Vegas: IEEE 2018. p. 438–41. https://doi.org/10.1109/BHI.2018.8333462

19. Poh M-Z, Loddenkemper T, Reinsberger C, Swenson NC, Goyal S, Sabtala MC, et al. Convulsive seizure detection using a wrist-worn electrodermal activity and accelerometry biosensor. Epilepsia. 2012;53(5):e93–e97. https://doi.org/10.1111/J.1528-1167.2012.03444.X
20. Kusmakar S, Karmakar CK, Yan B, O'Brien TJ, Muthuganapathy R, Palaniswami M. Onset detection of epileptic seizures from accelerometry signal. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, 2018;2018:1–4. https://doi.org/10.1109/EMBC.2018.8513669
21. Johansson D, Ohlsson F, Krýsl D, Rydenhag B, Czarnecki M, Gustafsson N, et al. Tonic-clonic seizure detection using accelerometry-based wearable sensors: a prospective, video-EEG controlled study. Seizure – Eur J Epilepsy. 2019;65:48–54. https://doi.org/10.1016/J.SEIZURE.2018.12.024
22. Li X, Tao S, Lhatoo SD, Cui L, Huang Y, Hampson JP, et al. A multimodal clinical data resource for personalized risk assessment of sudden unexpected death in epilepsy. Front Big Data. 2022;5:965715. https://doi.org/10.3389/FDATA.2022.965715
23. Keevers TL. Cross-validation is insufficient for model validation approved for public release. Fishermans Bend:Joint and Operations Analysis Division Defence Science and Technology Group; 2019.
24. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32. https://doi.org/10.1023/A:1010933404324
25. Li X, Tao S, Jamal-Omidi S, Huang Y, Lhatoo SD, Zhang GQ, et al. Detection of postictal generalized electroencephalogram suppression: random forest approach. JMIR Med Inform. 2020;8(2):e17061. https://doi.org/10.2196/17061
26. Frans S, Leijten S, the Dutch TeleEpilepsy Consortium. Multimodal seizure detection: A review. Epilepsia open. 2018;59:42–7. https://doi.org/10.1111/epi.14047
27. Ting KM. Precision and Recall. In: Sammut, C., Webb, G.I. (eds), Encyclopedia of machine learning. Boston:Springer: 2011. https://doi.org/10.1007/978-0-387-30164-8-652
28. Powers DMW. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. 2020 http://arxiv.org/abs/2010.16061
29. van Andel J, Ungureanu C, Arends J, Tan F, van Dijk J, Petkov G, et al. Multimodal, automated detection of nocturnal motor seizures at home: is a reliable seizure detector feasible? Epilepsia Open. 2017;2(4):424–31. https://doi.org/10.1002/EPI4.12076
30. Böttcher S, Bruno E, Manyakov NV, Epitashvili N, Claes K, Glasstetter M, et al. Detecting tonic-clonic seizures in multimodal biosignal data from wearables: methodology design and validation. JMIR Mhealth Uhealth. 2021;9(11):e27674. https://doi.org/10.2196/27674
31. Jeppesen J, Fuglsang-Frederiksen A, Johansen P, Christensen J, Wüstenhagen S, Tankisi H, et al. Seizure detection using heart rate variability: a prospective validation study. Epilepsia. 2020;61(S1):S41–S46. https://doi.org/10.1111/epi.16511
32. Callegari D, Conte E, Ferreto T, Fernandes D, Moraes F, Burmeister F, et al. EpiCare – a home care platform based on mobile cloud computing to assist epilepsy diagnosis. Proceedings of the 2014 4th international conference on wireless Mobile communication and healthcare - "transforming healthcare through innovations in Mobile and wireless technologies", MOBIHEALTH 2014. Volume 2015. Piscataway: Institute of Electrical and Electronics Engineers Inc. p. 148–51. https://doi.org/10.1109/MOBIHEALTH.2014.7015931
33. Hadady L, Klivényi P, Fabó D, Beniczky S. Real-world user experience with seizure detection wearable devices in the home environment. Epilepsia. 2023;64;Suppl 4.S4;S72–S77. https://doi.org/10.1111/EPI.17189
34. Mittlesteadt J, Bambach S, Dawes A, Wentzel E, Debs A, Sezgin E, et al. Evaluation of an activity tracker to detect seizures using machine learning. J Child Neurol. 2020;35(13):873–8. https://doi.org/10.1177/0883073820937515
35. Naganur V, Sivathamboo S, Chen Z, Kusmakar S, Antonic-Baker A, O'Brien TJ, et al. Automated seizure detection with noninvasive wearable devices: a systematic review and meta-analysis. Epilepsia. 2022;63(8):1930–41. https://doi.org/10.1111/epi.17297
36. Gubbi J, Kusmakar S, Rao AS, Yan B, O'Brien T, Palaniswami M. Automatic detection and classification of convulsive psychogenic nonepileptic seizures using a wearable device. IEEE J Biomed Heal Informatics. 2016;20(4):1061–72.
37. Meritam P, Ryvlin P, Beniczky S. User-based evaluation of applicability and usability of a wearable accelerometer device for detecting bilateral tonic–clonic seizures: a field study. Epilepsia. 2018;59:48–52. https://doi.org/10.1111/epi.14051
38. Velez M, Fisher RS, Bartlett V, Le S. Tracking generalized tonic-clonic seizures with a wrist accelerometer linked to an online database. Seizure. 2016;39:13–8. https://doi.org/10.1016/J.SEIZURE.2016.04.009
39. Patterson AL, Mudigoudar B, Fulton S, McGregor A, Van Poppel K, Wheless MC, 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(4):309–11. https://doi.org/10.1016/j.pediatrneurol.2015.07.002
40. Kramer U, Kipervasser S, Shlitner A, Kuzniecky R. A novel portable seizure detection alarm system: preliminary results. J Clin Neurophysiol. 2011;28(1):36–8. https://doi.org/10.1097/WNP.0B013E3182051320

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Vakilna YS, Li X, Hampson JS, Huang Y, Mosher JC, Dabaghian Y, et al. Reliable detection of generalized convulsive seizures using an off-the-shelf digital watch: A multisite phase 2 study. Epilepsia. 2024;65:2054–2068. https://doi.org/10.1111/epi.17974