Received: 21 February 2025 | Revised: 26 March 2025 | Accepted: 26 March 2025

DOI: 10.1111/epi.18406

Epilepsia®

RESEARCH ARTICLE

Deep learning–based detection of generalized convulsive seizures using a wrist-worn accelerometer

Antoine Spahr[1] | **Adriano Bernini**[1] | **Pauline Ducouret**[1] | **Christoph Baumgartner**[2,3] | **Johannes P. Koren**[2,3] | **Lukas Imbach**[4] | **Sàndor Beniczky**[5,6] | **Sidsel A. Larsen**[5,6] | **Sylvain Rheims**[7,8] | **Martin Fabricius**[9] | **Margitta Seeck**[10] | **Berhard J. Steinhoff**[11,12] | **Isabelle Beuchat**[1] | **Jonathan Dan**[13] | **David A. Atienza**[13] | **Charles-Edouard Bardyn**[1] | **Philippe Ryvlin**[1]

[1]NeuroDigital@NeuroTech, Department of Clinical Neurosciences, Lausanne University Hospital (CHUV), University of Lausanne, Lausanne, Switzerland

[2]Department of Neurology, Clinic Hietzing, Vienna, Austria

[3]Karl Landsteiner Institute for Clinical Epilepsy Research and Cognitive Neurology, Medical Faculty Sigmund Freud University, Vienna, Austria

[4]Swiss Epilepsy Center, Klinik Lengg, Zurich, Switzerland

[5]Department of Clinical Neurophysiology, Aarhus University Hospital, Aarhus, Denmark

[6]Danish Epilepsy Centre, Dianalund, Denmark

[7]Department of Functional Neurology and Epileptology, Hospices Civils de Lyon, Lyon 1 University, Lyon, France

[8]Lyon Neuroscience Research Center, Institut National de la Santé et de la Recherche Médicale U1028/CNRS UMR 5292 Epilepsy Institute, Lyon, France

[9]Department of Clinical Neurophysiology, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark

[10]EEG and Epilepsy Unit, University Hospitals and Faculty of Medicine, University of Geneva, Geneva, Switzerland

[11]Epilepsiezentrum Kork, Kehl-Kork, Germany

[12]Clinic of Neurology and Clinical Neurophysiology, Albert-Ludwigs University of Freiburg, Freiburg, Germany

[13]Embedded Systems Laboratory, EPFL, Lausanne, Switzerland

Correspondence
Antoine Spahr and Philippe Ryvlin, NeuroDigital@NeuroTech, Department of Clinical Neurosciences, Lausanne University Hospita (CHUV), University of Lausanne, Lausanne, Switzerland.
Email: antoine.spahr@chuv.ch and philipperyvlin@gmail.com

Abstract

Objective: To develop and validate a wrist-worn accelerometer-based, deep-learning tunable algorithm for the automated detection of generalized or bilateral convulsive seizures (CSs) to be integrated with off-the-shelf smartwatches.

Methods: We conducted a prospective multi-center study across eight European epilepsy monitoring units, collecting data from 384 patients undergoing video electroencephalography (vEEG) monitoring with a wrist-worn three dimensional (3D)–accelerometer sensor. We developed an ensemble-based convolutional

Antoine Spahr and Adriano Bernini contributed equally to this work.

Charles-Edouard Bardyn and Philippe Ryvlin contributed equally to this work.

Sàndor Beniczky, Sylvain Rheims and Philippe Ryvlin are members of the European Reference Network EPICARE.

Funding information
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: 10.002.812, 180365, 320030_179240 and CRSII5_193813

neural network architecture with tunable sensitivity through quantile-based aggregation. The model, referred to as Episave, used accelerometer amplitude as input. It was trained on data from 37 patients who had 54 CSs and evaluated on an independent dataset comprising 347 patients, including 33 who had 49 CSs.

Results: Cross-validation on the training set showed that optimal performance was obtained with an aggregation quantile of 60, with a 98% sensitivity, and a false alarm rate (FAR) of 1/6 days. Using this quantile on the independent test set, the model achieved a 96% sensitivity (95% confidence interval [CI]: 90%–100%), a FAR of <1/8 days (95% CI: 1/9–1/7 days) with 1 FA/61 nights, and a median detection latency of 26 s. One of the two missed CSs could be explained by the patient's arm, which was wearing the sensor, being trapped in the bed rail. Other quantiles provided up to 100% sensitivity at the cost of a greater FAR (1/2 days) or very low FAR (1/100 days) at the cost of lower sensitivity (86%).

Significance: This Phase 2 clinical validation study suggests that deep learning techniques applied to single-sensor accelerometer data can achieve high CS detection performance while enabling tunable sensitivity.

KEYWORDS
deep learning, epilepsy, focal-to-bilateral tonic–clonic seizures, generalized tonic–clonic seizure, seizure detection, wearable

1 | INTRODUCTION

Focal to bilateral tonic–clonic seizures and generalized tonic–clonic seizures (collectively referred to hereafter as convulsive seizures [CSs]) represent some of the most severe forms of epileptic seizures, affecting ~25%–30% of the estimated 50 million people with epilepsy (PWE) worldwide.[1–3] These seizures are associated with increased risks of injury[4–6] and sudden unexpected death in epilepsy (SUDEP).[7,8] Automated detection of such seizures enables timely and potentially life-saving intervention,[9,10] as well as accurate documentation of seizure occurrence,[11–13] particularly valuable for PWE with unreliable self-reporting or sleep-related seizures.[14–16]

Motion analysis has emerged as a promising approach for automated detection of CSs,[17] with several studies demonstrating the feasibility of such detection using only wrist-worn accelerometers.[18–21] Other CS-detecting devices are based on changes in electromyography (EMG),[22–24] the combination of accelerometer and heart rate,[25,26] or accelerometer and electrodermal activity (EDA).[27–29]

In this study, we investigated a novel deep-learning approach to CS detection using solely accelerometer data from a wrist-worn device, with the view to producing an algorithm that offers adaptive sensitivity settings and can be integrated with off-the-shelf smartwatches. To this purpose, we developed a method that combines a deep convolutional neural network (CNN) architecture with an ensemble-based approach and a tunable parameter. We evaluate this approach, referred to as Episave, through a large-scale, multi-center study involving eight European epilepsy monitoring units (EMUs).

Key points

- A novel deep learning approach was developed to detect convulsive seizures (CSs) using wrist accelerometer amplitude data.
- The fixed-and-frozen model referred to as Episave, achieved 96% sensitivity with one false alarm every 8 days on a fully independent test set that included 347 patients and 49 CSs.
- The ensemble-based architecture of the model with tunable sensitivity enables adaptive detection thresholds without requiring model retraining, with the potential to increase sensitivity to 100% or decrease false alarms to one per 100 days.

2 | MATERIALS AND METHODS

2.1 | Study design and population

This is a prospective observational multi-center study across eight European epilepsy centers from Austria (Vienna), Denmark (Copenhagen, Dianalund), France

(Lyon), Germany (Kork), and Switzerland (Geneva, Lausanne, Zurich).

The inclusion criteria were PWE who were 12 years and older and had a documented history of CSs. All participants underwent continuous video-electroencephalography (vEEG) monitoring during an EMU stay while wearing a multisensor wristband after having consented. The study protocol received approval from the local ethics committees of all participating centers (lead center reference: CER-VD 2019-00437).

2.2 | Data collection

The Empatica E4 wristband (Empatica Inc., Boston, MA, USA) was used for data collection. This research-grade device incorporates multiple sensors for comprehensive physiological monitoring, including a three-axis accelerometer (3D-acc) operating at 32 Hz with a ±2 g range, an EDA sensor, a blood volume pulse sensor, and a temperature sensor. For this study, only 3D-acc data were used for the development of the CS detection algorithm.

All recorded CS events were initially reported by the technicians and epileptologists of each EMU. Subsequently, these events underwent an independent review and validation process conducted by a senior epileptologist (P.R.), adhering to the International League Against Epilepsy (ILAE) seizure classification.[30] CS onset was defined clinically based on video recording.

2.3 | Wearable data preprocessing

The E4 clock proved to suffer from a previously unreported linear time drift, which manifested in two distinct patterns: recordings before December 2021 exhibited a drift rate of 0.0955 ms/s, whereas those after December 2021 showed a drift of ~1.5 ms/s. This finding was used to realign vEEG labels with E4 data and was complemented by a visual inspection of the last clonic movements of each recording CS on the video, EEG, EMG, and 3D-acc data. Indeed, the last clonic movements can be precisely delineated and distinguished from the postictal phase on each of these signals, enabling a robust final realignment.

For the model development with the 3D-acc data, we converted all measurements to G units (where 1 G = 9.81 m/s^2). To simplify the input features while maintaining signal integrity, we transformed the 3D-acc signals into a single amplitude time series by computing their Euclidean norm. No further processing, like filtering, was performed on the 3D-acc data.

2.4 | Dataset partitioning

To ensure robust model evaluation and prevent data leakage, we implemented a patient-level partitioning strategy. The complete dataset was divided into two distinct sets irrespective of the CS characteristics: the training set, and the independent test set, with strict separation ensuring no patient overlap between the sets.

2.5 | Model overview and architecture

Our detection system employs an ensemble-based approach utilizing multiple neural networks trained on different subsets of the data. Each neural network performs binary classification of time-series windows (CS vs no CS), and the final prediction is obtained by aggregating all predictions using the Harrell-Davis quantile function,[31] where the quantile value is a tunable parameter depending on the desired sensitivity/false alarm rate (FAR) trade-off.

An overview of the complete model architecture and training pipeline is provided in Figure 1. Each model of our ensemble framework takes the 3D-acc amplitude as input and consists of a 1D CNN comprising 14 convolutional layers. The filter depths progressively increase through the network ($1 \rightarrow 16 \rightarrow 32(\times 11) \rightarrow 64 \rightarrow 64$). Each convolutional layer is configured with a kernel size of 5, and the network uses varying strides of [1, 2] with no zero padding and a dilation of 1. Following each convolutional layer, we applied rectifier linear unit (ReLU) activation and batch normalization. The convolutional output was processed through a global average pooling layer to produce a 64-dimensional representation. This representation was then fed into a feed-forward network with layer dimensions $64 \rightarrow 64 \rightarrow 32 \rightarrow 16 \rightarrow 2$. The feed-forward network incorporates ReLU activation functions and batch normalization at each layer, with a dropout layer ($p = .025$) for regularization.

In the ensemble framework, we used the Harrell–Davis quantile function to obtain the final prediction score.[31] A quantile value of 0.7, for example, requires at least 30% of the models to predict a CS event for the window to be classified as CS. The system declares a CS detection when the final prediction score exceeds 0.5.

2.6 | Ensemble model training pipeline

2.6.1 | Cross-validation scheme

Our training approach involved training 30 models through a cross-validation scheme using different random

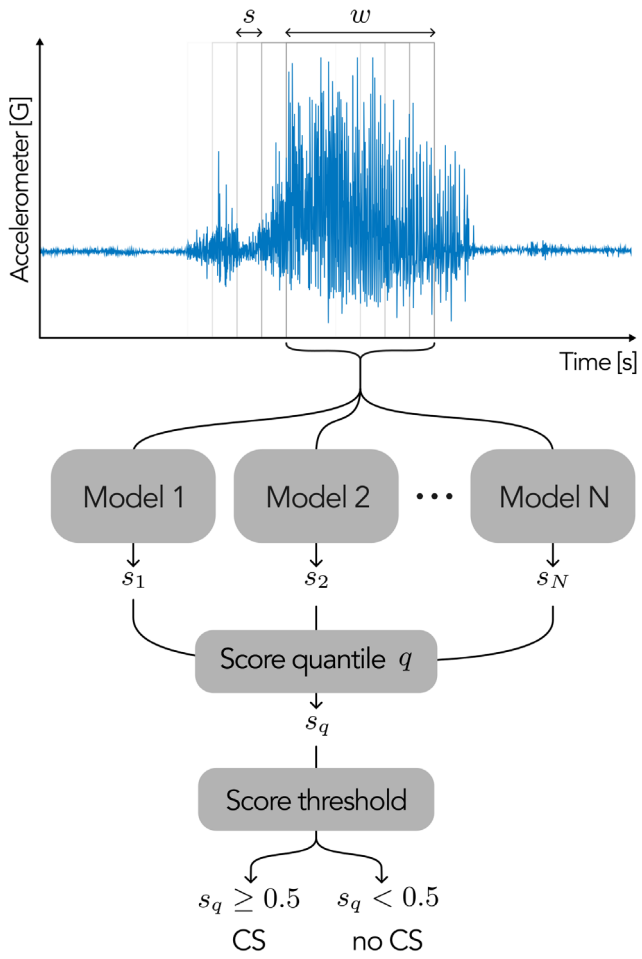

FIGURE 1 Scheme of the ensemble model. The accelerometer amplitude time–series is fed to the model by windows of width $w = 30$ s with a stride of $s = 5$ s. Each window is passed in the N models that give N predictions scores (s_1, ..., s_N). The final decision score, s_q is obtained with the q^{th} Harrell–Davis quantile, where q is a personalizable parameter. A CS (convulsive seizure) is detected when the final score is larger than a threshold of 0.5.

seeds and data partitioning. Within the training set, for each fold (1 per model) the data were split at the patient level into training and validation sets with a fixed number of CSs per fold. To sample the validation set, we employed a weighted sampling strategy that prioritizes patients who were underrepresented in previous validation sets, ensuring that all the patients are used evenly in both sets.

This extensive training served two purposes. First, it enabled us to estimate confidence intervals (CIs) on model performance by randomly sampling 10 models from the pool of 30 and computing the metrics for various aggregation quantiles. This sampling was repeated 50 times to generate a sample of 50 metric values for each quantile, from which the minimum and maximum values yield the CIs. Second, it allowed us to perform a selection of the 10 best models on validation performance to use as our final ensemble. The best models selected were the ones displaying the lowest Euclidean distance between their validation performances and the optimal performance point (sensitivity = 1, FAR = 0).

2.6.2 | Single model training

For the training of a single model of the ensemble, we employed a binary cross-entropy loss function and focused only on windows that either fully overlapped with CS events or windows that did not overlap with the CS events. The class imbalance was addressed by oversampling the windows with CS events to achieve balanced batches by randomly sampling with replacement as many windows with and without CS. However, for validation and inference phases, we used the entire dataset.

We used the AdamW optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and no weight decay for optimization.[32] We implemented a learning rate schedule beginning at 1e−5 and following a three-phase progression: a warm-up phase with a linear increase to 1e−3 over 2500 steps, followed by cosine annealing to 3e−5 over 45 000 steps, and concluding with a linear cool-down to 0 over 2500 steps. The total training ran for 50 000 steps, without early stopping, with a batch size of 512.

Throughout training, we evaluated the model performance every 5000 steps and saved the model's weights as the current best under two conditions: either it showed improvement in both sensitivity and FAR, or it achieved a reduction in FAR >10% while maintaining sensitivity within 5% of the previous best model. This allowed us to keep a good balance between sensitivity and FAR.

2.7 | Performance evaluation

Each recording was processed by the ensemble model with 30 s windows and a stride of 5 s (0.2 Hz), which we chose empirically as a sweet spot between computational efficiency and signal coverage. Scores from different aggregation quantiles were recovered (0, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100) to evaluate the model's performances at different sensitivity/FAR trade-offs. The quantile found to provide the optimal performance on the cross-validation training set was used to assess the primary model performance on the held out independent test set.

Although the models yield a prediction at the window level, their performance was evaluated through clinically-relevant metrics at the event level, including true positive (TP; CS effectively detected by the models, Figure 2A), false negative (FN; CS not detected by the models, Figure 2B), false positive (FP; non-ictal event falsely interpreted by the

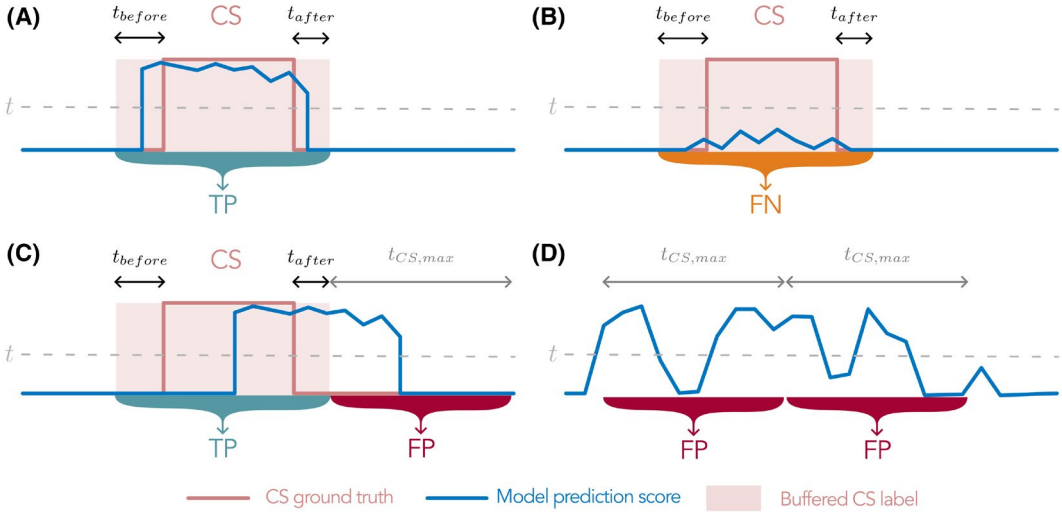

FIGURE 2 Event-based metric computation. Illustration of different situations in our event-based metric computation approach. (A) Model's prediction within the convulsive seizure (CS) bounds (with buffer t_{before} and t_{after}) giving a true positive (TP). (B) Model's prediction within the buffered CS bounds giving a false negative (FN). (C) Model's prediction overlapping with the CS-buffered bounds but overflowing those bounds. The CS is detected (TP) but it also yields a false positive (FP). Note that this specific case never occurred in our analysis. (D) Model's predictions outside of any CS bounds yielding FP. In the scenario of a long period of false alarms, this period is divided by the maximum expected CS time ($t_{CS,max}=150$ s) to yield multiple FPs. For instance, a model triggering a false alarm over 10 min will result in 4 FPs (10 min/150 s).

models as a CS, Figure 2C,D), sensitivity (TP/(TP + FN)), and FAR (number of FPs per day).

To compute these event-level metrics, we first augmented the CS event with a $t_{before}=120$ s buffer pre-event to cover detection that could occur in advance and reflect a focal motor seizure preceding the CS, and a $t_{after}=10$ s buffer post-event to account for potential temporal uncertainty and residual misalignment. The prediction scores were binarized using a threshold of $t=0.5$. If any prediction overpasses the threshold within the augmented CS bounds, we consider the CSs as detected. We defined another parameter $t_{CS,max}=150$ s, which represents the maximum event duration (as most CSs last <120 s, we allocated extra time capacity to safeguard against unexpected event extensions[33]). This parameter represents both the minimum duration between events and the maximum duration for a single FP (longer FPs are thus 'cut' into multiple events of 150 s). The system classified detection scenarios into four categories, as illustrated in Figure 2.

2.8 | Pilot smartwatch implementation

We implemented the algorithm on a commercially available Android WearOS smartwatch (TicWatch Pro 3, Mobvoi, Beijing, China) to perform a timing analysis of the computational requirements of the algorithm. The algorithm was ported to Kotlin. The CNN was ported to TFLite. The experiment was conducted over 30 min of 3D-acc recording.

3 | RESULTS

We enrolled 398 PWE between June 2019 and December 2024 (median age: 34 years, interquartile range 25–75: 24–42 years; 200 male, 198 female), 84 of whom experienced at least one CS during EMU monitoring for a total of 121 recorded CSs.

Usable E4 3D-acc data were available in 384 PWE, cumulating 36 401 h (~1517 days) of recording. One hundred three CSs were captured by the E4 in 70 of these patients. These included 55% type 1, 27% type 2, and 18% type 3 according to Alexandre et al. classification.[34] The mean ± standard deviation (SD) duration of CSs was 57 ± 17 s.

The 18 CSs not captured by the E4 resulted from the patient not wearing the device (one patient) at the time of the seizure or from technical issues hampering the recording or retrieval of the data (e.g., E4 battery down, data not found on Empatica cloud, 17 patients).

E4 data were split into a training and an independent test set. The training set included only PWE with at least one recorded CS and comprised 37 patients and 54 CSs across 4922 h (~206 days) of recording. The independent test set included 347 PWE across 31 479 h (~1312 days) of recording, 33 of whom had at least one recorded CS and a total of 49 CSs.

For each model training in our cross-validation scheme, the training set was further split at the patient level into a training and validation set, as described in the Methods, with 34 CSs allocated for training and 20 CSs for validation.

The performance of the model on the cross-validation training set varied with the quantile used, with optimal performance for quantile 60 that provided a sensitivity of 98% and FAR of 1/6 days. At the highest quantile (100), the sensitivity did not increase and the FAR reached its maximum (1/2 days). The lowest FAR (1/50 days) was reached at a quantile 10 at the cost of a lower sensitivity (80%). The performances of all other quantiles are provided in Table 1.

Using quantile 60, the performance of the model on the independent test set provided a sensitivity of 96% (95% CI: 90%–100%) and a FAR of 1/8 days (95% CI: 1/9–1/7 days). In one of the two undetected CSs, the patient's arm wearing the E4 was trapped in the bed rail, preventing any movement detectable by the accelerometer. In the other CS, the patient turned prone at the onset of the clonic phase, which limited the movement of the arm wearing the accelerometer. The distribution of FAR showed that only 63 of the 347 patients (18.2%) experienced FAs, whereas 284 (81.8%) of PWE had no FAs. Furthermore, 35 of the 63 patients (55%) with FAs had only one FA (see Figure 3B). Only 22 FAs occurred during nighttime (corresponding to 7 h: 23:00 to 06:00) over 9337 h of night recording, translating into a FAR of 1/61 nights. On the other hand, 147 FAs occurred during daytime (corresponding to 17 h: 06:00–23:00) over 22 142 h of day recording, corresponding to a FAR 1/9 days. Peaks in FAs were observed around 9:00, 14:00, and 19:00, possibly in relation to specific activities such as toothbrushing and meals (see Figure 3A).

Performance associated with other quantiles on the independent test set are provided in Table 1. Overall, the model's performance remained consistent across the different subsets of patients, with quantile 60 offering the best trade-off between sensitivity and FAR. Furthermore, there are no strong discrepancies between the cross-validation and the independent test set performances indicating an absence of overfitting (see Figure 4).

CS median detection latency, measured as the time difference between the end of the first detection window and the clinically marked CS onset was 26 s (min = −21 s, max = 42 s), when using quantile 60. For some focal motor seizures that evolved into a CS, detection occurred during the focal phase accounting for negative latencies. Detection always occurred before the end of the CS, with a median latency of −26 s (min = −105 s, max = −4 s) (Figure 5).

When operating on the smartwatch, a single CNN had an average inference time for one window (= 30 s) of 3D-acc data of 112 ms (min = 90 ms, max = 141 ms). For the 10-model ensemble, this inference time scales linearly to 1.12 s. These timings are well below the stride duration (5 s), ensuring real-time execution of the algorithm.

4 | DISCUSSION

This study found that our deep-learning based model, Episave, can achieve high sensitivity (96%) and very low FAR (<1/8 days) using only 3D-acc data captured at the wrist. This finding opens the door to the integration of this CS detection algorithm with low-cost off-the-shelf smartwatches, with the potential to expand accessibility to CS monitoring and detection, particularly for PWE from low- and middle-income countries. Furthermore, we introduced the possibility of adjusting one parameter (ensemble aggregation quantile) of the algorithm to adapt its sensitivity and FAR according to the patient's needs and CS characteristics, with the capacity to either raise the model's sensitivity to 100% or decrease its FAR to 1/100 days.

The ensemble-based CNN architecture used in our study represents a significant advancement in single-sensor CS detection. Although previous studies have utilized machine learning approaches,[29] our method introduces two key innovations. First, the ensemble approach provides inherent robustness against individual model variations. Second, the quantile-based aggregation enables adaptive sensitivity tuning without requiring model retraining, offering clinicians and patients flexibility in balancing detection sensitivity against FAs.

Our performance metrics compare well with those of other CS detection systems and are in line with the requirements usually accepted in the epilepsy community. Using the setting that proved optimal on the cross-validation set (quantile 60), we achieved a 96% sensitivity on the primary independent test set, whereas all other quantiles achieved sensitivities ≥86% and up to 100%. Current medically-certified CS detection devices have reported sensitivities in the same range: 85%–90% for the EpiCare free/mobile (Danish care technology, Sorø, Denmark),[19,35] 80%–100% for the Nightwatch (LivAssured B.V., Leiden, The Netherlands),[17,36,37] 95%–98% for the Embrace/E4 (Empatica Inc, Cambridge, MA, USA), and 76%–95% for SPEAC (Brain Sentinel Inc, San Antonio, TX, USA).[22,38] An Apple watch–based algorithm has also recently reported a 100% sensitivity.[25]

Our primary setting was associated with 1 FA/8 days, whereas the above-cited solutions reported FAR ranging from ~3/day to 1/50 days. We could further reduce FAR to 1 FA/100 days at the cost of a lower sensitivity (86% – quantile 0), whereas a 100% sensitivity resulted in a FAR of ~1 FA/2 days (quantile 90).

The median detection latency of our algorithm was 26 s after CS onset, which is also comparable to that of other CS detection devices,[19,20,28,29] except those using surface EMG, which achieve latencies of ~10 s.[22,38] Most importantly, our detection always occurred before

TABLE 1 Model's performances on the cross-validation training set and the independent test set according to its various quantiles.

	Cross-validation patient set					Independent test patient set				
	FP	FN	TP	Sensitivity (95% CI)	FAR/day (95% CI)	FP	FN	TP	Sensitivity (95% CI)	FAR/day (95% CI)
Quantile 0	5	14	40	.74 (.62–.86)	.02 (.00–.07)	19	7	42	.86 (.76–.96)	.01 (.00–.03)
Quantile 10	5	11	43	.80 (.69–.90)	.02 (.00–.07)	20	7	42	.86 (.76–.96)	.02 (.00–.03)
Quantile 20	10	5	49	.91 (.83–.99)	.05 (.00–.12)	44	6	43	.88 (.79–.97)	.03 (.01–.06)
Quantile 30	15	3	51	.94 (.88–1.00)	.07 (.00–.16)	63	5	44	.90 (.81–.98)	.05 (.02–.07)
Quantile 40	19	3	51	.94 (.88–1.00)	.09 (.00–.19)	90	3	46	.94 (.87–1.00)	.07 (.04–.10)
Quantile 50	27	2	52	.96 (.91–1.00)	.13 (.02–.24)	123	3	46	.94 (.87–1.00)	.09 (.06–.13)
Quantile 60	37	1	53	.98 (.95–1.00)	.18 (.05–.31)	169	2	47	.96 (.90–1.00)	.13 (.09–.17)
Quantile 70	43	1	53	.98 (.95–1.00)	.21 (.07–.35)	249	2	47	.96 (.90–1.00)	.19 (.14–.24)
Quantile 80	58	1	53	.98 (.95–1.00)	.28 (.12–.45)	366	1	48	.98 (.94–1.00)	.28 (.22–.34)
Quantile 90	107	1	53	.98 (.95–1.00)	.52 (.30–.74)	755	0	49	1.00 (1.00–1.00)	.58 (.48–.67)
Quantile 100	120	1	53	.98 (.95–1.00)	.59 (.35–.82)	854	0	49	1.00 (1.00–1.00)	.65 (.55–.75)
N patients	37					347				
Duration, h	4921.67					31478.91				
Duration, days	205.07					1311.62				

Note: The 95% confidence interval displayed for the sensitivity and the false alarm rate represents the clinical variability and is estimated with the Wald method using the normal distribution.

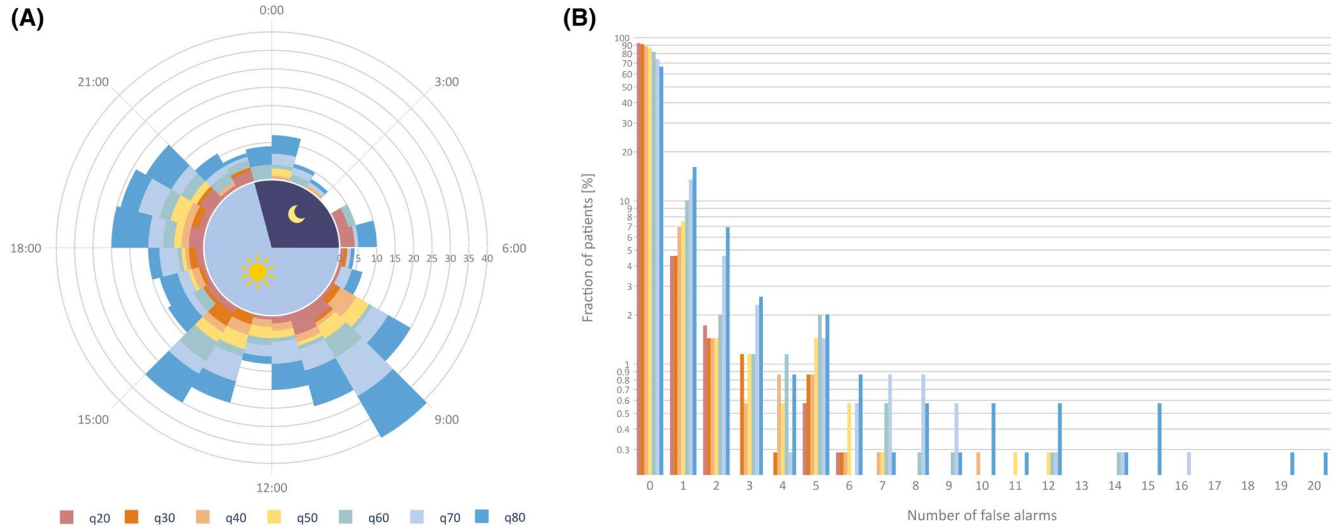

FIGURE 3 Distribution of false alarms for quantiles 20–80 (extreme quantiles were excluded for a better visualization). (A) Distribution around the clock: The number of false alarms on the full test set is shown by hour of the day. The histograms for the different quantiles are overlaid and not stacked. (B) Distribution per patient. Each bar represents the fraction of the test patients that had N false alarm (N being the x-axis).

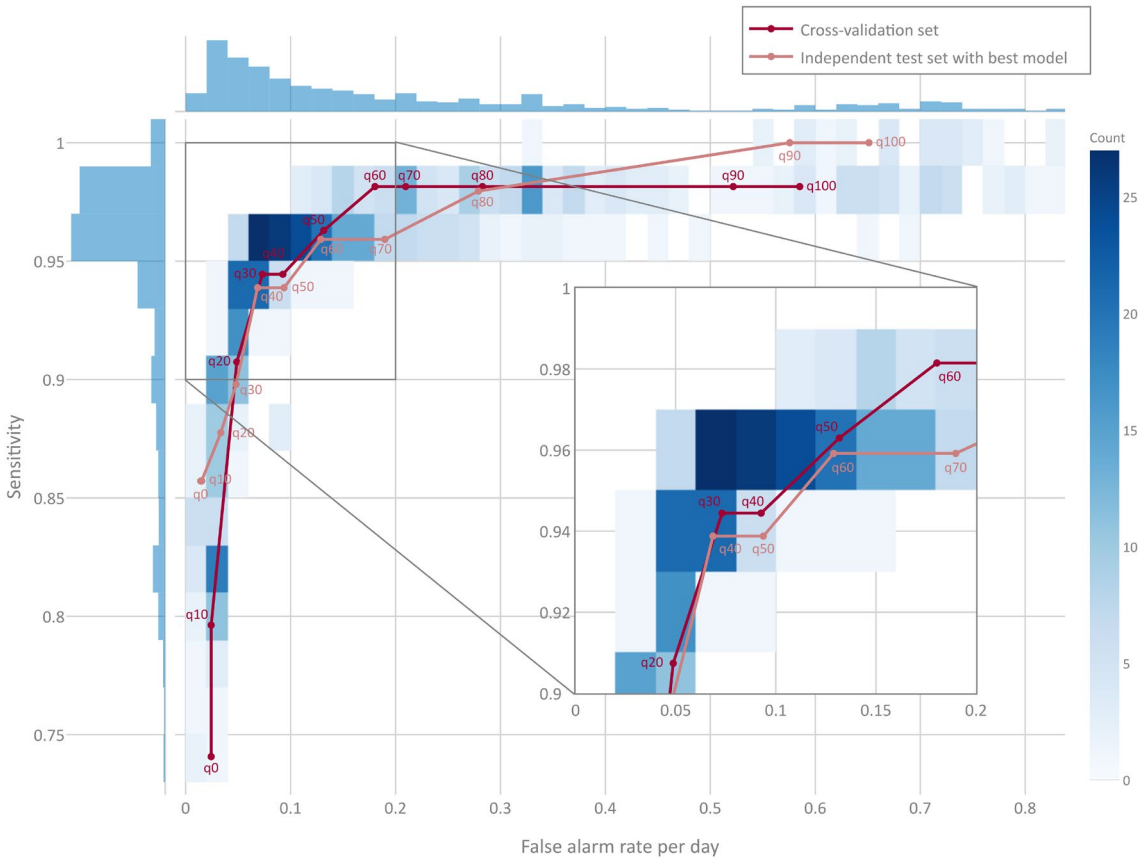

FIGURE 4 Model performances across the different quantiles on the independent test set. The center plot shows the distribution of pairs (false alarm rate, sensitivity) for all the quantiles over the $N = 50$ realizations. In addition, the individual 1D distribution for the sensitivity and false alarm rate are displayed on the top and left of each plot. The performance obtained on the cross-validation set and with the best set of 10 models on the independent test set is overlaid on the distribution and annotated with the quantile associated with each point. A zoom-in view of the middle range is given in the inset plot.

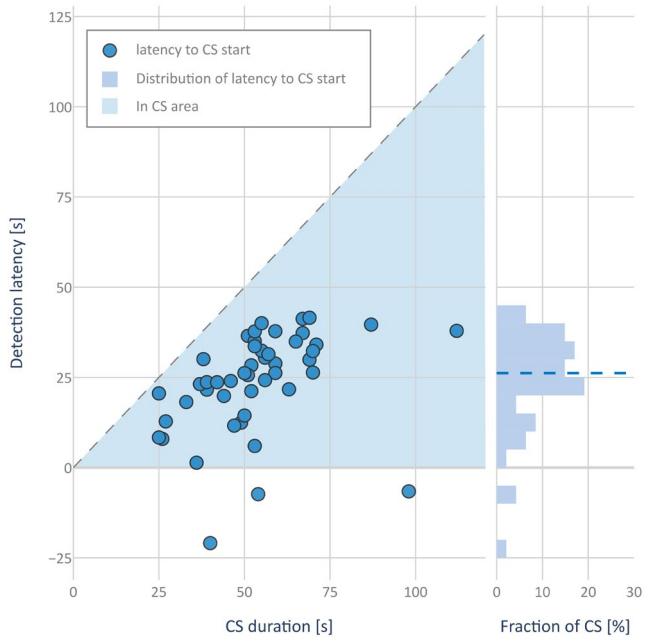

FIGURE 5 Latency of CS (convulsive seizure) detection for quantile 60 with the best model relative to CS onset. Latencies are computed for each detected CS as the time difference between the end time of the first window detecting the CS and the labels of the CS start from the vEEG. The distribution is displayed as a function of CS duration. The gray dashed line is the unitary line that represents points where the latency equals the CS duration. It is thus the upper bound of the region where latencies are within the CS bounds (names *In CS area* in the plot). If a point is above the *In CS area*, it means the CS is detected after the seizure, if the point is below the *In CS area*, it means the CS is detected in advance. In addition, the distribution of latencies is presented as a histogram on the right side of the plot with the median highlighted by a colored dashed line.

the end of the CS (median = −26 s), a feature that would enable timely response from caregivers with respect to the risk of SUDEP. Indeed, all recorded CS-triggered SUDEP were found to occur in the post-ictal phase, with the first apneic episode starting between 30 s and 3 min post-ictal.[7]

The above findings were obtained in one of the largest cohorts of PWE with concomitant vEEG and wearable data, with a total of 384 patients, including 70 with at least one CS captured, and a total of 103 exploitable CSs. Only data from 37 of these patients were used for training and cross-validation. Data from the remaining 347 PWE were kept independent and used solely for testing the performance of the model. In comparison, other series reported independent sets ranging from 27 to a maximum of 152 patients.[25–29] The strict separation between the training and independent test datasets reduces the risk of overfitting and provides a more realistic assessment of real-world performance.[39] The multi-center origin of the data from eight EMUs and five countries is also likely to promote a greater generalizability of our findings.

Most medically certified CS detection devices use dual sensors, coupling 3D-acc with EDA or photoplethysmography (PPG).[25,26,28,29,40] Our choice of developing a CS detection algorithm based on 3D-acc only stems from our goal to provide a solution integrated into low-cost Android-based off-the-shelf smartwatches. Indeed, most smartwatches available to the public include 3D-acc sensors, whereas very few integrate EDA. Furthermore, 3D-acc demonstrates superior artifact resilience and energy efficiency compared to EDA and PPG sensors, enabling continuous 24-h monitoring without requiring battery recharge on a standard off-the-shelf smartwatch. We have successfully validated this capability by implementing our algorithm in such a commercial device (TicWatch Pro 3, Mobvoi, Beijing, China), achieving sustained operation throughout nearly a full day.

Several limitations of this study must be acknowledged. First, our data collection occurred in the controlled environment of EMUs, which does not fully represent real-world conditions. Movement patterns and environmental factors in home settings are likely to differ from those observed in EMUs, which could affect the performance of our algorithm.[36] Second, all analyses were performed off-line and not in a smartwatch. Although we have already been able to integrate and operate our algorithm in an off-the-shelf smartwatch, its performance on such devices might differ from what is reported herein. 3D-acc data do not vary much across wearables but differences in signal amplitude and sampling rate might have an impact on algorithm performances. Thus, future studies are needed to test on-line CS detection with a complete solution, both in EMUs and in ambulatory patients. Ultimately, longitudinal studies assessing the system's impact on clinical outcomes, including SUDEP prevention and quality of life, will be crucial in demonstrating its value in everyday clinical practice.

5 | CONCLUSION

This Phase 2 clinical validation study suggests that robust CS detection might be achieved using a 3D-acc single sensor. The system's combination of strong performance metrics, tunable sensitivity, and implementation capacity into off-the-shelf smartwatches suggests significant potential for clinical application. Future validation in real-world settings will be crucial to fully establish its practical utility.

AUTHOR CONTRIBUTIONS

Antoine Spahr: Conceptualization, methodology, data curation, analysis, visualization, and writing. **Adriano Bernini:** Conceptualization, project management,

methodology, data curation, analysis, and writing. **Pauline Ducouret:** Conceptualization, project management, and data curation. **Christophe Baumgartner:** Patient recruitment. **Johannes P. Koren:** Patient recruitment. **Lukas Imbach:** Patient recruitment. **Sàndor Beniczky:** Patient recruitment. **Sidsel A. Larsen:** patients recruitment. **Sylvain Rheims:** Patient recruitment. **Martin Fabricius:** Patient recruitment. **Margitta Seeck:** Patient recruitment. **Berhard J. Steinhoff:** Patient recruitment. **Isabelle Beuchat:** Patient recruitment. **Jonathan Dan:** Integration of Episave into a smartwatch. **David A. Atienza:** Integration of Episave into a smartwatch. **Charles-Edouard Bardyn:** Conceptualization, methodology, analysis, visualization, writing – review and editing, and supervision. **Philippe Ryvlin:** Project leader, conceptualization, methodology, writing – review and editing supervision, and funding acquisition.

ACKNOWLEDGMENTS

We thank France Ravey and Ilona Hubbard for their assistance in coordinating the multi-center study protocol and Jan Novy and Andrea Rossetti for screening potential candidates at Lausanne university hospital. We are grateful to all epilepsy monitoring units' technicians and clinicians who participated in data collection during the standard clinical routine. We also express our gratitude to all persons with epilepsy who consented to share their data for research purposes.

FUNDING INFORMATION

This work was supported by the Swiss National Science Foundation (SNSF) grants #320030_179240: SEVERITY – Quantifying the severity of generalized tonic–clonic seizures (GTCS) with connected devices; and #CRSII5_193813: PEDESITE – Personalized detection of epileptic seizure in the Internet of Things (IoT) era. M.S. was supported by SNSF 180365. This work was supported in part by the Swiss SNSF, grant no. 10.002.812: "Edge-Companions: Hardware/Software Co-Optimization Toward Energy-Minimal Health Monitoring at the Edge."

CONFLICT OF INTEREST STATEMENT

None of the 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 authors upon reasonable request.

ETHICS STATEMENT

We confirm that we have read the Journal 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 written informed consent.

ORCID

Adriano Bernini https://orcid.org/0000-0003-0207-0064
Johannes P. Koren https://orcid.org/0000-0002-5290-2837
Lukas Imbach https://orcid.org/0000-0002-6135-8642
Sàndor Beniczky https://orcid.org/0000-0002-6035-6581
Sylvain Rheims https://orcid.org/0000-0002-4663-8515
Martin Fabricius https://orcid.org/0000-0002-6959-6848
Berhard J. Steinhoff https://orcid.org/0000-0001-5995-5862
Isabelle Beuchat https://orcid.org/0000-0002-9300-4443
Jonathan Dan https://orcid.org/0000-0002-2338-572X
Philippe Ryvlin https://orcid.org/0000-0001-7775-6576

REFERENCES

1. World Health Organization. Epilepsy. Geneva: World Health Organization; 2024.
2. Beghi E. The epidemiology of epilepsy. Neuroepidemiology. 2020;54(2):185–91.
3. Shan T, Zhu Y, Fan H, Liu Z, Xie J, Li M, et al. Global, regional, and national time trends in the burden of epilepsy, 1990–2019: an age-period-cohort analysis for the global burden of disease 2019 study. Front Neurol. 2024;15:1418926. https://doi.org/10.3389/fneur.2024.1418926
4. Tan M, D'Souza W. Seizure-related injuries, drowning and vehicular crashes – a critical review of the literature. Curr Neurol Neurosci Rep. 2013;13(7):361.
5. Sveinsson O, Andersson T, Mattsson P, Carlsson S, Tomson T. Clinical risk factors in SUDEP: a nationwide population-based case-control study. Neurology. 2020;94(4):e419–e429.
6. Salas-Puig X, Iniesta M, Abraira L, Puig J, QUIN-GTC Study Group. Accidental injuries in patients with generalized tonic-clonic seizures. A multicenter, observational, cross-sectional study (QUIN-GTC study). Epilepsy Behav. 2019;92:135–9. https://doi.org/10.1016/j.yebeh.2018.10.043
7. Ryvlin P, Nashef L, Lhatoo SD, Bateman LM, Bird J, Bleasel A, et al. Incidence and mechanisms of cardiorespiratory arrests in epilepsy monitoring units (MORTEMUS): a retrospective study. Lancet Neurol. 2013;12(10):966–77.
8. Devinsky O, Hesdorffer DC, Thurman DJ, Lhatoo S, Richerson G. Sudden unexpected death in epilepsy: epidemiology, mechanisms, and prevention. Lancet Neurol. 2016;15(10):1075–88.
9. 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.
10. Beniczky S, Ryvlin P. Standards for testing and clinical validation of seizure detection devices. Epilepsia. 2018;59 Suppl 1:9–13.
11. Ramgopal S, Thome-Souza S, Jackson M, Kadish NE, Sanchez Fernandez I, Klehm J, et al. Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy. Epilepsy Behav. 2014;37:291–307.

12. Zhao X, Lhatoo SD. Seizure detection: do current devices work? And when can they be useful? Curr Neurol Neurosci Rep. 2018;18(7):40.
13. 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 the international league against epilepsy and the International Federation of Clinical Neurophysiology. Epilepsia. 2021;62(3):632–46.
14. 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.
15. Jory C, Shankar R, Coker D, McLean B, Hanna J, Newman C. Safe and sound? A systematic literature review of seizure detection methods for personal use. Seizure. 2016;36:4–15.
16. 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.
17. Arends J, Thijs RD, Gutter T, Ungureanu C, Cluitmans P, Van Dijk J, et al. Multimodal nocturnal seizure detection in a residential care setting: a long-term prospective trial. Neurology. 2018;91(21):e2010–e2019.
18. Lockman J, Fisher RS, Olson DM. Detection of seizure-like movements using a wrist accelerometer. Epilepsy Behav. 2011;20(4):638–41.
19. 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.
20. 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.
21. Johansson D, Ohlsson F, Krysl 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. 2019;65:48–54.
22. Szabó C, Morgan LC, Karkar KM, Leary LD, Lie OV, Girouard M, et al. Electromyography-based seizure detector: preliminary results comparing a generalized tonic-clonic seizure detection algorithm to video-EEG recordings. Epilepsia. 2015;56(9):1432–7.
23. Beniczky S, Conradsen I, Pressler R, Wolf P. Quantitative analysis of surface electromyography: biomarkers for convulsive seizures. Clin Neurophysiol. 2016;127(8):2900–7.
24. Zibrandtsen IC, Kidmose P, Kjaer TW. Detection of generalized tonic-clonic seizures from ear-EEG based on EMG analysis. Seizure. 2018;59:54–9.
25. Shah S, Gonzalez Gutierrez E, Hopp JL, Wheless J, Gil-Nagel A, Krauss GL, et al. Prospective multicenter study of continuous tonic-clonic seizure monitoring on apple watch in epilepsy monitoring units and ambulatory environments. Epilepsy Behav. 2024;158:109908.
26. 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(7):2054–68.
27. 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–e97.
28. 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.
29. 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;12:724904. https://doi.org/10.3389/fneur.2021.724904
30. Fisher RS, Cross JH, French JA, Higurashi N, Hirsch E, Jansen FE, et al. Operational classification of seizure types by the international league against epilepsy: position paper of the ILAE Commission for Classification and Terminology. Epilepsia. 2017;58(4):522–30.
31. Harrell FE, Davis CE. A new distribution-free quantile estimator. Biometrika. 1982;69(3):635–40.
32. Loshchilov I, Hutter F. Decoupled weight decay regularization. 2019. *arXiv*. https://doi.org/10.48550/arXiv.1711.05101
33. Pan S, Wang F, Wang J, Li X, Liu X. Factors influencing the duration of generalized tonic-clonic seizure. Seizure. 2016;34:44–7. https://doi.org/10.1016/j.seizure.2015.11.008
34. Alexandre V, Mercedes B, Valton L, Maillard L, Bartolomei F, Szurhaj W, et al. Risk factors of postictal generalized EEG suppression in generalized convulsive seizures. Neurology. 2015;85(18):1598–603.
35. 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 Suppl 1:48–52.
36. van Westrhenen A, Lazeron RHC, van Dijk JP, Leijten FSS, Thijs RD, Dutch TeleEpilepsy Consortium. Multimodal nocturnal seizure detection in children with epilepsy: a prospective, multicenter, long-term, in-home trial. Epilepsia. 2023;64(8):2137–52. https://doi.org/10.1111/epi.17654
37. Lazeron RHC, Thijs RD, Arends J, Gutter T, Cluitmans P, Van Dijk J, et al. Multimodal nocturnal seizure detection: do we need to adapt algorithms for children? Epilepsia Open. 2022;7(3):406–13.
38. 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
39. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. PLoS One. 2019;14(11):e0224365.
40. Regalia G, Onorati F, Lai M, Caborni C, Picard RW. Multimodal wrist-worn devices for seizure detection and advancing research: focus on the Empatica wristbands. Epilepsy Res. 2019;153:79–82.

How to cite this article: Spahr A, Bernini A, Ducouret P, Baumgartner C, Koren JP, Imbach L, et al. Deep learning–based detection of generalized convulsive seizures using a wrist-worn accelerometer. Epilepsia. 2025;66(Suppl. 3):53–63. https://doi.org/10.1111/epi.18406