Received: 5 November 2020 | Revised: 2 June 2021 | Accepted: 2 June 2021

DOI: 10.1111/epi.16967

Epilepsia®

FULL-LENGTH ORIGINAL RESEARCH

Seizure detection using wearable sensors and machine learning: Setting a benchmark

Jianbin Tang[1] | **Rima El Atrache**[2] | **Shuang Yu**[1] | **Umar Asif**[1] | **Michele Jackson**[2] | **Subhrajit Roy**[1,3] | **Mahtab Mirmomeni**[1] | **Sarah Cantley**[2] | **Theodore Sheehan**[2] | **Sarah Schubach**[2] | **Claire Ufongene**[2] | **Solveig Vieluf**[2] | **Christian Meisel**[4,5] | **Stefan Harrer**[1,6] | **Tobias Loddenkemper**[2]

[1]IBM Research Australia, Melbourne, Victoria, Australia

[2]Boston Children's Hospital and Harvard Medical School, Boston, Massachusetts, USA

[3]Google Brain, London, UK

[4]Department of Neurology, Charité-Universitätsmedizin Berlin, Berlin, Germany

[5]Berlin Institute of Health, Berlin, Germany

[6]Digital Health Cooperative Research Centre, Melbourne, Victoria, Australia

Correspondence
Jianbin Tang, IBM Research Australia, Level 23, IBM Tower, 60 City Road, Southbank, 3006, Melbourne, VIC, Australia.
Email: jbtang@au1.ibm.com

Funding information
Epilepsy Research Fund

Abstract

Objective: Tracking seizures is crucial for epilepsy monitoring and treatment evaluation. Current epilepsy care relies on caretaker seizure diaries, but clinical seizure monitoring may miss seizures. Wearable devices may be better tolerated and more suitable for long-term ambulatory monitoring. This study evaluates the seizure detection performance of custom-developed machine learning (ML) algorithms across a broad spectrum of epileptic seizures utilizing wrist- and ankle-worn multisignal biosensors.

Methods: We enrolled patients admitted to the epilepsy monitoring unit and asked them to wear a wearable sensor on either their wrists or ankles. The sensor recorded body temperature, electrodermal activity, accelerometry (ACC), and photoplethysmography, which provides blood volume pulse (BVP). We used electroencephalographic seizure onset and offset as determined by a board-certified epileptologist as a standard comparison. We trained and validated ML for two different algorithms: Algorithm 1, ML methods for developing seizure type-specific detection models for nine individual seizure types; and Algorithm 2, ML methods for building general seizure type-agnostic detection, lumping together all seizure types.

Results: We included 94 patients (57.4% female, median age = 9.9 years) and 548 epileptic seizures (11 066 h of sensor data) for a total of 930 seizures and nine seizure types. Algorithm 1 detected eight of nine seizure types better than chance (area under the receiver operating characteristic curve [AUC-ROC] = .648–.976). Algorithm 2 detected all nine seizure types better than chance (AUC-ROC = .642–.995); a fusion of ACC and BVP modalities achieved the best AUC-ROC (.752) when combining all seizure types together.

Significance: Automatic seizure detection using ML from multimodal wearable sensor data is feasible across a broad spectrum of epileptic seizures. Preliminary results

Jianbin Tang, Rima El Atrache, Stefan Harrer, and Tobias Loddenkemper contributed equally to this study.

show better than chance seizure detection. The next steps include validation of our results in larger datasets, evaluation of the detection utility tool for additional clinical seizure types, and integration of additional clinical information.

KEYWORDS

deep learning, epilepsy, machine learning, multisensor recordings, wearable devices

1 | INTRODUCTION

Epilepsy is a common cause of morbidity and mortality, especially among children, despite advances in management regimens.[1,2] Accurate monitoring and tracking of seizures are important to evaluate seizure burden, recurrence risk, and response to treatment. Outside the hospital, seizure tracking relies on patients' and families' self-reporting, which is often unreliable due to underreporting, seizures missed by caregivers, and patients' difficulties recalling seizures.[3–6] Although long-term video-electroencephalography (EEG) in the epilepsy monitoring unit (EMU) is the gold standard for accurately diagnosing and evaluating epilepsy,[7] it is also time-consuming and costly, can be perceived as stigmatizing, and places a greater burden on patients and caregivers than seizure monitoring with wearable devices. Based on prior studies, there exists a large clinical gap and urgent medical need to detect a broad range of seizures, beyond focal to bilateral tonic–clonic seizures (FBTCSs) and generalized tonic–clonic seizures (GTCSs), with wearable devices.[3,8–10]

Recent advances in the use and development of non-EEG-based seizure detection devices utilizing a variety of sensors and modalities provided innovative opportunities to fill this gap and to monitor patients continuously in the outpatient setting. Examples include wrist-worn or arm-worn devices, devices worn on the chest, and mattresses,[11–19] which, among others, are less stigmatizing and better tolerated by patients during extended use.[11,20–23] Furthermore, analyzing signal recordings using artificial intelligence learning algorithms led to improved performance of seizure detection systems.

Machine learning (ML) algorithms are trained to automatically detect signal patterns of epileptic seizures. Previous studies have demonstrated the feasibility of using ML, and in particular deep learning models, to automatically detect seizures and classify seizure types based on EEG data.[24,25] Although similar ML approaches applied to monitoring data from mobile devices can detect generalized tonic and tonic–clonic seizures,[26] detection of other seizure types using data from wearable devices is limited. There exists a critical need for reliable, automatic, nonintrusive methods to detect additional clinical seizure types. Based on the analysis of a large clinical seizure dataset of wrist- and ankle-worn wearable device data confirmed by time-synchronized EEG, we aimed to evaluate seizure detection with wearable devices for a broad spectrum of epileptic seizures. We hypothesized that detection of a variety of seizure types solely based on wrist- and ankle-worn sensor data is feasible, utilizing EEG seizure onset and offset annotations as a gold standard comparison.

Key Points

- We evaluated the seizure detection performance of custom-developed machine learning algorithms across a spectrum of nine epileptic seizure types
- We analyzed electrodermal activity, accelerometry, and photoplethysmography recorded by wrist- and ankle-worn wearable devices
- ACC and BVP data fusion with CNN algorithms outperformed single data modality, with an overall AUC-ROC of .752 when applied to the complete dataset
- Machine learning models trained using all seizures regardless of their type performed better overall than models trained on specific seizure types
- Automatic epileptic seizure detection using machine learning and wearable device data is feasible across a broad spectrum of motor and nonmotor seizures

2 | MATERIALS AND METHODS

We obtained approval from the institutional review board of Boston Children's Hospital before enrollment and data acquisition. We prospectively enrolled patients admitted to the Boston Children's Hospital EMU between February 2015 and November 2017. We asked participants to wear sensors (E4; Empatica) on either the left or right wrist or ankle for long-term recording during their admission. The wearable devices recorded body temperature (TEMP), electrodermal activity (EDA), accelerometry (ACC), and photoplethysmography to provide blood volume pulse (BVP; Figure S1a). We obtained written informed consent or assent from all participants or their guardians.

2.1 | Patient and seizure selection criteria

We extracted clinical data from the electronic medical records with a standardized data collection tool in Microsoft Excel and REDCap (Research Electronic Data Capture; Vanderbilt University), a secure web-based application that facilitates data acquisition and storage. We collected demographic data, clinical patient data, and seizure characteristics. Epileptic seizure semiology and etiology were classified according to the International League Against Epilepsy criteria,[27] and patients with a high hourly frequency of seizures or clusters of four or more seizures in 15-min time windows were excluded.

2.2 | Seizure data collection

This study utilized the conventional 10–20 electrode scalp EEG system during video-EEG monitoring.[28] Video-EEG review by a board-certified clinical epileptologist confirmed electrographic seizure onset and offset annotations as well as semiology of seizure types.

2.3 | Synchronization of EEG and wearable device clocks

The wearable device and EEG monitor record with independent clocks. To compensate for the time drift between the device and EEG clocks, we synchronized the device clocks at the start of the device recording. After we turned on the recording device, we simultaneously pressed the device button and EEG event marker button. This created a time tag on both the device and EEG recordings. When we placed two wearable devices on the patient, the button press was done simultaneously for both. We recorded a video of the syncing process and used video timestamps to verify device placement times in cases where we could not find EEG event markers due to button press failures. When patients agreed to wear the recording device again on a second day or in another admission, they were enrolled again, and the process was repeated.

We defined the start timing error between the wristband and EEG clock as the time difference between the wristband and EEG clock at the beginning of the experiment (Figure S1b). Figure 1A shows the distribution of start timing errors observed in our study; absolute errors follow approximately a Gaussian distribution and are predominantly shorter than 20 s.

The wearable devices also have button press failures, which are indicated by a large start timing error. We set the start timing error to 0 when there is a button press failure. To ensure the seizure segment is within the annotation, we added 20-s preseizure and 20-s postseizure end windows. Figure 1B shows the timing drift of the Empatica E4 for 11 different devices at different time points, confirming device-agnostic frequency stability of 1 ppm[12] and a constant timing drift of 13 s in 24 h. These observations allowed us to derive the following three-step timing error compensation scheme: (1) 13-s offline timing drift compensation was added to the signal data; (2) the start timing error was calculated (Figure S1b), and labels were adjusted accordingly; and (3) 20-s preseizure and 20-s postseizure end windows were added. In a final step, we added verified seizure annotations to the signal data.

2.4 | Data quality check

We downloaded EEGs and wearable sensor device data separately and manually labeled each dataset. Subsequently, we reviewed data for possible missing or incorrect annotations, recording failures, or missing data by running seizure annotations and wearable sensor device data through custom-developed quality check processes (Figure S2).

Patients occasionally removed wristbands. In this study, we used TEMP sensor data to detect whether the wristband had been removed. We used a 10-min moving average to smoothen the TEMP recordings. For temperature values higher than 45° or lower than 27°, we added 10-min safety windows to both ends of the segment, and we excluded this time window from the study.

2.5 | Data analysis

2.5.1 | ML techniques

We used a convolutional neural network (CNN) to classify raw time series data and applied random undersampling to balance the raw data. We first trained different detection models separately on individual ACC, EDA, and BVP data. Then we serially trialed different combinations of sensor modalities, including ACC and EDA, ACC and BVP, EDA and BVP, and ACC, BVP, and EDA. The entire ML framework and the architectures and functionalities of the developed CNN models are shown in Figure 2.

Our CNN model consists of two convolutional layers. Convolutional Layer 1 is followed by Rectified Linear Unit (ReLU) and dropout operations. Convolutional Layer 2 is followed by ReLU and max pooling. The final layer is a global average pooling[29] layer. The network outputs are probabilistic distributions related to the target classes.

In Poh et al.,[30] data reduction and support vector machines (SVMs) were used to detect secondarily generalized tonic–clonic seizures with 94% sensitivity and a .74/24-h false alarm rate (FAR). We applied this method to tonic–clonic

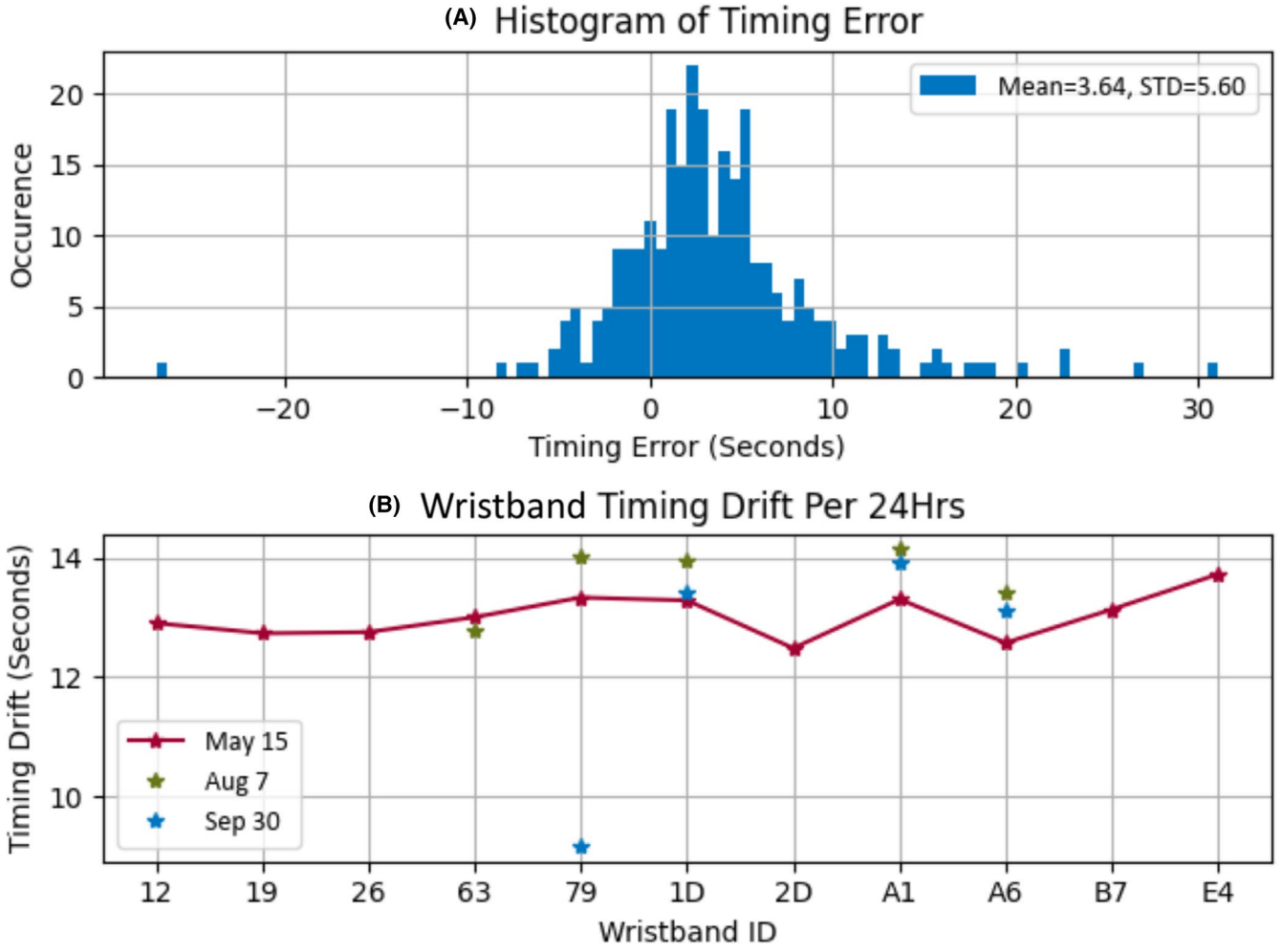

FIGURE 1 Wristband and electroencephalographic (EEG) monitor clocks show different drift rates over time. For seizure detection with seconds-level accuracy, the timing error must be measured and compensated to enable consistent data labeling and analysis consistent with video-EEG recordings. (A) Relative start timing error distribution between the wristband and EEG monitor. (B) Timing drift of the wristband mapped over 24 h is plotted. STD, standard deviation

seizure data in our dataset and compared the performance to that of our CNN model for the same data.

2.5.2 | Seizure detection scenarios

We trained and validated ML methods for two different scenarios: (1) ML methods for developing seizure type-specific detection models for nine individual seizure types; and (2) ML methods for building general type-agnostic seizure detection models, in which case seizures from different seizure types were lumped together.

To ensure the availability of sufficient data to train the seizure type-specific ML models (Scenario 1), we chose to apply leave-one-subject-out cross-validation to those seizure types for which datasets contained more than five patients and in total 10 seizures. This allowed us to test the performance of the developed ML models for nine seizure types (Table 1).

For validating the general seizure detection method (Scenario 2), we applied 10-fold cross-validation to evaluate the performance of the developed ML models for all patients and seizure types. For each fold, the patients in the validation dataset are excluded from the training dataset.

The majority of patients in our cohort wore two functioning sensor devices on opposite sides of the body on their wrists and/or ankles. For some patients, one device stopped working during the study due to battery life limitations. Additionally, some patients chose to wear only one device. We treated each wearable sensor device recording of a seizure as an individual sample during training, which allowed us to maximize the amount of training data. During testing, we used only one wristband recording for those seizures with two simultaneous wristband recordings. We strictly excluded all seizure samples contained in the training dataset from the respective test dataset at all times, and therefore model performance was never tested on any data that had been used for training such models.

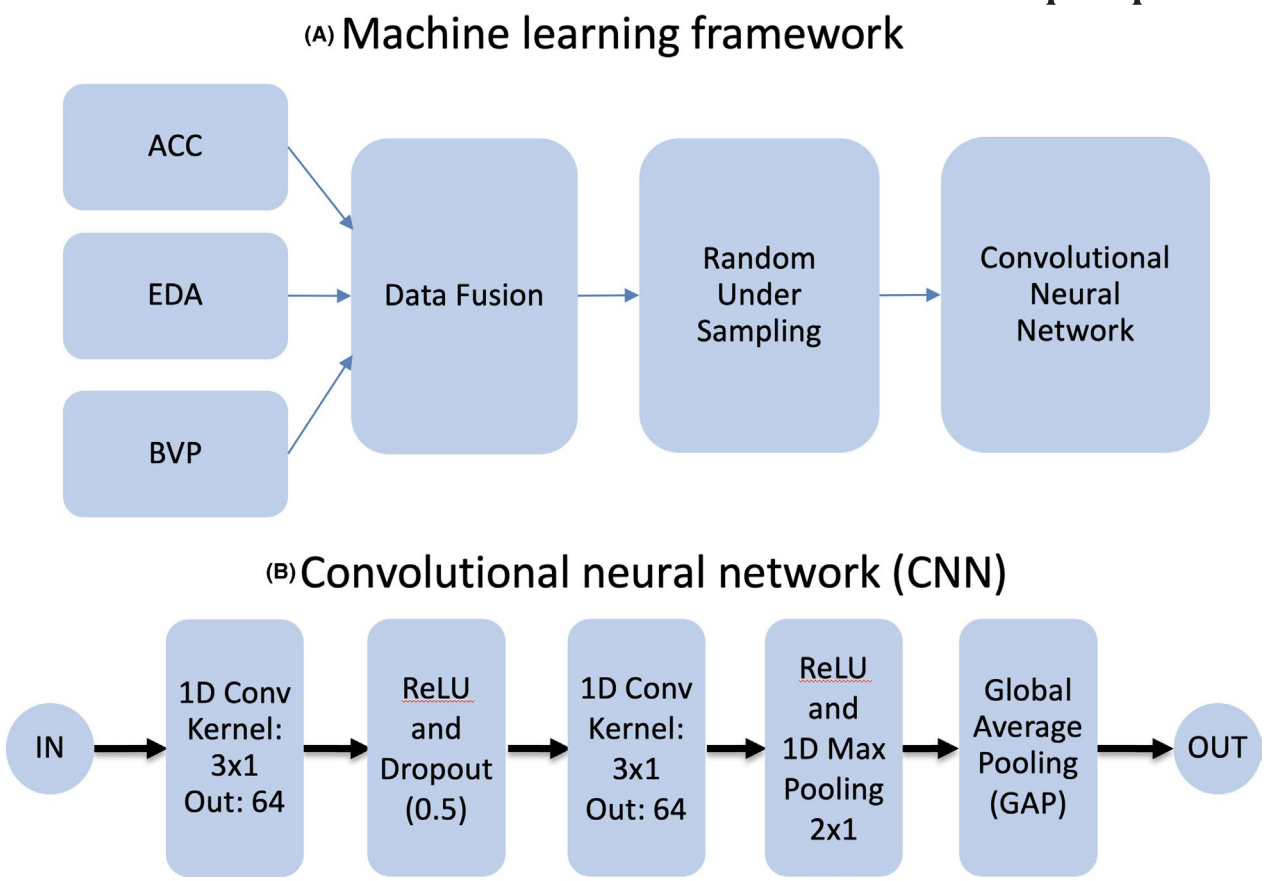

FIGURE 2 FIGURE (A) Machine learning (ML) framework for computing seizure detection baseline performance. The full pipeline depicted columnwise from left to right consists of the following modules: (1) individual sensor modalities generating raw time series data, (2) modality fusion techniques, (3) data sampling methods to compensate for data imbalance, and (4) ML techniques adapted to raw time series. (B) A convolutional neural network (Conv: Convolutional, ReLU: Rectified Linear Unit) is used to analyze raw time series data. ACC, accelerometry; BVP, blood volume pulse; EDA, electrodermal activity

2.6 | Statistical methods

Previous studies[26,31] utilized custom-defined seizurewise performance metrics to account for specific experimental scenarios, data formats, and clinical use cases. We also custom-defined and applied seizurewise performance metrics. We added 1-min preictal and postictal buffers before and after each ictal segment. We consider a seizure detected if it is correctly labeled within the preictal, ictal, and postictal windows.

We divide the interictal period into nonoverlapping 1-min intervals. If multiple false alarms occurred within a single 1-min interval, they are reported as one single false alarm.

The sensitivity measures apply to the proportion of correctly identified seizures. FAR is the number of incorrectly detected seizures over 24 h. We defined detection delay as the difference between the seizure onset and detector recognition of seizure activity.

We calculated seizurewise area under the curve (AUC)–receiver operating characteristic (ROC) value, sensitivity, and FAR per 24 h. The ROC curve is created by plotting the true positive rate (also known as sensitivity) against the FAR at various threshold settings. AUC is the area under the ROC curve. The higher the AUC, the better the model is at distinguishing seizures from nonseizure events.

For leave-one-subject-out cross-validation, we calculated sensitivity, FAR, detection delay, and AUC-ROC by combining all subjects' prediction results. For leave-one-subject-out cross-validation, we calculated sensitivity, FAR, detection delay, and AUC-ROC by combining all folds' prediction results.

3 | RESULTS

We included 94 patients who met our inclusion criteria (median age = 9.9 years, range = 27.2, interquartile range = 9.2; 54 [57.4%] males; Table 1, Figure S3), and 548 seizures. For training purposes, we had seizure onset and offset annotations for 930 seizures captured by both left and right sensor devices. The dataset is highly imbalanced, containing 153 h of ictal and 10 913 h of normal interictal brain activity signals.

TABLE 1 Demographic and clinical characteristics for patients with seizures during enrollment ($N = 94$ patients, $n = 548$ seizures[a])

Demographic and clinical characteristics	Value
Age at first enrollment, median years (range, IQR)	9.9 (27.2, 9.2)
Male, *n* (%)	54 (57.4)
Ethnicity, *n* (%)	
Not Hispanic or Latino	71 (75.5)
Unknown	9 (9.6)
Not reported	8 (8.5)
Hispanic or Latino	6 (6.4)
Race, *n* (%)	
White	64 (68.1)
Unknown	17 (18.1)
Black or African American	6 (6.4)
Not reported	6 (6.4)
Asian	1 (1.1)
History of clinical epilepsy characteristics	
Diagnosis of epilepsy, *n* (%)	92 (97.9)
Age at first seizure, median years (range, IQR)	1.8 (16.0, 6.6)
Seizure frequency [$n = 82$], median seizures/30 days (range, IQR)[b]	32.2 (2880.0, 119.2)
Epilepsy etiology [$n = 92$], *n* (%)	
Unknown	39 (42.4)
Structural	34 (37.0)
Genetic	13 (14.1)
Infectious	3 (3.3)
Metabolic	2 (2.2)
Immune	1 (1.1)
2017 ILAE seizure semiology, *n* patients (*n* seizures)[c]	
Focal onset	62 (548)
Focal to bilateral tonic–clonic[e]	21 (38)
Subclinical[e]	14 (66)
Awareness and motor semiology unavailable	6 (7)
Aware	6 (9)
Motor	4 (4)
Tonic[e]	2 (2)
Clonic[e]	1 (1)
Automatisms[e]	1 (1)
Nonmotor	2 (5)
Sensory	2 (5)
Impaired awareness	28 (69)
Motor	21 (49)
Automatisms[e]	10 (21)
Tonic[e]	7 (18)
Hyperkinetic	4 (6)
Clonic[e]	4 (4)
Nonmotor	10 (20)
Behavior arrest[e]	10 (16)
Cognitive	1 (4)
Unclassified awareness	22 (145)
Motor	16 (78)
Tonic[e]	6 (47)
Automatisms[e]	3 (4)
Hyperkinetic	3 (3)
Myoclonic	3 (14)
Clonic[e]	2 (5)
Atonic	1 (5)
Nonmotor	5 (57)
Behavior arrest[e]	3 (5)
Autonomic	1 (50)
Unclassified	1 (1)
Sensory	1 (1)
Unclassified movement	5 (10)
Generalized onset	35 (213)
Subclinical	2 (2)
Motor semiology unavailable	4 (5)
Motor	30 (174)
Tonic[e]	15 (90)
Epileptic spasms[e]	8 (47)
Tonic–clonic[e]	6 (15)
Myoclonic	3 (9)
Clonic	2 (7)
Atonic	2 (5)
Unclassified	1 (1)
Nonmotor	2 (16)
Typical absence	2 (16)
Unclassified movement	6 (16)
Unknown onset/unclassified movement	1 (1)
Wristband location, *n* patients[c,d]	
Left wrist	35
Right wrist	40
Left ankle	94
Right ankle	59

Abbreviations: ILAE, International League Against Epilepsy; IQR, interquartile range.

[a]In total, 930 seizures were captured when seizures from both left and right sensor devices are combined.

[b]Median number of seizures in 30 days before first enrollment.

[c]Patients may be represented in more than one category.

[d]Wristband location may change over the course of an enrollment period.

[e]Seizure type included in the seizure-specific analysis.

In Algorithm 1, we trained seizure-type specific detection models for nine seizure types based on the following inclusion criteria: FBTCSs, GTCSs, focal tonic seizures, focal subclinical seizures, focal automatisms, focal behavior arrest, focal clonic seizures, generalized tonic seizures, and generalized epileptic spasms (Table 2). We found better than chance detection for eight of nine seizure types.

In Algorithm 2, we lumped all seizures from different seizure types together to build a general type-agnostic seizure detection system. For this scenario, we also display individual seizure type-specific detection performance and compare it to Algorithm 1 (Table 3). We found better than chance detection potential for all nine seizure types, including seizures with and without robust motor components.

3.1 | Algorithm 1: Specific ML model for individual seizure types

We trained a specific model for each seizure type with leave-one-subject-out cross-validation (Table 2). ACC performed best for focal tonic seizures and BVP for focal behavior arrest and generalized tonic seizures. ACC + BVP data fusion provided the best averaged AUC-ROC performance. The model detected GTCSs and FBTCSs with an AUC-ROC of .976 and .932, respectively. The algorithm detected focal tonic seizures, focal automatisms, focal behavior arrest, generalized tonic seizures, and focal clonic seizures with an AUC-ROC of more than .6, suggesting better than chance detection. AUC-ROC for focal subclinical seizures was less than .6, indicating a poor, close to chance detection performance of our algorithm for this seizure type.

3.2 | Algorithm 2: Generalized ML model for all seizure types

We trained a model with all 94 patients and 548 seizures with 10-fold cross-validation, shown in Table 3. Except for EDA, the overall AUC-ROC performance was greater than .6 for all modalities, yielding better than chance detection performance. ACC and BVP modalities reached AUC-ROC values of .72 and .744, respectively, and ACC + BVP data fusion reached the highest overall AUC-ROC of .752.

3.3 | Comparison of Algorithms 1 and 2

We calculated the AUC-ROC performance comparison between Algorithm 1 (type-specific ML model for individual seizure types) and Algorithm 2 (type-agnostic generalized ML model) for all seizure types (Figure 3). For all three sensors, the averaged results show that the generalized ML model for all seizure types performs better than the specific ML model for individual seizure types.

3.4 | Impact on EDA shift

We visualized seizure annotations and sensor data. EDA shows a delayed response time. Figure S4a shows two examples of EDA increasing after a seizure. We compared the performance of the EDA sensor without shifting and with shifting 120 s ahead. Figure S4b shows AUC-ROC for different seizure types, and EDA shifting shows an AUC-ROC improvement for FBTCSs (.716 vs. .516) and GTCSs (.830 vs. .519) without a major impact on other seizure types.

TABLE 2 Leave-one-subject-out performance

Seizure type	ACC	EDA	BVP	ACC + BVP	ACC + EDA	BVP + EDA	ACC + EDA + BVP
Focal to bilateral tonic–clonic	.921	.712	.888	.921	**.932**	.876	.910
Focal tonic	**.786**	.570	.751	.776	.671	.603	.754
Focal subclinical	.548	.550	.496	.528	.488	.537	.504
Focal automatisms	.688	.728	.682	.750	.743	**.806**	.795
Focal behavior arrest	.635	.415	**.706**	.678	.557	.619	.594
Focal clonic	.516	.268	.648	.534	.420	.534	.396
Generalized epileptic spasms	.594	.480	.627	**.633**	.507	.617	.583
Generalized tonic	.588	.507	**.814**	.770	.519	.741	.687
Generalized tonic–clonic	.975	.830	.904	.945	**.976**	.933	.965
All nine seizure types	.673	.559	.716	**.721**	.613	.679	.682

Note: Leave-one-subject-out performance of detection models trained on individual modality data (Columns 1–3) and multimodality data fusion (Columns 4–7). In each row the best AUR-ROC value is highlighted in bold. An AUC-ROC less than .6 is not significantly better than random guess. Although ACC and BVP modalities performed best for some specific seizure types, in general, ACC + BVP data fusion provided the best averaged AUC-ROC performance.

Abbreviations: ACC, accelerometry; AUC-ROC, area under the receiver operating characteristic curve; BVP, blood volume pulse; EDA, electrodermal activity.

TABLE 3 10-fold cross-validation performance

Seizure type	ACC	EDA	BVP	ACC + BVP	ACC + EDA	BVP + EDA	ACC + EDA + BVP
Focal to bilateral tonic–clonic	**.919**	.662	.886	.910	.905	.862	.890
Focal tonic	**.812**	.624	.736	.772	.789	.719	.758
Focal subclinical	.555	.429	**.642**	.623	.520	.603	.568
Focal automatisms	.541	.699	**.811**	.761	.772	.807	.780
Focal behavior arrest	**.765**	.532	.693	.713	.730	.593	.737
Focal clonic	.564	.588	.830	**.762**	.593	.758	.668
Generalized epileptic spasms	**.840**	.450	.711	.831	.796	.632	.789
Generalized tonic	.662	.565	.779	.746	.698	.661	.704
Generalized tonic–clonic	**.995**	.802	.889	.992	.987	.939	.990
All nine seizure types	.720	.549	.744	**.752**	.695	.672	.705

Note: 10-fold cross-validation performance of detection models trained on individual modality data (Columns 1–3) and multimodality data fusion (Columns 4–7). In each row the best AUR-ROC value is highlighted in bold. An AUC-ROC less than .6 is not significantly better than random guess. Although ACC and BVP performed best for selected seizure types, in general, ACC + BVP data fusion provided the best overall AUC-ROC performance, as shown in the last row.

Abbreviations: ACC, accelerometry; AUC-ROC, area under the receiver operating characteristic curve; BVP, blood volume pulse; EDA, electrodermal activity.

We applied EDA shifting for 120 s to all fusions of sensor modalities.

4 | DISCUSSION

4.1 | Summary

In our study, we used a seizure-level evaluation for seizure detection and found that fusing ACC and BVP data modalities achieved the best AUC-ROC of .752 applied to all seizure samples lumped together across all seizure types when we trained one generalized model with all seizure types. Results show the feasibility of seizure detection across a broad spectrum of seizure types, including but not limited to GTCSs, with wrist-worn wearable sensors in a large cohort of patients. Our study expands current literature by demonstrating that noninvasive, wrist- and ankle-worn sensors and custom-developed deep learning techniques can automatically detect a variety of epileptic seizure types.

Numerous previous studies have shown the feasibility of using ML and, in particular, deep learning models to automatically detect,[24] classify,[25] and predict[5,32] epileptic seizure episodes in EEG signals. Initial studies using electrodes implanted in the brain have increasingly been followed with less intrusive systems measuring scalp EEG data.[33] The advent of commercially available medical-grade wearable sensors and mobile smart devices, such as smartwatches, provided an opportunity to assess the suitability of other data modalities for seizure detection.[8] Some of these explorations have shown promise for detecting GTCSs,[23,26,31,34] and based thereon smartwatches have first received US Food and Drug Administration approval and medical clearance for use as epilepsy monitoring devices.[35,36]

4.2 | Performance comparison for GTCSs

We compared a previously utilized detection method[30] with a CNN to detect GTCSs. In the previous study, the authors extracted 19 features from ACC and EDA, then applied SVM, and reached an AUC-ROC of .896, a sensitivity of 73%, and a FAR of 12.52/24 h. Our CNN reached an AUC-ROC of .957, a sensitivity of 80%, and a FAR of 13.63/24 h. The CNN performed better than Poh et al.'s method, as shown by the 5.9% higher AUC-ROC value. However, the performance of the CNN is lower than the SVM reported in the mentioned study[30] (94% sensitivity and .74/24h FAR). One main reason could be that our dataset contains mainly children, a cohort that tends to exhibit pronounced movement activity during seizure-free times. Additionally, the study reported a median electrographic latency of 42.95 s. In our study, the average latency is 51.67 s. Onorati et al.[26] included 55 convulsive epileptic seizures (six focal tonic–clonic seizures and 49 FBTCSs) from 22 patients and achieved 83.6% sensitivity and .29/24h FAR. Their study[26] included 24 children and 45 adults, with an age range of 4–18 and 19–60 years, respectively, including patients who did not experience seizures during the experiment. In comparison, our patients are children, with only three patients older than 20 years. Once monitoring data from children and adolescents are included in a dataset, seizure detection may become more challenging. The algorithm in Onorati et al.[26] failed to detect all three seizures of a 4-year-old child.

4.3 | Effect of detection modality

We expected that models trained on specific seizure types would perform better than models trained on the combined

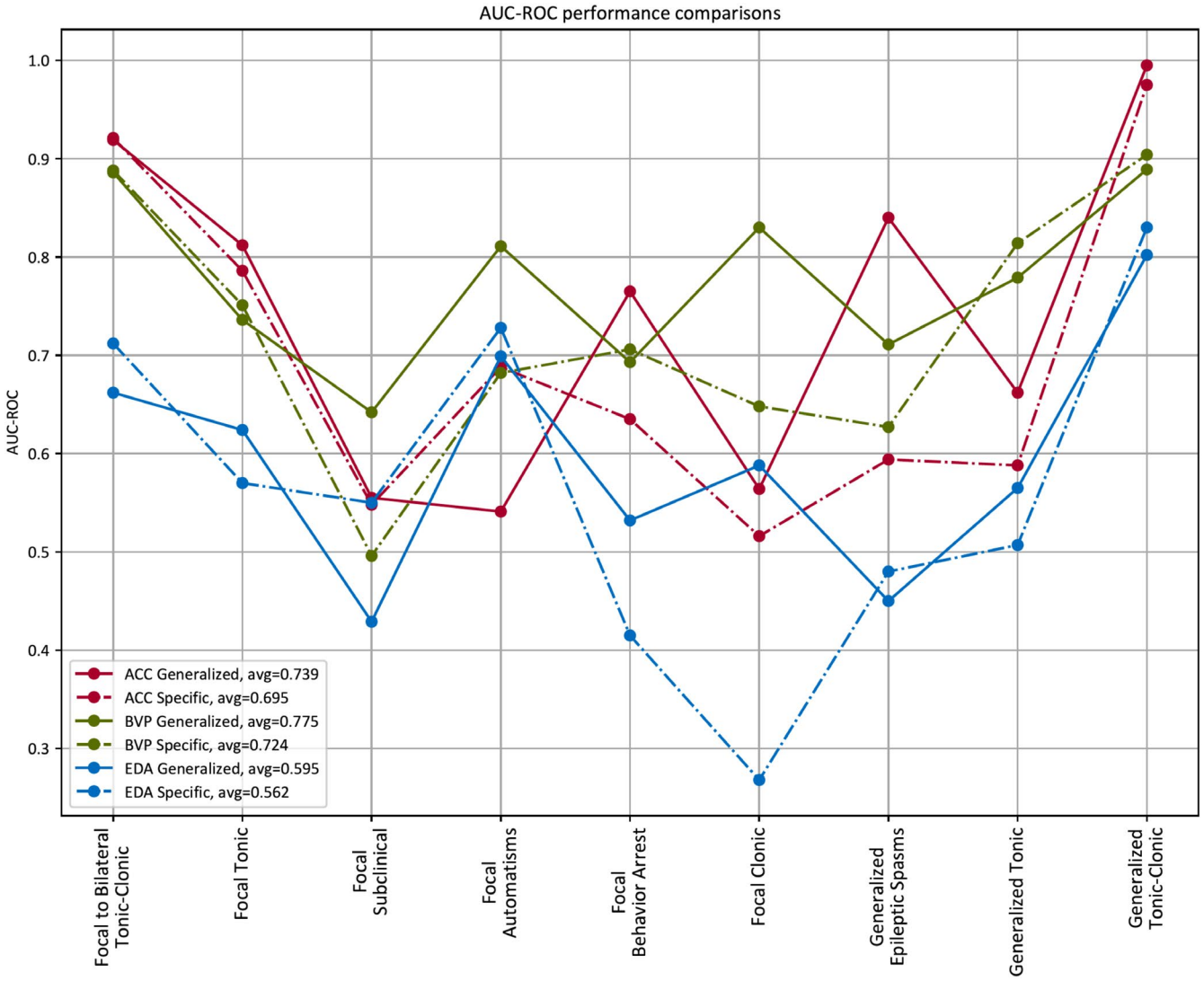

FIGURE 3 Area under the receiver operating characteristic curve (AUC-ROC) performance comparisons between a generalized type-agnostic machine learning (ML) model for all seizure types and a type-specific ML model for individual seizure types. An AUC-ROC less than .6 is not significant compared to random guess. For AUC-ROC levels greater than .6, accelerometry (ACC) performed similarly for focal to bilateral tonic–clonic seizures (FBTCSs), focal tonic seizures, generalized tonic seizures, and generalized tonic–clonic seizures (GTCSs); blood volume pulse (BVP) performed similarly for FBTCSs, focal tonic seizures, focal behavior arrest, generalized tonic seizures, and GTCSs; electrodermal activity (EDA) performed similarly for FBTCSs, focal automatisms, and GTCSs. For all three sensors, the average performance of the generalized ML model for all seizure types is better than the performance of the specific ML model for individual seizure types

data of all seizure types, but the results show the contrary. The main reasons for this are that (1) we have lower patient and seizure numbers per seizure type for each type compared to the lumped dataset; these lower sample numbers may not be enough to train a high-performance ML model; and (2) different sensor modalities may perform similarly across different seizure types.

We found better than chance detection potential for all nine studied seizure types, including focal and generalized seizures. ACC and BVP performed better than chance, with an AUC-ROC of .72 and .744, respectively; ACC and BVP data fusion achieved the best AUC-ROC of .752 when applied to the entire dataset consisting of all seizure samples lumped together from all nine seizure types. All nine clinical seizure types performed better than chance, with an AUC-ROC ranging from .648 to .995. ACC sensors achieve the best performance for most seizure types, similar to other studies that employ commercial smartwatches[37] using the standard ACC sensors and running the developed analytical models, showing utility in detecting seizures with motor components, especially GTCSs.[14,15] BVP performed best for some seizure types, namely focal motor seizures with automatisms and subclinical seizures. Given the sensitivity of BVP to motion, which may result in artifactual signals,[34,38]

the seizure detection performance of BVP may improve in seizures with a limited motor component. BVP may therefore assist in the detection of the peri-ictal period of tonic–clonic seizures and focal unaware seizures.[39,40] EDA with or without ACC has shown promise in the evaluation of convulsive seizures,[26,30,41] and may also be able to detect focal seizures.[41] In our study, EDA's overall AUC-ROC of .549 is not significantly better than chance, but EDA AUC-ROC values reached .802, .699, and .662 for GTCSs, focal automatisms, and FBTCSs. Although adding EDA to data fusions yielded an overall decreasing performance of all combinations (ACC + EDA, BVP + EDA, ACC + BVP + EDA), the detection performance for some seizure types increased when EDA was included. For example, in focal automatisms, the performance of ACC + EDA is better than ACC only (.772 vs. .541). Shifting EDA by 120 s showed significant improvement, especially for GTCSs and FBTCSs. Because the delay of EDA can vary for different seizures, applying an adaptive detection model may further improve its performance.

Whereas we used the same types of analytical models and data processing methods for all data modalities and fusion datasets in this study, future work will focus on developing advanced deep learning architectures and data preprocessing schemes individually for each data modality and integrating additional clinical features. We expect that this will allow us to leverage clinical and data modality-related features, thus further improving detection performance and cross-patient generalizability of detection models.

4.4 | Individualized seizure detection

Seizures are associated with altered autonomic nervous system activity, manifesting as system changes, including changes in heart rate, blood pressure, respiration, and sweating responses.[42] Monitoring of such physiologic signals may allow for tracking of seizure-related autonomic changes and seizure detection, and ultimately also provide information regarding sudden unexpected death in epilepsy (SUDEP) risk and seizure severity.[41] In particular, multimodal signal analysis has shown promise in this area.[43–47] Findings may also complement the detection of nonconvulsive seizures, which are harder to monitor and detect with non-EEG devices.[8]

Tracking seizures and evaluating the seizure burden remains challenging, as these tasks rely heavily on patient or caregiver reporting that may be incomplete.[3,6] Proper tracking of seizures, on an individual level, is crucial for disease management, improving outcomes, injury prevention, and potentially decreasing the risk of SUDEP. A user-friendly, portable, noninvasive, nonstigmatizing tool that reliably detects seizures can improve patients' quality of life and their health outcomes and may improve the evaluation of treatment outcomes based on seizure frequency.[11,45,47–49] Closed-loop seizure detection systems are aimed at seizure detection and prediction to provide early warnings to control or prevent seizures.[8,11] Modalities with slower time to detection, such as EDA in our study, may be more suitable for seizure diary purposes or clinical trials looking at seizure frequency and recurrence assessment with greater sensitivity and specificity. Our results suggest that wrist-worn and commercially available sensors, running advanced deep learning models and data preprocessing techniques, could be a feasible out-of-the-box starting alternative to custom-developed monitoring devices. Our results also suggest that individualized customization of detection modalities based on clinical features, including seizure semiology, may improve detection performance in selected patients.

4.5 | Challenges

We need to interpret our results in the setting of data acquisition. Specifically, we included pediatric patients admitted for in-hospital long-term EEG monitoring; activities and behaviors may differ from those at home and manifest differently in wearable device signals. Although the patients were admitted to the hospital, they were free to move around and perform certain activities such as playing video games, drawing, pedaling, and watching TV as feasible in this setting. Longer recordings over several days may help account for daily patterns and variations of seizure characteristics. We collected signals from one or two devices, depending on availability, placed on the left or right wrist or ankle. This may interfere with the consistency of the signals, as different parts and sides of the body may generate signals that differ in quality and characteristics. We included various seizure types in our analysis, yet there are additional seizure types that we have not investigated.

During the data collection phase, EEG labels were manually inputted into the annotation system, and wearable signal data were manually downloaded and saved. Performed manually, these processes are prone to human error. Furthermore, signal segments may be compromised by switched-off wearable devices and battery failures. To account for these shortcomings, we ran a stepwise quality check process (Figure S2).

EEG monitors and wearable devices run on independent clocks. Hence, we synchronized EEG and wearable device times at the start of each recording and accounted for possible time drift between the clocks.

4.6 | Conclusions

Automatic epileptic seizure detection of a broad variety of seizure types using ML and wearable data is feasible.

Preliminary results show better than chance seizure detection across a range of nine seizure types. Future improvements may consider clinical chronoepileptological variables, such as seizure duration and etiology or syndrome, as well as alternative data balancing, pre- and postprocessing, fusion, and ensemble learning methods. Thus, although our findings suggest feasibility, future adjustments may further improve detection performance.

ACKNOWLEDGMENTS

This study was funded in part by the Epilepsy Research Fund. S.V. was supported by the German Research Foundation under the grant VI 1088/1-1. We thank Rahul Herur and Sarah Salzman for clinical data collection and assistance with demographics tables, S. M. Shariar for assistance with the semiology table, Dr. Xiaofan Wang for data cleaning, Dr. Bo Zhang for his expert advice on statistical methods, and Drs. Stefan Maetschke and Fatemeh Jalali for their advice in developing ML methods.

CONFLICT OF INTEREST

T.L. serves on the Council of the American Clinical Neurophysiology Society, on the American Board of Clinical Neurophysiology, as founder and consortium principal investigator of pSERG (Pediatric Status Epilepticus Research Group), as an associate editor for Wyllie's Treatment of Epilepsy 6th and 7th editions, and as a member of the NORSE Institute, PACS1 Foundation, and CCEMRC. He served as associate editor of *Seizure* and served on the Laboratory Accreditation Board for Long Term (Epilepsy and Intensive Care Unit) Monitoring in the past. He is part of patent applications to detect and predict clinical outcomes and to manage, diagnose, and treat neurological conditions, epilepsy, and seizures. T.L. is coinventor of the TriVox Health technology, and T.L. and Boston Children's Hospital might receive financial benefits from this technology in the form of compensation in the future. He has received research support from the Epilepsy Research Fund, the National Institutes of Health, the Epilepsy Foundation of America, the Epilepsy Therapy Project, and the Pediatric Epilepsy Research Foundation, and has received research grants from Lundbeck, Eisai, Upsher-Smith, Mallinckrodt, Sunovion, Sage, Empatica, and Pfizer, including past device donations from various companies, including Empatica, SmartWatch, and Neuro-electrics. In the past, he served as a consultant for Zogenix, Upsher-Smith, Amzell, Engage, Elsevier, UCB, Grand Rounds, Advance Medical, and Sunovion. He performs video-electroencephalographic long-term and intensive care unit (ICU) monitoring, electroencephalograms, and other electrophysiological studies at Boston Children's Hospital and affiliated hospitals and bills for these procedures, and he evaluates pediatric neurology patients and bills for clinical care. He has received speaker honorariums from national societies, including the AAN, AES, and ACNS, and grand rounds at various academic centers. His wife, Dr. Karen Stannard, is a pediatric neurologist. She performs video-electroencephalographic long-term and ICU monitoring, electroencephalograms, and other electrophysiological studies and bills for these procedures, and she evaluates pediatric neurology patients and bills for clinical care. C.M. is part of patent applications covering technology for detecting and predicting clinical outcomes and for managing, diagnosing, and treating neurological conditions. S.H., J.T., U.A., and S.R. are part of patent applications describing technology to detect and classify epileptic seizures using EEG data and video data. S.V. is part of a patent application covering technology for seizure forecasting. In support of two studies investigating the use of artificial intelligence technology for epileptic seizure detection S.H. serves as a scientific consultant for Massachusetts General Hospital and as a member of the scientific advisory board of NeuroPace. None of the other authors have any conflict of interest to disclose. 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.

ORCID

Jianbin Tang https://orcid.org/0000-0001-5440-0796
Rima El Atrache https://orcid.org/0000-0002-5600-1120
Shuang Yu https://orcid.org/0000-0001-7428-2600
Umar Asif https://orcid.org/0000-0001-5209-7084
Michele Jackson https://orcid.org/0000-0001-9869-7333
Subhrajit Roy https://orcid.org/0000-0002-6072-5500
Mahtab Mirmomeni https://orcid.org/0000-0002-0192-995X
Sarah Cantley https://orcid.org/0000-0001-9786-0960
Theodore Sheehan https://orcid.org/0000-0002-8059-9388
Sarah Schubach https://orcid.org/0000-0002-6329-5761
Solveig Vieluf https://orcid.org/0000-0002-5532-8690
Christian Meisel https://orcid.org/0000-0003-2984-5480
Stefan Harrer https://orcid.org/0000-0001-7947-330X
Tobias Loddenkemper https://orcid.org/0000-0003-2074-0674

REFERENCES

1. Beghi E. The epidemiology of epilepsy. Neuroepidemiology. 2020;54(2):185–91.
2. Sander JW. The epidemiology of epilepsy revisited. Curr Opin Neurol. 2003;16(2):165–70.
3. Elger CE, Hoppe C. Diagnostic challenges in epilepsy: seizure under-reporting and seizure detection. Lancet Neurol. 2018;17(3):279–88.
4. Hoppe C, Poepel A, Elger CE. Epilepsy: accuracy of patient seizure counts. Arch Neurol. 2007;64(11):1595–9.
5. Cook MJ, O'Brien TJ, Berkovic SF, Murphy M, Morokoff A, Fabinyi G, et al. Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study. Lancet Neurol. 2013;12(6):563–71.

6. Blachut B, Hoppe C, Surges R, Elger C, Helmstaedter C. Subjective seizure counts by epilepsy clinical drug trial participants are not reliable. Epilepsy Behav. 2017;67:122–7.
7. Shih JJ, Fountain NB, Herman ST, Bagic A, Lado F, Arnold S, et al. Indications and methodology for video-electroencephalographic studies in the epilepsy monitoring unit. Epilepsia. 2018;59(1):27–36.
8. Ulate-Campos A, Coughlin F, Gaínza-Lein M, Fernández IS, Pearl Pl, Loddenkemper T. Automated seizure detection systems and their effectiveness for each type of seizure. Seizure. 2016;40:88–101.
9. Verdru J, Van Paesschen W. Wearable seizure detection devices in refractory epilepsy. Acta Neurol Belg. 2020;120(6):1271–81.
10. Leijten FSS. Multimodal seizure detection: a review. Epilepsia. 2018;59(Suppl 1):42–7.
11. Ramgopal S, Thome-Souza S, Jackson M, Kadish NE, Sánchez Fernández I, Klehm J, et al. Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy. Epilepsy Behav. 2014;37:291–307.
12. Garbarino M, Lai M, Bender D, Picard RW, Tognetti S. Empatica E3—A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. In 2014 4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH). IEEE. 2014; p. 39–42.
13. Poh MZ, Swenson NC, Picard RW. A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. IEEE Trans Biomed Eng. 2010;57(5):1243–52.
14. 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):e58–61.
15. Lockman J, Fisher RS, Olson DM. Detection of seizure-like movements using a wrist accelerometer. Epilepsy Behav. 2011;20(4):638–41.
16. Rajan S, Leonard N, Fletcher R, Casarjian B, Casarjian R, Cisse C, et al. Ambulatory autonomic activity monitoring among at-risk adolescent mothers. J Mobile Technol Med. 2012;1(3):25–31.
17. 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):e428–34.
18. Nijsen TM, Cluitmans PJ, Arends JB, Griep PAM. Detection of subtle nocturnal motor activity from 3-D accelerometry recordings in epilepsy patients. IEEE Trans Biomed Eng. 2007;54(11):2073–81.
19. Narechania AP, Garic II, Sen-Gupta I, Macken MP, Gerard EE, Schuele SU. Assessment of a quasi-piezoelectric mattress monitor as a detection system for generalized convulsions. Epilepsy Behav. 2013;28(2):172–6.
20. Bruno E, Simblett S, Lang A, Biondi A, Odoi C, Schulze-Bonhage A, et al. Wearable technology in epilepsy: the views of patients, caregivers, and healthcare professionals. Epilepsy Behav. 2018;85:141–9.
21. Ulate-Campos A, Coughlin F, Gainza-Lein M, Fernández IS, Pearl Pl, Loddenkemper T. Automated seizure detection systems and their effectiveness for each type of seizure. Seizure. 2016;40:88–101.
22. Simblett SK, Biondi A, Bruno E, Ballard D, Stoneman A, Lees S, et al. Patients' experience of wearing multimodal sensor devices intended to detect epileptic seizures: a qualitative analysis. Epilepsy Behav. 2020;102:e106717.
23. Beniczky S, Karoly P, Nurse E, Ryvlin P, Cook M. Machine learning and wearable devices of the future. Epilepsia. 2021;62(S2):e16555.
24. Saab K, Dunnmon J, Ré C, Rubin D, Lee-Messer C. Weak supervision as an efficient approach for automated seizure detection in electroencephalography. NPJ Digital Med. 2020;3(1):59.
25. Asif U, Roy S, Tang J, Harrer S. SeizureNet: Multi-spectral deep feature learning for seizure type classification. In: Machine learning in clinical neuroimaging and radiogenomics in neuro-oncology. Cham: Springer; 2020; p. 77-87.
26. 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.
27. Fisher RS, Cross JH, D'Souza C, French JA, Haut SR, Higurashi N, et al. Instruction manual for the ILAE 2017 operational classification of seizure types. Epilepsia. 2017;58(4):531–42.
28. Acharya JN, Hani A, Cheek J, Thirumala P, Tsuchida TN. American Clinical Neurophysiology Society Guideline 2: guidelines for standard electrode position nomenclature. J Clin Neurophysiol. 2016;33(4):308–11.
29. Lin M, Chen Q, Yan S. Network in network. arXiv preprint. 2013; arXiv:1312.4400.
30. Poh MZ, 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):e93–7.
31. Heldberg BE, Kautz T, Leutheuser H, Hopfengartner R, Kasper BS, Eskofier BM. Using wearable sensors for semiology-independent seizure detection—towards ambulatory monitoring of epilepsy. Conf Proc IEEE Eng Med Biol Soc. 2015;2015:5593–6.
32. Kiral-Kornek I, Roy S, Nurse E, Mashford B, Karoly P, Carroll T, et al. Epileptic seizure prediction using big data and deep learning: toward a mobile system. EBioMedicine. 2018;27:103–11.
33. Loeckx D, Buckinx T. In-field validation of epihunter for the detection of absence seizures using wearable EEG. Paper presented at 2nd International Congress on Mobile Devices and Seizure Detection in Epilepsy. Lausanne, Switzerland; September 6–7, 2019.
34. Vandecasteele K, De Cooman T, Gu Y, Cleeren E, Claes K, Paesschen W, et al. Automated epileptic seizure detection based on wearable ECG and PPG in a hospital environment. Sensors (Basel). 2017;17(10):2338.
35. Voelker R. Smart watch detects seizures. JAMA. 2018;319(11):1086.
36. Drug and device news. P T. 2019;44(3):95–107.
37. Harrer S, Shah P, Antony B, Hu J. Artificial intelligence for clinical trial design. Trends Pharmacol Sci. 2019;40(8):577–91.
38. Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas. 2007;28(3):R1–39.
39. El Atrache R, Tamilia E, Mohammadpour Touserkani F, Hammond S, Papadelis C, Kapur K, et al. Photoplethysmography: a measure for the function of the autonomic nervous system in focal impaired awareness seizures. Epilepsia. 2020;61(8):1617–26.
40. Mohammadpour Touserkani F, Tamilia E, Coughlin F, Hammond S, El Atrache R, Jackson M, et al. Photoplethysmographic evaluation of generalized tonic-clonic seizures. Epilepsia. 2020;61(8):1606–16.
41. Poh MZ, Loddenkemper T, Reinsberger C, Swenson NC, Goyal S, Madsen JR, et al. Autonomic changes with seizures correlate with postictal EEG suppression. Neurology. 2012;78(23):1868–76.

42. Devinsky O. Effects of seizures on autonomic and cardiovascular function. Epilepsy Curr. 2004;4(2):43–6.
43. Vieluf S, El Atrache R, Hammond S, Touserkani FM, Loddenkemper T, Reinsberger C. Peripheral multimodal monitoring of ANS changes related to epilepsy. Epilepsy Behav. 2019;96:69–79.
44. Vieluf S, Reinsberger C, El Atrache R, Jackson M, Schubach S, Ufongene C, et al. Autonomic nervous system changes detected with peripheral sensors in the setting of epileptic seizures. Sci Rep. 2020;10(1):11560.
45. Van de Vel A, Cuppens K, Bonroy B, Milosevic M, Jansen K, Van Huffel S, et al. Non-EEG seizure detection systems and potential SUDEP prevention: state of the art: review and update. Seizure. 2016;41:141–53.
46. van Westrhenen A, De Cooman T, Lazeron RHC, Van Huffel S, Thijs RD. Ictal autonomic changes as a tool for seizure detection: a systematic review. Clin Auton Res. 2019;29(2):161–81.
47. Amengual-Gual M, Ulate-Campos A, Loddenkemper T. Status epilepticus prevention, ambulatory monitoring, early seizure detection and prediction in at-risk patients. Seizure. 2019;68:31–7.
48. Thompson ME, Langer J, Kinfe M. Seizure detection watch improves quality of life for adolescents and their families. Epilepsy Behav. 2019;98(Pt A):188–94.
49. Gutierrez EG, Crone NE, Kang JY, Carmenate YI, Krauss GL. Strategies for non-EEG seizure detection and timing for alerting and interventions with tonic-clonic seizures. Epilepsia. 2018;59(Suppl 1):36–41.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: 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;62:1807–1819. https://doi.org/10.1111/epi.16967