Computer Methods and Programs in Biomedicine 279 (2026) 109295

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine

# Reliable detection of focal onset impaired awareness seizures in patients with epilepsy using wearable ECG: Development and validation study

Mohamed Alhaskir [a,b], Ekaterina Kutafina [b,c], Florian Linke [a], Florian P. Fischer [a], Elisabeth Schriewer [a], Stephan Lauxmann [d], Kevin Klett [d,g], Julian Hofmeister [d,g], Florian Lutz [d,g], Lukas Burow [e], Michal Cicanic [e], Sara Khosrawikatoli [e], Stefan Wolking [a], Thomas Mayer [e], Sandor Beniczky [f], Josua Kegele [d], Rainer Röhrig [b], Henner Koch [a], Yvonne Weber [a,*]

[a] Section of Epileptology, Department of Neurology, RWTH Aachen University Hospital, Aachen, Germany
[b] Institute for Medical Informatics, Medical Faculty, RWTH Aachen University, Aachen, Germany
[c] Institute for Biomedical Informatics, University of Cologne, Faculty of Medicine and University Hospital Cologne, Germany
[d] Department of Neurology and Epileptology, Hertie Institute for Clinical Brain Research, University of Tübingen, Germany
[e] Epilepsy Center Kleinwachau, Germany
[f] Danish Epilepsy Centre, Dianalund, Denmark and Aarhus University, Aarhus, Denmark
[g] Life Science Incubator Bonn, Germany

ARTICLE INFO

*Keywords:*
Data quality
Neural network
Electrocardiogram
Heart rate variability
Outpatient
Monitoring

ABSTRACT

*Background:* Underreporting of seizures, particularly focal onset impaired awareness seizures (FIAS), compromises the effectiveness of patient care and condition management in patients with epilepsy. Traditional reliance on patient self-reporting can lead to inaccuracies, hindering effective treatment. Wearable-based seizure detection algorithms offer a promising solution, however, developing an efficient method for detecting FIAS remains a challenge. Additionally, as data quality can vary in wearable settings, the absence of continuous data quality assessment poses a concern for the reliability of such algorithms.

*Objective:* The objective of our study is to develop and evaluate the performance and feasibility of FIAS detection algorithm with automatic data quality assessment (ADQA) using a wearable electrocardiography (ECG) device. We will also conduct an exploratory analysis of inter-individual variability in autonomic seizure signatures to identify potential future candidates, or "responders" to this system. Performance will be evaluated using sensitivity, false alarm rate per 24 h (FAR/24), positive predictive value, and F1-Score.

*Methods:* A multicenter study was conducted across three epilepsy centers and recruited patients of all ages who were admitted to video-EEG monitoring for a minimum of 24 h consecutively. Data were collected using a wearable ECG device. The algorithm involved R-peak detection to identify heartbeats, extraction of knowledge domain heart rate variability features, ADQA, heart rate (HR) filter to address class imbalance, and a deep learning model for the final detection step. The algorithm was validated in a leave-one-patient-out (LOPO) approach using expert-labeled ictal events from video-EEG monitoring as ground truth.

*Results:* A total of 236 patients were recruited, of whom 49 patients experienced at least one FIAS, resulting in 3278 h of ECG data and 260 seizures. Two patients with 33 seizures were excluded due to a technical error in the recording files, leaving 47 patients for analysis. After data quality screening, 161 seizures from 38 patients met the quality criteria. In this group, the median sensitivity was 66.6% (95% CI:33.3%-100%) with a median FAR/24 of 5.2 (95% CI:3.5-8.2). An exploratory responder analysis identified 20 patients with a detection sensitivity of ≥66.6%, for whom the median sensitivity was 100% (95% CI: 92%–100%) and the median FAR/24 was 4.3

*Abbreviations:* EEG, Electroencephalography; Vem, Video-EEG Monitoring; Fias, Focal Onset Impaired Awareness Seizure; Ecg, Electrocardiography; Hrv, Heart Rate Variability; Adqa, Automatic Data Quality Assessment; Ml, Machine Learning; Dl, Deep Learning; Far/24, False Alarm Rate in 24 H; Tp, True Positive; Fp, False Positive; FN, False Negative; PPV, Positive Predictive Value; LOPO, Leave-One-Patient-Out; bpm, Beats Per Minute; CI, Confidence Interval; HR, Heart Rate; $HR_{max}$, Maximum Heart Rate; $HR_{min}$, Minimum Heart Rate; RR, R-R Intervals (time between two successive R-wave peaks in ECG).

* Corresponding author.
*E-mail address:* yweber@ukaachen.de (Y. Weber).

https://doi.org/10.1016/j.cmpb.2026.109295
Received 22 May 2025; Received in revised form 28 January 2026; Accepted 22 February 2026
Available online 23 February 2026
0169-2607/© 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

(95% CI: 3–7). Finally, removing ADQA from the test data reduced the algorithm's reliability, while removing it from training and test data reduced sensitivity, robustness, and reliability.
*Conclusions:* The proposed algorithm demonstrated reasonable performance in patients whose wearable ECG data met the ADQA quality criteria (n = 38), with the highest detection performance observed in an exploratory responder subgroup (n = 20). These findings highlight the potential of ECG-based wearable systems for improving FIAS monitoring and underscore the importance of data quality in ensuring reliable algorithm performance.
*Trial Registration:* German Clinical Trials Register: DRKS00026939

## Introduction

Epilepsy is a global neurological disorder affecting the health and quality of life of more than 50 million people worldwide [1]. The disease is defined by the appearance of repetitive, unprovoked seizures which can be focal (starting in a defined area of the brain) or generalized [2]. Despite successful clinical treatment, appropriate monitoring of the condition progression and patient safety requires accurate detection and documentation of seizures in daily life environment [3]. Continuous video-EEG monitoring (VEM) is the gold standard for seizure detection [4], however, technical limitations, its stigmatizing nature and the need for professional oversight limit its suitability for long-term ambulatory use [3]. Patient self-reporting, on the other hand, while commonly used [3], proves to be inefficient and inaccurate, especially with focal onset impaired awareness seizures (FIAS) [5]. For example, in the case of bilateral tonic-clonic seizures, a study has reported that up to 55% of the seizures remain undocumented, whereas in FIAS, the proportion rises to 88% [5]. Seizures may go undetected due to a lack of witnesses, or when patients are unaware during sleep [3]. Consequently, healthcare providers are unable to assess seizures frequency and the effectiveness of treatment [3,6]. In addition, seizures can arise during activities requiring alertness, such as driving or operating machinery, putting patients' lives at risk [7]. As a result, interest in using automated systems for seizure monitoring in daily life has been growing [8,9]. Several approaches using different modalities, such as accelerometry, electrodermal activity, and electromyography, have been proposed [10–12]. For detecting seizures that do not always involve a strong physical component such as FIAS [13], wearable electrocardiography (ECG) has been shown to be a feasible solution. Epileptic seizures can cause related changes in heart rhythms which can be picked up through heart rate variability (HRV) measurements [14–17]. There have been multiple studies that have proposed ECG-based solutions for seizure detection or prediction [18–22]. However, only algorithms detecting tonic-clonic seizures have demonstrated readiness for real-world application, whereas more proof is needed to verify the effectiveness of algorithms in detecting FIAS [23]. This is particularly important when considering that focal onset epilepsy makes up most epilepsy cases, with FIAS being its most common type [24,25]. Another important challenge is the reliability of seizure detection systems in general [26]. This concern is linked to a lack of automatic data quality assessment (ADQA) in wearable detectors when data could exhibit variation in its quality depending on the user activity [26,27]. Although the importance of data quality has been explored in other fields [28–30], this subject has not been widely addressed in seizure detection systems [27]. A seizure detection algorithm needs to automatically recognize that its output may be compromised when input data cannot be trusted due to missing physiological information or artifacts [26]. In addition, ensuring data quality is not only vital for applications involving wearable devices but also for machine learning (ML) and deep learning (DL). High-quality data can reflect positively on the model's performance and robustness [29,30]. Also, implementing a clear strategy for data quality measurement ensures reproducibility and transparency, thus increasing the model's trustworthiness [30–32]. Therefore, in this work, we aim to evaluate the performance and feasibility of a seizure detection algorithm for FIAS, a common and frequently underreported seizure type for which detection algorithms remain limited [23–25], using data from a chest-worn wearable ECG device with ADQA. This study serves as an initial validation step within the controlled VEM environment prior to real-world deployment. Furthermore, we explore inter-individual variability in autonomic seizure responses to identify a patient subgroup, refer to as 'responders,' who may be more likely to benefit from this technology.

## Methods

### *Ethics statement*

The study received approval from the ethical committees of the University of Tübingen (Vo.-No. 186/2016BO1), University Hospital Aachen (Vo.-No. 381/19) and Kleinwachau (Vo.-No. EK-BR-71/18-1), Germany. Each patient received an oral and written explanation of the study design and purpose. Patients could participate in the study after signing an informed consent form. In the case of minors capable of giving consent or non-consent-capable adults, at least one guardian informed about the research intentions must consent to the participation. The data was stored with no personal identifiers, on password-protected and encrypted servers that are only accessible to authorized members of the research team. The study was registered at the German Clinical Trials Register (DRKS-ID: DRKS00026939).

### *Recruitment*

This diagnostic study recruited patients with epilepsy of all ages consecutively, at three locations from 2016 to 2021: the University of Tübingen, Germany; University Hospital Aachen, Germany; and the Kleinwachau Epilepsy Center, Germany. Patients were approached for inclusion in the study if they were scheduled to remain in the monitoring units for at least 24 h and had provided informed consent to participate. Patients with severe cardiac comorbidities were excluded from enrollment due to safety concerns and the potential for ECG data interference. Due to the unknown prevalence of FIAS among these patients, the likelihood of patients experiencing FIAS during the recording period with the wearable device, and their willingness to participate in the study, it was challenging to estimate the sample size. Given the undefined sample size and the observational character of the study, a minimum of at least 10 patients with 15 seizures was intended to ensure an appropriate validation of the proposed algorithm [33]. Finally, for the algorithm development and evaluation, only patients who experienced at least one FIAS during monitoring were included in the analysis.

### *Data collection*

ECG data used in the algorithm training and evaluation were captured using the wearable EcgMove4 (movisens GmbH, Karlsruhe, Germany), a scientific research-grade tool. This device is available to the public for research use only and operates at a sampling frequency of 1024 Hz, with an amplitude range of ±5mV and a 12-bit analog-to-digital converter. The sensor was applied by research staff and positioned on the lower chest area during the patients' stay in the VEM unit, where patients were predominantly resting in bed but allowed limited mobility. Raw data were exported using SensorManager software

version 1.12.7 and no additional signal preprocessing or filtering (e.g., bandpass filtering, artifact removal) was applied beyond device export. In addition to ECG recordings, we documented clinical variables such as age at epilepsy onset, epilepsy type, seizure type observed during VEM, seizure onset zone, and antiseizure medication administered during data acquisition.

*Data labeling*

Simultaneous VEM was performed during the data acquisition process using the wearable ECG system. Ictal events timestamps (seizure onset, seizure type, end of seizure in EEG) were defined by expert neurologists and used as ground truth for algorithm validation. These reference timestamps were also accessible to the assessor of the algorithm output.

*Feature extraction methods*

During the data preprocessing phase, R-peaks were identified using the Hamilton R-peak detection algorithm [34]. Then, to avoid increasing model complexity [35], we incorporated domain knowledge features for their effectiveness in detecting autonomic changes during seizures [14, 17–19]. Here, HRV parameters, including time-domain and non-linear features, were calculated using a sliding window approach with a 5-second step. The window size for time-domain features was 30 s and for non-linear features 60 s (Supplementary-Table 1). These two window sizes were chosen to balance the need for timely detection of transient changes associated with ictal events, which have a median duration of approximately 42.5 s in FIAS cases [36], and the reliability of feature computation. Studies have shown that the time-domain HRV measures can be computed from recordings as short as 30 s [37], whereas non-linear metrics such as HR Diff, CSI, and CSIM require longer recordings [19]. Following HRV parameter computation, the features were segmented into 300-second epochs with a 5-second moving step. According to literature, HRV changes can begin several minutes before a seizure and persist afterward [14–16]. Thus, this 300-second window ensures the model has enough temporal context to detect these patterns. Finally, to standardize the number of data points within each 300-second segment of each feature set, the features computed over 60-second windows were up-sampled using linear interpolation to match the number of data points of those computed over 30-second windows.

*Automatic data quality assessment*

The main objective of this step is to ensure the model does not correlate corrupted features arising from noise or artifacts with seizures in the training phase while also ensuring a reliable level of the algorithm's performance across patients when the data quality threshold is met in the testing phase. In this regard, we employed an adapted ADQA algorithm to evaluate the quality of wearable ECG data [38]. Similar to the original work, we applied the quality assessment initially in a moving window of 10 s with a step size of 5 s. The ADQA relied first on the three following physiological criteria:

1. The heart rate (HR) must range between 40 and 180 beats per minute (bpm).
2. A single interval between successive R-peaks should not exceed three s.
3. The ratio of the maximum to minimum RR interval within the segment must be less than 2.2.

In the original work, the fourth condition calculates the average correlation between a QRS template and all QRS complexes within a 10-second window. We replaced this step with a beat detection signal quality index, which quantifies the agreement between two R-peak detection algorithms (Hamilton [34] vs. slope sum function [39]) applied to the same segment [40]. This change was made because our analysis relies primarily on accurate R-peak detection and HRV features rather than the morphology of the QRS complex. To compute the quality score for each segment, the algorithm first applies the three initial conditions. When a segment does not fulfill any of these conditions, it is assigned a score of 0. Otherwise, if the segment passes, it is then analyzed using the beat detection signal quality index. The score is determined by calculating the ratio of matched R-peaks between the two algorithms, expressed as a percentage. A perfect agreement between the two algorithms results in a score of 100%, while lower percentages reflect higher disagreement and therefore lower data quality. Finally, we computed the average quality for each 300-second epoch, and only segments scoring $\geq$75% were labeled as "good". This threshold was chosen as an operating point to clearly separate low-quality from high-quality wearable ECG epochs and to quantify the impact of signal quality on model performance and reliability. Segments that did not pass the quality filter were added to the algorithm deficiency time, defined as the time during which the algorithm cannot produce reliable output and is therefore not functional [41].

*Seizure detection algorithm*

The detection algorithm consisted of two steps (Fig. 1). The first is an HR filter, which identifies segments of interest acting as a simple but highly sensitive classifier. This is followed by a DL model for a more conservative classification. The logic behind this approach is to address the class imbalance between seizure and non-seizure events, thereby avoiding bias towards predicting the majority class, which would result in poor detection of seizures (the minority class). This approach also allows to train the model on segments that could exhibit similar changes in HRV parameters. These two steps are further clarified as follows:

1. *HR Filter*

In order to address class imbalance, 300-second epochs that exhibited prolonged tachycardia were identified based on a threshold of 85 beats per minute (bpm). This approach was motivated by a previous study showing that during epileptic seizure, tachycardia, defined as an increase in HR of more than 10 bpm above baseline, frequently surpassed 85 bpm [14]. Accordingly, HR values were derived from the computed MeanNN as follow:

$$HR = \frac{60}{MeanNN}(bpm)$$

As shown in Supplementary-Table 1, MeanNN was calculated over 30-second windows with a 5-second step within each 300-second segment, resulting in 55 HR estimates per segment. Corresponding, if at least 10% of derived HR estimates exceeded 85 bpm, the segment was passed to the DL model for classification. Otherwise, it was filtered out and labeled as non-seizure.

2. *Deep learning model*

Several studies have investigated various ML techniques, including support vector machines, random forests, and k-nearest neighbors, alongside DL models, for seizure detection in wearable devices [42–44]. However, the main aim of this work was not to provide an ML/DL model comparison but rather to introduce and evaluate an approach for seizure detection that incorporates ADQA with a transformer-based architecture [45]. This approach leverages the transformer attention mechanism to capture dependencies more effectively than traditional models that rely on fixed or sequential temporal representations. Indeed, recent studies have shown the value of attention-based neural networks for capturing seizure-related changes using HRV features, underscoring the potential of attention in this area [21,22]. The proposed model was implemented as an encoder component of a transformer-based neural network, trained from scratch using ECG-derived HRV features. Each input sequence $X \in \mathbb{R}^{55\times 28}$ represented 28 HRV features extracted over

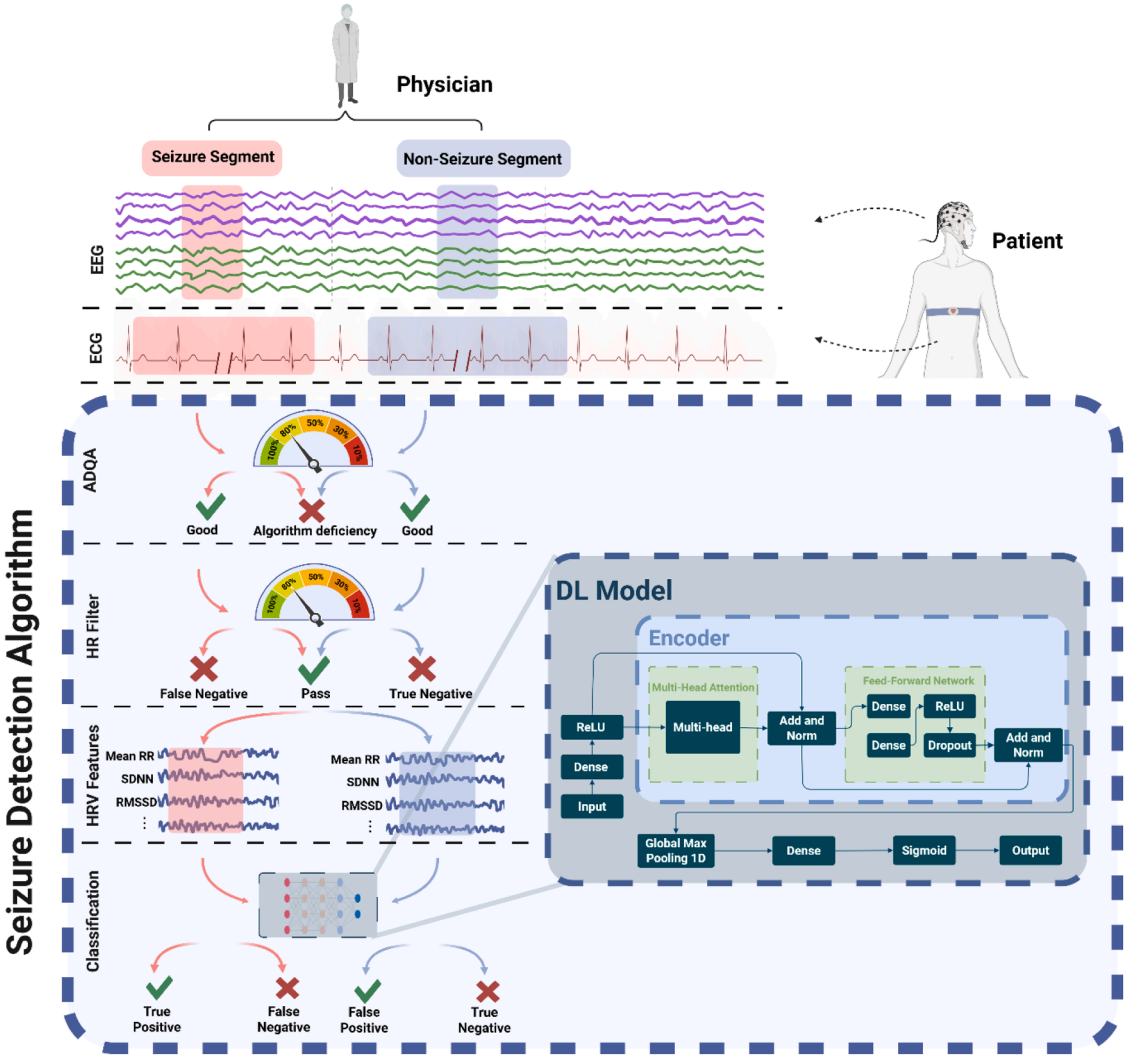


**Fig. 1.** Study and Algorithm Evaluation Workflow. Simultaneous EEG and ECG data recording is performed in the VEM unit during the patient's stay at the clinic. Seizure (pink) and non-seizure (blue) segments are labeled based on ictal events identified continuously by an expert neurologist. ADQA is performed on each ECG segment, and segments not meeting the quality criteria are assigned to the algorithm deficiency time. Segments that pass the quality check are processed by the HR filter for potential event detection. If a seizure (pink) is not detected by the HR filter, it is considered FN. If a segment passes through the HR filter, HRV features undergo final classification by the DL model. The DL model includes an encoder with a multi-head attention mechanism and a feed-forward network, which processes the HRV features to determine the presence or absence of a seizure. The DL model incorporated ReLU, dense layers, dropout, add and norm layers, and a sigmoid function for output activation. ADQA: Automatic Data Quality Assessment, DL: Deep Learning, EEG: Electroencephalography, ECG: Electrocardiogram, FN: False Negative, FP: False Positive, TN: True Negative TP: True Positive, HR: Heart Rate, HRV: Heart Rate Variability, ReLU: Rectified Linear Unit, outputs the input if positive; otherwise, zero, Dense: Fully connected neural network layer, Dropout: Randomly zeros inputs to prevent overfitting, Add and Norm: Combines and standardizes inputs, Sigmoid: applies a sigmoid function to inputs VEM: Video-EEG monitoring.

55-time steps from a 300-second window. Each time step $x_t \in \mathbb{R}^{28}$ was projected into a 64-dimensional latent space via a fully connected embedding layer with ReLU activation:

$$h_t = ReLU(W_e x_t + b_e), W_e \in \mathbb{R}^{64\times 28}$$

This produced an embedded sequence $H \in \mathbb{R}^{55\times 64}$, which was passed to a single Transformer encoder layer consisting of:

1. Multi-head self-attention with 2 heads, each computed as:

$$Attention(Q,K,V) = softmax\left(\frac{QK^t}{\sqrt{d_k}}\right)V, d_k = 64$$

Where Q, K, V are learned projections of H.

2. For each time step $t$, the output vector from the self-attention block, denoted $z = H_t' \in \mathbb{R}^{64}$, was passed through a Feedforward network (FFN):

$$FFN(z) = W_2 ReLU(W_1 z + b_1) + b_2$$

Where $W_1 \in \mathbb{R}^{32\times 64}$ and $W_2 \in \mathbb{R}^{64\times 32}$ and $b_1$, $b_2$ are bias terms.

3. Add and Normalize layers were applied after both the self-attention and feed-forward network (FFN) blocks and dropout with rate of 0.2 applied after the FFN to mitigate overfitting.

To aggregate temporal information, a global max pooling was applied to the output of the encoder, denoted as $H_{trans} \in \mathbb{R}^{55\times 28}$ across time steps:

$$z_{pool} = max_t(H_{trans}[t]), z_{pool} \in \mathbb{R}^{64}$$

Finally, the pooled vector was passed through a fully connected layer with sigmoid activation for binary classification:

$$\hat{y} = \sigma\left({w_{out}}^T z_{pool} + b_{out}\right), with \hat{y} \in [0,1]$$

The model was trained using the Adam optimizer (learning rate =

0.001), with a batch size of 32 and a maximum of 20 epochs, employing early stopping based on validation loss. Hyperparameter choices were informed by prior work on transformer-based time-series models [46, 47], where similar configurations were shown to be effective, and were further refined using grid search. Also, to further address class imbalance, we used a weighted binary cross-entropy loss during training, assigning a weight of 8:1 to FIAS versus non-FIAS segments. This penalized misclassification of seizure events more heavily and encouraged the model to focus on ictal patterns.

*Training and testing strategy*

We used a Leave-One-Patient-Out (LOPO) variation of k-fold cross-validation for training and testing, where k equals the number of patients in the dataset. This approach allowed a comprehensive assessment of algorithm performance under varying signal quality conditions and across diverse patient profiles. For each fold, data from all patients except one were used for model training (Supplementary-Fig. 1). Training data consisted of HRV features from a single 300-second segment per FIAS, together with non-seizure segments, all of which had passed the HR and data quality filters and were forwarded to the DL model for training. During testing, the complete detection pipeline was applied to the held-out patient. Specifically, the extracted features segmented into overlapping 300-second windows of the test patient, that also passed the HR and data quality filters were forwarded to the DL model for classification, enabling continuous detection. This process was repeated until all patients had been tested. As the model assigned a value between 0 and 1 for each 300-second epoch, indicating the probability of the segment being a seizure, we performed a parameter scan from 0 - 1 to estimate the algorithm's overall performance at different probability thresholds.

*Algorithm evaluation*

In this section, we describe the assessment strategy of the algorithm's performance. First, the following metrics were defined:

- True positive (TP): A detection was considered a true positive if it occurred within a ±300-second window surrounding the labeled seizure onset. This window accounts for seizure-related physiological changes, which may begin before or persist after the electrographic onset [14–16].
- False positive (FP): were all the events that were classified as seizures by the algorithm but were not within ±300 s of any seizure onset.
- False negative (FN): are seizures that were not detected by the algorithm.

It is important to note that the HR filter was considered part of the algorithm's output. Seizures that were not detected by the HR filter, and in turn not passed to the DL model, were treated as FNs. Thus, the final results presented in this work will reflect the output of the whole pipeline. The algorithm's efficacy was further assessed using sensitivity, false alarm rate per 24 h (FAR/24), positive predictive value (PPV), and F1-Score. We employed event-based scoring, treating each seizure as a single event [48]. In this approach, a seizure was counted only once as a TP if it was correctly detected, regardless of how many overlapping windows identified it. Furthermore, as this work is focused on the detection of FIAS seizures across all patients, if other seizure types were present, they were not included in the calculation of true positives (TPs) or false negatives (FNs). In a secondary, exploratory analysis, and consistent with a previous study [19], we defined responders as patients with ≥66.6% of their FIAS detected, whereas patients with less than two-thirds of their seizures detected were classified as non-responders. Only FIAS was considered to label a patient as a responder. If a patient had three seizures, of which one was a FIAS, and this seizure was not detected, then this patient is also a non-responder. To account for the wide variability in the number of seizures per patient, patient-level performance metrics were computed by first aggregating detections within each patient and then summarizing results across patients (e.g. using medians), such that each patient contributed equally regardless of seizure count. In contrast, cohort-level metrics were computed by pooling all seizures and recording time across patients, thereby reflecting overall event-level performance. Metrics computed at the cohort level are referred to as 'total [metric]'. Conversely, metrics computed at the patient level are preceded by either 'median' or 'mean,' depending on the method used for computation. Parametric confidence intervals were used for mean-based summaries, whereas rank-based confidence intervals were used for median-based summaries. Finally, to examine the effect of ADQA on the model's performance, we conducted additional evaluations by removing the quality filter. This allowed us to assess the model on low-quality data segments by first excluding the filter from the test data only and then from both the training and test data.

## Results

*Study Cohort and dataset characteristics*

1. *Patients*

Out of the 236 patients included in the study, 49 experienced at least a single FIAS during the data collection period. None of these patients experienced non-epileptic seizures. Two patients were excluded due to technical errors in the recording files. The diagram in Fig. 2 illustrates the flow of participants with FIAS through the study, from initial recording to final inclusion in the analysis.

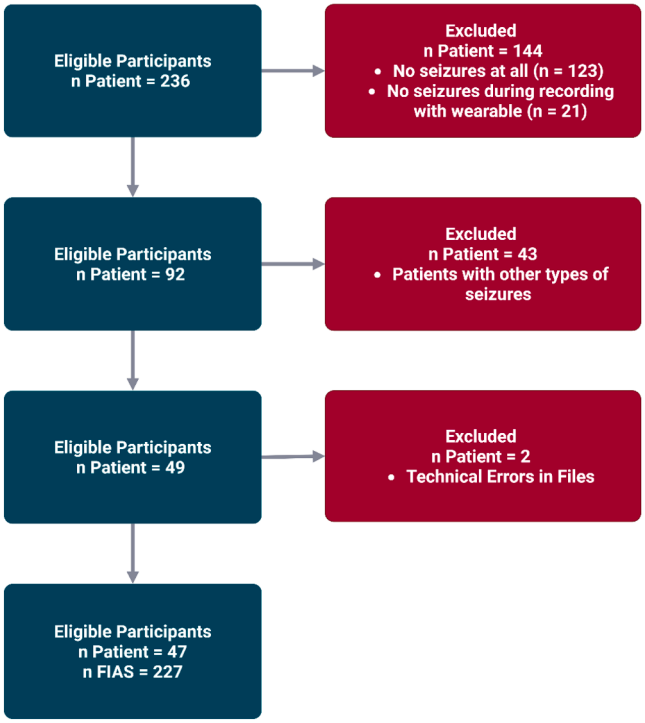


**Fig. 2.** Participant selection flowchart. This diagram illustrates the flow of participants through the stages of the study. A total of N = 236 patients were initially recorded. Of these, N = 49 patients experienced at least a single FIAS during the wearable recording period. N = 2 patients were excluded from the analysis due to technical errors in the recording files, leaving N = 47 patients with a total of n = 227 seizures that were included in the final analysis. FIAS: focal onset impaired awareness seizure.

## 2. Dataset

The total recording duration of the 47 patients was 3141 h, with a median of 67 h per patient, ranging between 24 and 146 h. No adverse events or systematic technical difficulties were observed from wearing the wearable ECG device. From the 47 included patients, 25 were females and 22 were males. The median age was 37 years (mean = 40, standard deviation= 16, range = 13 - 81), including one pediatric case. Further, the median number of FIAS was 3 seizures (mean = 4.8, standard deviation= 5.2, range = 1 - 30). Table 1 includes a summary of the demographic and clinical background of all patients that were included in the analysis. Specifically, information about the type of epilepsy, the affected regions, and the medications administered to the patients during the data collection process is provided. Overall, we included 33 patients with lesional focal epilepsy versus 14 patients with non-lesional epilepsy. 28 patients had temporal seizures, 11 had frontal seizures, 8 had fronto-temporal seizures, 1 bilateral diffuse seizure, and 2 had parieto-occipital seizures.

**Table 1**
Patients Demographics and Clinical Characteristics:
Information about the distribution of sex, age at the time of the study, and age at seizure onset along with the type of focal epilepsy (lesional or non-lesional), the region and side of the seizure onset zone, and the antiseizure medications administered during the study. Percentages may exceed 100% for seizure onset zone region and side, and for antiseizure medications, because categories were not mutually exclusive.

| Total Patient (n = 47) | No. (%) | Median (Range) |
|---|---|---|
| **Sex** | | |
| **Male** | 22 (47) | |
| **Female** | 25 (53) | |
| **Age, y** | | 37 (13-81) |
| **Age at Seizure Onset, y** | | 18 (1-73) |
| **Focal Epilepsy Type** | | |
| **Lesional** | 33 (70) | |
| **Non-lesional** | 14 (30) | |
| **Region of Seizure Onset Zone** | | |
| **Temporal** | 28 (60) | |
| **Frontal** | 11 (23) | |
| **Parieto-occipital** | 2 (4) | |
| **Fronto-temporal** | 8 (17) | |
| **Diffuse** | 1 (2) | |
| **Side of Seizure Onset Zone** | | |
| **Right** | 16 (34) | |
| **Left** | 23 (49) | |
| **Central** | 2 (4) | |
| **Bilateral or Left and Right** | 8 (17) | |
| **Antiseizure Medication** | | |
| **LEV** | 18 (38) | |
| **LCM** | 15 (32) | |
| **LTG** | 13 (28) | |
| **VPA** | 6 (13) | |
| **OXC** | 8 (17) | |
| **ZNS** | 5 (11) | |
| **BRV** | 8 (17) | |
| **TPM** | 3 (6) | |
| **CLB** | 4 (8) | |
| **PER** | 2 (4) | |
| **ESL** | 2 (4) | |
| **CBD** | 2 (4) | |
| **THC** | 1 (2) | |
| **PHT** | 2 (4) | |
| **CBZ** | 1 (2) | |
| **CLZ** | 1 (2) | |

BRV: Brivaracetam, CBD: Cannabidiol, CBZ: Carbamazepine, CLB: Clobazam, CLZ: Clonazepam, ESL: Eslicarbazepine, LCM: Lacosamide, LEV: Levetiracetam, LTG: Lamotrigine, OXC: Oxcarbazepine, PER: Perampanel, PHT: Phenytoin, SOZ: Seizure Onset Zone, THC: Tetrahydrocannabinol, TPM: Topiramate, VPA: Valproic Acid, ZNS: Zonisamide.

### *Data quality assessment and analysis populations*

Following the ADQA (Supplementary-Fig. 2) and quality filtering using the 75% threshold, the algorithm experienced a deficiency time of 100 h, representing 3.2% of the total 3,141 h. Nevertheless, this period included 66 seizures, indicating that 161 out of the 227 total FIAS (71%) in the dataset met the quality criteria by achieving an average quality of at least 75%. Since the distribution of seizures across patients varied, some patients who had only one or a few seizures, all of which occurred during poor-quality segments, were excluded. Consequently, the 161 seizures corresponded to 38 out of the 47 patients in the study.

### *Seizure detection performance*

First, we performed our examination using training and testing data that met the previously described data quality criteria. Following the detection threshold scan, the mean F1-score increased with the detection threshold, increasing from 3% at a threshold of 0 to a maximum of 22% at a threshold of 0.9 before dropping to 0 at threshold 1 (Fig. 3A). When we adjusted the performance metric in our secondary analysis for responders only, F1-score increased from 3.7% at 0 to 42% at 0.9 (Fig. 3B). However, it is also important to note that the number of responders decreased simultaneously from 26 at a threshold of 0 to 17 at 0.9. These observations highlight a trade-off between detection performance and the proportion of patients benefiting from the algorithm across different thresholds.

Based on this trade-off, we selected a detection threshold of 0.84 that balanced sensitivity and false alarm burden in the cohort of patients whose data met the ADQA quality criterion. At this threshold, the median sensitivity across the 38 patients who passed quality filtering was 66.6% (95% CI: 33.3%–100%) with a median FAR/24 of 5.2 (95% CI: 3.5–8.2). Using the same threshold, the responder subgroup (n = 20) exhibited a median sensitivity of 100% (95% CI: 92%–100%) with a median FAR/24 of 4.3 (95% CI: 3–7). We then re-evaluated the algorithm performance after removing the quality filtering from the test data. We observed a drop in median and total sensitivity between the test group of patients with data quality $\geq$ 75% (n =38) and all patients together (n =47). This decrease was larger in the median sensitivity (from 66.6% to 33.3%) than in total sensitivity (56% to 42.7%). There was also an increase in the total FAR/24 from 7.47 to 8.74 in the group that included all patients (Fig. 4A). Detailed results for the 0.84 threshold for each patient group, including the number of seizures, recording time, and performance metrics, are provided in Table 2.

On the other hand, when training was performed on the entire dataset without excluding low-quality segments, we observed a general increase in FAR/24 across all test groups, accompanied by a decrease in sensitivity (Fig. 4B). This was evident both on the cohort level (total sensitivity and FAR/24) and the patient level (median sensitivity and FAR/24).

### *Post-Hoc analysis*

Following the algorithm's outcome, we conducted a classification of seizure events based on their time of day (day (6:00– 22:00) vs. night (22:00 – 6:00)). The analysis did not reveal a difference in the rate of seizure passage through the HR filter between night- and daytime events (Supplementary-Fig. 3). We then compared algorithm performance between responders and non-responders within the cohort of patients whose data met the ADQA quality criterion (data quality $\geq$ 75%). As previously mentioned, responders (n = 20) demonstrated substantially higher performance, with a median sensitivity of 100% (95% CI: 92%–100%) and a median FAR/24 of 4.3 (95% CI: 3–7). In contrast, non-responders (n = 18) exhibited markedly poorer performance, with a median sensitivity of 8.3% (95% CI: 0%–33.3%) and a median FAR/24 of 5.6 (95% CI: 3.5–13.3). To explore potential physiological differences underlying this performance gap, we compared the median ictal HR

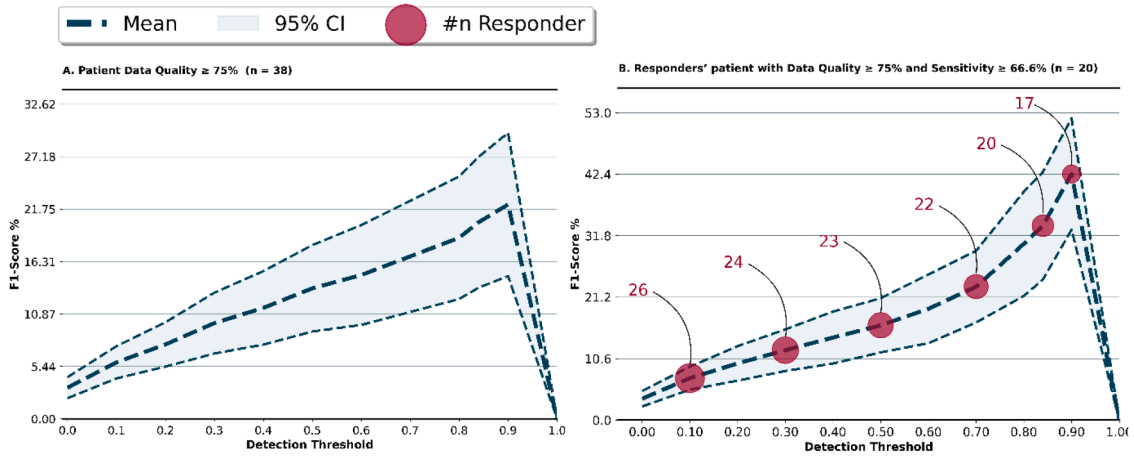


**Fig. 3.** Impact of Detection Threshold on Algorithm Performance Across Patients who passed the quality filtering (A) and After Correction for Responders (B). The figure illustrates the effect of varying detection thresholds (from 0 to 1) on the algorithm's performance, measured by the F1-Score, across patients with data quality ≥ 75% (A) and after correcting for the remaining responders at each threshold level (B). The mean F1-Score is represented by a dark blue line, while the light blue shading denotes the 95%CI of the mean. Red circles indicate the number of responders (patients with sensitivity of ≥ 66.6%) at each threshold level. CI: confidence interval.

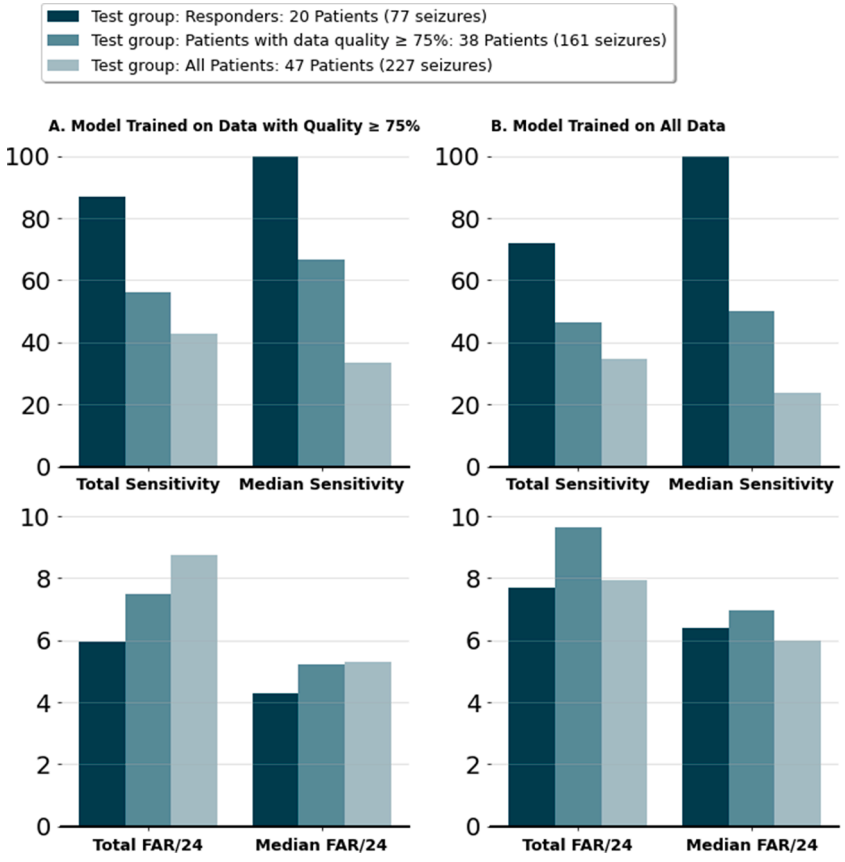


**Fig. 4.** Effects of ADQA on Algorithm's Performance and Reliability. This figure compares the total and median sensitivity and FAR/24 across three test groups: Responders (20 patients, 77 seizures), patients with data quality ≥ 75% (38 patients, 161 seizures), and all patients (47 patients, 227 seizures). Panel (A) presents results where the training was conducted using data with quality ≥ 75%, while panel (B) shows the results when all available data were used for training. The y-axis represents the value of each metric, while the x-axis indicates the method by which the metric was computed (i.e., Total across the cohort vs. Median across patients). FAR/24: false alarm rate per 24 h.

change between responders and non-responders during seizures (Supplementary-Fig. 4). Responders showed a significantly larger HR increase (>40 bpm; median = 49.83 bpm, 95% CI: 38.54–68.76) than non-responders (median = 24.85 bpm, 95% CI: 19.43–26.79; Mann–Whitney $U = 16$, $p < 0.0005$; rank-biserial $r = 0.91$). However, this pattern was not uniform across all responders, as six responders exhibited HR changes below 40 bpm, comparable to those observed in non-responders. Finally, we performed feature importance analysis using Shapley Additive exPlanations (SHAP) [49]. Features that reflect an increase in short-term variability, such as high SD1, along with an

**Table 2**
Algorithm Performance Metrics at Threshold 0.84 for each Patient Group.
For each group, the total, median, and 95% confidence interval (CI) for recording time (RT hours), Heart-rate filter detection (HR detected), number of seizures (N Seizure), true positives (TP), false positives (FP), false negatives (FN), and the false alarm rate calculated both per 24 h (FAR/24). Additionally, performance indicators such as the positive predictive value (PPV%), Sensitivity (Sen%), and F1-Score are reported.

| Group | Metric | RT (h) | N Seizure | HR detected | TP | FP | FN | FAR/24 | PPV% | Sen% | F1% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| All Patients included in the Study (n = 47) | Total | 3141 | 227 | 180 | 97 | 1144 | 130 | 8.74 | 7.8 | 42.7 | 13.2 |
| | Median | 66.81 | 3 | - | - | - | - | 5.27 | 7.1 | 33.3 | 10.00 |
| | 95%CI | 51.55-72.25 | 2-4 | - | - | - | - | 3.82-7.32 | 0-12.5 | 0-50 | 0-18.4 |
| Patients with Data Quality ≥ 75% (n = 38) | Total | 2465.72 | 161 | 139 | 90 | 768 | 71 | 7.47 | 10.4 | 56 | 17.6 |
| | Median | 65.72 | 2.5 | - | - | - | - | 5.20 | 10.1 | 66.6 | 17.4 |
| | 95%CI | 50.42-72.0 | 2-5 | - | - | - | - | 3.50-8.20 | 5.0-15.3 | 33.3-100 | 8.3-26.3 |
| Responders' patient with Data Quality ≥ 75% and Sensitivity ≥ 66.6% (n = 20) | Total | 1318.53 | 77 | 67 | 67 | 326 | 10 | 5.93 | 17.0 | 87 | 28.5 |
| | Median | 66.55 | 2 | - | - | - | - | 4.30 | 16.7 | 100 | 28.07 |
| | 95%CI | 43.81-72.74 | 1-5 | - | - | - | - | 3-7 | 14.2-28.2 | 92-100 | 25-43.3 |

overall reduction in RR intervals variability (Quantile 25 and Variance), seem to contribute more to a positive detection compared to measures that capture the overall balance between sympathetic and parasympathetic activity, such as CSI (Supplementary-Fig. 5).

## Discussion

### *Main findings*

This study evaluated a wearable-based seizure detection algorithm that integrates ADQA as a core component of the detection pipeline with a transformer encoder architecture [45] in a multicenter cohort of n = 47 patients with FIAS. When training and testing the model on data that met the predefined data quality threshold ≥ 75%, the algorithm achieved a median false alarm rate per 24 h (FAR/24) of 5.2 (95% CI: 3.5–8.2) and a median sensitivity of 66.6% (95% CI: 33.3%–100%) across the 38 patients. In our secondary responder analysis, the median FAR/24 decreased to 4.3 (95% CI: 3–7), with a median sensitivity of 100% (95% CI: 92%–100%) in 20 patients. A reduction in sensitivity was observed when the testing was done on all the patients without quality labeling. This was followed by a larger increase in FAR/24 when the training were performed on all the data without excluding low-quality segments across the same groups. Together, these findings demonstrate the importance of integrating ADQA into wearable seizure detection systems, enabling explicit quantification of the impact of signal quality on algorithm performance and transparent identification of periods in which algorithm output may be unreliable.

### *Comparison with literature*

Numerous studies have explored the detection of FIAS [19,50–53]. However, comparing the current work to these studies can be challenging. This difficulty arises because performance metrics were often reported across various seizure types without distinguishing between them, making direct comparisons of the detection of FIAS complex. Among these studies, the performance of ECG-based seizure detection algorithms is summarized in Table 3.

To the best of our knowledge, only a few studies have focused explicitly on detecting FIAS while none have applied data quality assessment as part of their seizure detection algorithm. In one such study, a behind-the-ear EEG sensor was used, reporting a sensitivity of 64.1% and a FAR of 2.8 per 24 h in their patient-independent model. The study included 42 patients with seizures, of which 90% were FIAS [54]. In another study, a wearable EEG device was used to monitor nine patients with seizures, yielding a sensitivity of 52% and a FAR/24h of 171 [55]. A third study employed a multimodal wearable system incorporating EEG, ECG, and accelerometry, and reported a median sensitivity of 100% with a FAR/24h of 8 [56]. When comparing these reports, our current study while applying the ADQA showed similar or improved performance. However, it's important to recognize that, unlike the algorithm in [51] and similar to the work in [54,56], our results were based on LOPO validation. Also, none of these studies included automatic data quality screening. On the other hand, when considering the algorithm's performance without applying the ADQA filtering, the median sensitivity drops to 33% with a slight increase of the median FAR/24 to 5.27, illustrating that the algorithm's performance was less reliable in this setting in comparison to the aforementioned studies [54–56]. Nevertheless, a FAR of 5.2 or 5.27 per 24 h remains relatively high compared to algorithms for detecting convulsive seizures (FAR/24 = 0.2) or non-convulsive seizures (FAR/24 = 1) [19,57].

### *Clinical and methodological implications*

In the context of data quality, applying an ADQA has demonstrated two key advantages, in addition to enhancing model transparency. Firstly, ADQA identified segments where the reliability of the algorithm's output could be compromised. This was evident in the difference in median sensitivity between the test group of 38 patients with data quality ≥ 75% and the test group of 47 patients with all the available data (Fig. 4A). Secondly, training the model on data that have met the quality threshold has improved algorithm robustness, reflected in the FAR/24 across all the test groups and sensitivity of the responders and patients with data quality ≥ 75% (Fig. 4A vs. 4B). This improvement highlights the benefits of data quality on the model's performance and that more data is not always better if quality is not controlled. Finally, the purpose of defining a responder group and comparing HR changes with non-responders was to explore inter-individual differences that may inform future patient stratification and potential use of the algorithm. In our post-hoc analysis, when stratifying the algorithm performance between responder and non-responder within the patient group with data quality ≥ 75%, we observed a large decrease in the median sensitivity from 100% to 8.3%. We then analyzed the HR change between the two groups and found that the responder group exhibited a significantly larger HR change during seizures compared to the non-responder group (Supplementary-Fig. 4). The smaller HR changes in non-responders suggest that this measurement could serve as a useful tool for identifying patients who would not benefit from the usage of the algorithm. However, since not all responders showed this pronounced HR increase, additional criteria will be needed to ensure more reliable identification of the target group. For quantifying the contribution of each HRV feature to the model's predictions, we applied SHAP analysis. The SHAP summary plot (Supplementary-Fig. 5) highlights the most influential features in detecting FIAS. Among these, Quantile 25, SD1, Variance, and SDNN demonstrated the highest impact. SD1, a marker of

**Table 3**
Overview of ECG-based seizure detection studies and performance metrics, compared with the present study.

| Study | Seizure Types Included | Sensitivity (%) | False Alarm Rate (/24 h) | No. of Patients |
|---|---|---|---|---|
| Jeppesen et al., 2019 [19] | Convulsive seizures (generalized tonic–clonic and focal-to-bilateral tonic–clonic) and focal seizures with impaired awareness and without impaired awareness; only responders' performance were reported | 93.1 (responders, all seizures); 90.5 (responders, non-convulsive seizures) | 1.0 (0.11/night) | 43 patients with seizures (53.5% classified as responders) |
| Jeppesen et al., 2020 [50] | Convulsive seizures (generalized tonic–clonic and focal-to-bilateral tonic–clonic) and focal seizures with impaired awareness and without impaired awareness | 100 (median) | 0.9/24 h (0.22/night) | 11 eligible patients where ictal heart rate increase >50 bpm) |
| Hegarty-Craver et al., 2021 [52] | Generalized tonic–clonic seizures and focal seizures | 92 (generalized); 54 (focal) | 1.0 | 18 children (+5 adults) |
| Jeppesen et al., 2025 [53] | Convulsive seizures (generalized tonic–clonic and focal-to-bilateral tonic–clonic) and focal seizures with impaired awareness and without impaired awareness | 100 (median) | 1.1/day; 0/night (median) | 17 eligible patients where ictal heart rate increase >50 bpm) |
| This study (total) | focal seizures with impaired awareness only | 33 (median) | 5.27 (median) | 47 patients |
| Subgroup with data quality ≥75% (this study) | focal seizures with impaired awareness only | 66.6 (median) | 5.2 (median) | 38 patients |
| Responder subgroup (this study) | focal seizures with impaired awareness only | 100 (median) | 4.3 (median) | 20 patients |

short-term HRV, contributed positively, suggesting localized, abrupt heart rate fluctuations during ictal events. In contrast, Variance and SDNN, both indicators of overall HRV and baseline autonomic tone, were negatively associated with seizure detection, pointing to a background of reduced variability. MeanNN, which measures the average interbeat interval, also emerged as an important feature. Lower MeanNN values (i.e., elevated heart rate) were positively associated with seizure detection, aligning with patterns of autonomic arousal or stress-like responses in preictal and ictal states. These patterns are consistent with previous findings on autonomic dysregulation in epilepsy, including elevated heart rate and suppressed HRV [15–17].

## Limitations

The main limitation of this work is that the training and validation were performed on the same dataset using the LOPO method. This approach may lead to overestimating the model's performance and ability to generalize to real-world scenarios. We acknowledge that external validation on an independent dataset is necessary to accurately assess the model's robustness and generalizability. However, it is also important to note that this study was specifically aimed at detecting a particular type of seizure. The strict inclusion criteria for the patient selection made this validation approach unavoidable. Furthermore, the relatively modest sample size in relation to the number of extracted HRV features introduces a potential risk of overfitting. Although several mitigation strategies were employed, including LOPO validation and regularization, this risk cannot be completely ruled out. Also, overlapping ECG windows were used to enable stable and continuous seizure detection, and event-based scoring ensured that each seizure was counted only once, however, this dense overlap may reduce the effective statistical independence of samples and affect uncertainty estimation. Another major limitation with direct implications for real-world deployment relates to data exclusion (approximately 100 h) by ADQA. While ADQA is critical for ensuring that only high-quality data are considered for seizure detection and in turn ensuring system reliability, it can lead to excluding segments containing seizures. This problem stems from how seizures manifest in physiological data and how the ADQA criteria are applied. During seizures, intense movements, irregular breathing patterns, and changes in posture could happen. These factors can lead to sensor misalignments, skin-electrode impedance fluctuations, and physiologically implausible heart rate intervals. This means that seizures could create conditions that increase the likelihood of poor signal quality. This bias highlights a limitation in the data acquisition conditions. In real-world use, such data loss may reduce effective monitoring time and limit algorithm availability during clinically relevant events. Therefore, while maintaining high standards for data quality is essential, minimizing the number of excluded data segments is equally important and requires methods that can contribute to improved data quality. Finally, we did not perform a systematic threshold sweep of the ADQA cutoff. As a result, we cannot quantify how alternative ADQA thresholds would trade off data retention (including potential seizure-containing segments) against model performance and robustness and false-positive detections were not analyzed qualitatively, limiting insight into their clinical relevance and underlying causes.

## Directions for future research

Even when an exploratory responder subgroup is considered, the long-term goal of this work is broad applicability of wearable seizure detection. Differences between responders and non-responders are intended to inform future methodological refinements (e.g., individualized baselines or adaptive thresholds) aimed at improving performance across a wider patient population. Achieving this broader applicability, however, requires addressing several practical and methodological challenges. One of the major challenges in developing seizure detection systems is data collection. This is mainly caused by the rarity of seizure occurrences in relation to the time spent at the monitoring unit. Thus, patients with less frequent seizures will always be underrepresented in the dataset. This underrepresentation poses a problem for developing and testing both patient-specific [58], and patient-independent models, as in this work, affecting the robustness and generalizability of these systems. Therefore, one possible approach

to these challenges could involve developing cloud-based and more energy-efficient solutions. This would reduce the number of bottlenecks in the data collection processes, such as data transfer and battery charge, allowing full utilization of the time spent by the patient at the monitoring unit. Another crucial factor was the data quality of sensors and artifacts arising from patients' movements. Enhancing signal by skin preparation, ensuring a secure fit of the sensor, providing continuous feedback to the user about the quality of the signal and prompting corrective action in case of signal quality degradation, would contribute to a more reliable system (Fig. 5). Also, as the 75% quality score was not optimized in this work as a hyperparameter. Future work should explore the potential of tuning the ADQA threshold to quantify the trade-off between data retention, false alarm burden, and assess algorithm generalizability across different ECG sensors to support broader real-world deployment. In addition to the quality of data, the nature of physiological responses during seizures could vary across patients and contexts. Although we did not observe a difference between the rate of seizure detection through the HR filter between night- and daytime events (Supplementary-Fig. 3), future work may benefit from incorporating relative HR increases from individual baselines. This could improve detection sensitivity, particularly in patients with blunted autonomic responses or during sleep-related seizures, where absolute thresholds may fail to capture subtle ictal changes. In the context of interpretability, SHAP values have shown the relative contributions of each feature to the model output, allowing a potential reduction of features in future work. Nevertheless, it would be still interesting to explore the effect of data quality on each feature output and, in turn on the balance between the model's true predictions and robustness. Finally, to enhance the generalizability, future research should aim to validate the model on independent external datasets from different clinical centers and patient populations to help mitigate the limitations posed by dataset-specific biases.

## Conclusions

In this study, we evaluated a wearable ECG–based FIAS detection algorithm incorporating ADQA within a controlled VEM environment. The algorithm's performance in patients whose data quality had met the defined criteria showed reasonable performance, with the highest detection performance observed in an exploratory responder subgroup. These results highlight the importance of integrating data quality assessment into wearable seizure detection systems and support further investigation of ECG-based approaches.

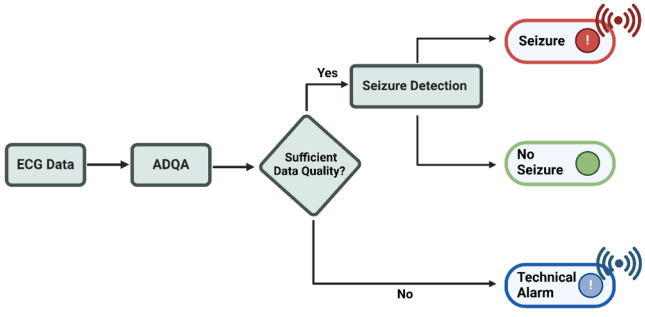


**Fig. 5.** Flowchart illustrating the seizure detection process with ADQA and its possible outcomes.ADQA analyzes ECG data and determines whether the data quality is sufficient for analysis. If quality is sufficient, seizure detection process is performed, leading to one of two outcomes: "Seizure Detected" (Red), where an alarm is triggered to alert the patient or caregiver, or "No Seizure Detected" (Green), confirming no seizure activity. If data quality is insufficient, a "Technical Alarm" (Blue) is triggered to notify the patient or caregiver, ensuring vigilance and preventing neglect of potential seizure events due to false negatives or noise. ADQA: Automatic Data Quality Assessment, ECG: Electrocardiogram.

## Data availability

The data sets analyzed during this study are available from the corresponding author on reasonable request.

## Ethics in publishing statement

I testify on behalf of all co-authors that our article submitted followed ethical principles in publishing.

Title: "Reliable Detection of Focal Onset Impaired Awareness Seizures in Patients with Epilepsy Using Wearable ECG: Development and Validation Study"

All authors agree that:

This research presents an accurate account of the work performed, all data presented are accurate and methodologies detailed enough to permit others to replicate the work.

This manuscript represents entirely original works and or if work and/or words of others have been used, that this has been appropriately cited or quoted and permission has been obtained where necessary.

This material has not been published in whole or in part elsewhere.

The manuscript is not currently being considered for publication in another journal.

That generative AI and AI-assisted technologies have not been utilized in the writing process or if used, disclosed in the manuscript the use of AI and AI-assisted technologies and a statement will appear in the published work.

That generative AI and AI-assisted technologies have not been used to create or alter images unless specifically used as part of the research design where such use must be described in a reproducible manner in the methods section.

All authors have been personally and actively involved in substantive work leading to the manuscript and will hold themselves jointly and individually responsible for its content.

Corresponding author's name: Yvonne Weber

Date: 19.05.2025

## Funding

This study has been funded by a grant from the Ministry of Economic Affairs, Industry, Climate Protection and Energy of the State of North Rhine-Westphalia (Funding No: 005-1911-0001. Project No: 1911LSI001) and the German Federal Ministry of Education and Research (BMBF, No 13GW0576C, Mondy).

## CRediT authorship contribution statement

**Mohamed Alhaskir:** Writing – original draft, Visualization, Methodology, Investigation, Formal analysis. **Ekaterina Kutafina:** Writing – review & editing, Supervision, Methodology. **Florian Linke:** Writing – review & editing, Data curation. **Florian P. Fischer:** Writing – review & editing, Resources, Data curation. **Elisabeth Schriewer:** Writing – review & editing, Resources. **Stephan Lauxmann:** Writing – review & editing, Resources, Data curation. **Kevin Klett:** Writing – review & editing, Resources. **Julian Hofmeister:** Writing – review & editing, Resources, Data curation. **Florian Lutz:** Writing – review & editing, Resources, Data curation. **Lukas Burow:** Writing – review & editing, Resources. **Michal Cicanic:** Writing – review & editing, Resources. **Sara Khosrawikatoli:** Writing – review & editing, Resources. **Stefan Wolking:** Writing – review & editing, Resources, Data curation. **Thomas Mayer:** Writing – review & editing, Resources. **Sandor Beniczky:** Writing – review & editing, Methodology. **Josua Kegele:** Writing – review & editing, Data curation. **Rainer Röhrig:** Writing – review & editing, Supervision, Methodology. **Henner Koch:** Writing – review & editing, Supervision, Resources, Data curation. **Yvonne Weber:** Writing – review & editing, Supervision, Resources, Project administration, Data curation.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Kevin Klett, Julian Hofmeister, Florian Lutz reports financial support was provided by Ministry of Economic Affairs, Industry, Climate Protection and Energy of the State of North Rhine-Westphalia. Yvonne Weber reports financial support was provided by German Federal Ministry of Education and Research. Kevin Klett, Julian Hofmeister and Florian Lutz reports a relationship with Monikit GmbH that includes: equity or stocks. K.K., J.H., and F.Lu are recent stakeholders of the Monikit GmbH, but not at the time the study was planned and conducted. Wearable ECG devices used in the study were leased from Monikit GmbH. K.K., J.H., F.Lu. and any other person of the Monikit GmbH had no influence on data analysis and interpretation. The other authors state that they have no conflict of interest. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

K.K., J.H., and F.Lu are recent stakeholders of the Monikit GmbH, but not at the time the study was planned and conducted. Wearable ECG devices used in the study were leased from Monikit GmbH. K.K., J.H., F. Lu. and any other person of the Monikit GmbH had no influence on data analysis and interpretation. The other authors state that they have no conflict of interest. We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

## Acknowledgments

We thank the patients for participating in the study and would like to acknowledge the University Hospital Aachen, University Hospital Tübingen, and Epilepsy Center Kleinwachau staff for helping with the special settings and recording the patients. We thank Alexander Ivanets for double programming the statistical analysis. We also appreciate the Monikit group for providing the ECG devices. Figs. 1 and 2 were created with BioRender.com.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2026.109295.

## References

[1] GBD, Epilepsy collaborators. Global, regional, and national burden of epilepsy, 1990-2016: a systematic analysis for the global burden of disease study, Lancet Neurol. 18 (4) (2016) 357–375, https://doi.org/10.1016/S1474-4422(18)30454-X, 2019.

[2] D.J. Thurman, E. Beghi, C.E. Begley, A.T. Berg, J.R. Buchhalter, D. Ding, D. C. Hesdorffer, W.A. Hauser, L. Kazis, R. Kobau, B. Kroner, Standards for epidemiologic studies and surveillance of epilepsy, Epilepsia 52 (2011) 2–6.

[3] C.E. Elger, C. Hoppe, Diagnostic challenges in epilepsy: seizure under-reporting and seizure detection, Lancet Neurol. 17 (3) (2018) 279–288, https://doi.org/10.1016/S1474-4422(18)30038-3.

[4] J.J. Shih, N.B. Fountain, S.T. Herman, et al., Indications and methodology for video-electroencephalographic studies in the epilepsy monitoring unit, Epilepsia 59 (1) (2018) 27–36, https://doi.org/10.1111/epi.13938.

[5] R.S. Fisher, D.E. Blum, B. DiVentura, et al., Seizure diaries for clinical research and practice: limitations and future prospects, Epilepsy Behav. 24 (3) (2012) 304–310, https://doi.org/10.1016/j.yebeh.2012.04.128.

[6] C. Hoppe, A. Poepel, CE. Elger, Epilepsy: accuracy of patient seizure counts, Arch. Neurol. 64 (11) (2007) 1595–1599, https://doi.org/10.1001/archneur.64.11.1595.

[7] B. Bases, S. Barnard, J.A. French, et al., Impact of seizures while driving prior to diagnosis in people with focal epilepsy: motor vehicle accidents and time to diagnosis, Neurology 101 (13) (2023) e1370–e1375, https://doi.org/10.1212/WNL.0000000000207464.

[8] Y. Paul, Various epileptic seizure detection techniques using biomedical signals: a review, Brain Inf. 5 (2) (2018 Jul 10) 6, https://doi.org/10.1186/s40708-018-0084-z. PMID: 29987692; PMCID: PMC6170938.

[9] E. Bruno, P.F. Viana, M.R. Sperling, MP. Richardson, Seizure detection at home: do devices on the market match the needs of people living with epilepsy and their caregivers? Epilepsia 61 (2020 Nov) https://doi.org/10.1111/epi.16521. Suppl 1: S11-S24Epub 2020 May 9. PMID: 32385909.

[10] R.H.C. Lazeron, R.D. Thijs, J. Arends, T. Gutter, P. Cluitmans, J. Van Dijk, F.I. Y. Tan, W. Hofstra, C. Donjacour, F. Leijten, Dutch tele-epilepsy consortium. multimodal nocturnal seizure detection: do we need to adapt algorithms for children? Epilepsia Open. 7 (3) (2022 Sep) 406–413, https://doi.org/10.1002/epi4.12618. Epub 2022 Jul 21. PMID: 35666848; PMCID: PMC9436288.

[11] S. Böttcher, E. Bruno, N.V. Manyakov, N. Epitashvili, K. Claes, M. Glasstetter, S. Thorpe, S. Lees, M. Dümpelmann, K. Van Laerhoven, M.P. Richardson, A. Schulze-Bonhage, R.-C.N.S. Consortium, Detecting tonic-clonic seizures in multimodal biosignal data from wearables: methodology design and validation, JMIR. MHealth UHealth 9 (11) (2021 Nov 19) e27674, https://doi.org/10.2196/27674. PMID: 34806993; PMCID: PMC8663471.

[12] S. Beniczky, I. Conradsen, O. Henning, M. Fabricius, P. Wolf, Automated real-time detection of tonic-clonic seizures using a wearable EMG device, Neurology 90 (5) (2018 Jan 30) e428–e434, https://doi.org/10.1212/WNL.0000000000004893. Epub 2018 Jan 5. PMID: 29305441; PMCID: PMC5791791.

[13] Y. Ghulaxe, A. Joshi, J. Chavada, S. Huse, B. Kalbande, PP. Sarda, Understanding focal seizures in adults: a comprehensive review, Cureus 15 (11) (2023 Nov 2) e48173, https://doi.org/10.7759/cureus.48173. PMID: 38046728; PMCID: PMC10693312.

[14] M. Zijlmans, D. Flanagan, J. Gotman, Heart rate changes and ECG abnormalities during epileptic seizures: prevalence and definition of an objective clinical sign, Epilepsia 43 (8) (2002 Aug) 847–854, https://doi.org/10.1046/j.1528-1157.2002.37801.x. PMID: 12181003.

[15] M.Z. Poh, T. Loddenkemper, C. Reinsberger, N.C. Swenson, S. Goyal, J.R. Madsen, RW. Picard, Autonomic changes with seizures correlate with postictal EEG suppression, Neurology 78 (23) (2012 Jun 5) 1868–1876, https://doi.org/10.1212/WNL.0b013e318258f7f1. Epub 2012 Apr 25. PMID: 22539579; PMCID: PMC3369522.

[16] W.J. van Elmpt, T.M. Nijsen, P.A. Griep, JB. Arends, A model of heart rate changes to detect seizures in severe epilepsy, Seizure 15 (6) (2006 Sep) 366–375, https://doi.org/10.1016/j.seizure.2006.03.005. Epub 2006 Jul 7. PMID: 16828317.

[17] H. Evrengül, H. Tanriverdi, D. Dursunoglu, A. Kaftan, O. Kuru, U. Unlu, M. Kilic, Time and frequency domain analyses of heart rate variability in patients with epilepsy, Epilepsy Res. 63 (2-3) (2005 Feb) 131–139, https://doi.org/10.1016/j.eplepsyres.2005.02.001. PMID: 15777689.

[18] J. Jeppesen, S. Beniczky, P. Johansen, P. Sidenius, A. Fuglsang-Frederiksen, Detection of epileptic seizures with a modified heart rate variability algorithm based on Lorenz plot, Seizure 24 (2015 Jan) 1–7, https://doi.org/10.1016/j.seizure.2014.11.004. Epub 2014 Nov 15. PMID: 25564311.

[19] J. Jeppesen, A. Fuglsang-Frederiksen, P. Johansen, J. Christensen, S. Wüstenhagen, H. Tankisi, E. Qerama, A. Hess, S. Beniczky, Seizure detection based on heart rate variability using a wearable electrocardiography device, Epilepsia 60 (10) (2019 Oct) 2105–2113, https://doi.org/10.1111/epi.16343. Epub 2019 Sep 20. PMID: 31538347.

[20] K. Fujiwara, M. Miyajima, T. Yamakawa, E. Abe, Y. Suzuki, Y. Sawada, M. Kano, T. Maehara, K. Ohta, T. Sasai-Sakuma, T. Sasano, M. Matsuura, E. Matsushima, Epileptic seizure prediction based on multivariate statistical process control of heart rate variability features, IEEE Trans. Biomed. Eng. 63 (6) (2016 Jun) 1321–1332, https://doi.org/10.1109/TBME.2015.2512276. Epub 2015 Dec 24. PMID: 26841385.

[21] Yamakawa T., Miyajima M., Fujiwara K., Kano M., Suzuki Y., Watanabe Y., Watanabe S., Hoshida T., Inaji M., Maehara T. Wearable epileptic seizure prediction system with machine-learning-based anomaly detection of heart rate variability. sensors. 2020 Jul 17;20(14):3987. doi: 10.3390/s20143987. PMID: 32709064; PMCID: PMC7411877.

[22] R. Ode, K. Fujiwara, M. Miyajima, T. Yamakawa, M. Kano, K. Jin, N. Nakasato, Y. Sawai, T. Hoshida, M. Iwasaki, Y. Murata, Development of an epileptic seizure prediction algorithm using R–R intervals with self-attentive autoencoder, Artif. Life Robot. 28 (2) (2023 May) 403–409.

[23] P. Meritam Larsen, S. Beniczky, Non-electroencephalogram-based seizure detection devices: state of the art and future perspectives, Epilepsy Behav. 148 (2023 Nov) 109486, https://doi.org/10.1016/j.yebeh.2023.109486. Epub 2023 Oct 17. PMID: 37857030.

[24] Asadi-Pooya A.A., Brigo F., Lattanzi S., Blumcke I. Adult epilepsy. Lancet. 2023 Jul 29;402(10399):412-424. doi: 10.1016/S0140-6736(23)01048-6. Epub 2023 Jul 14. PMID: 37459868.

[25] E. Beghi, The epidemiology of epilepsy, Neuroepidemiology 54 (2) (2020) 185–191, https://doi.org/10.1159/000503831. Epub 2019 Dec 18. PMID: 31852003.

[26] M. Nasseri, E. Nurse, M. Glasstetter, S. Böttcher, N.M. Gregg, A. Laks Nandakumar, B. Joseph, T. Pal Attia, P.F. Viana, E. Bruno, A. Biondi, M. Cook, G.A. Worrell, A. Schulze-Bonhage, M. Dümpelmann, D.R. Freestone, M.P. Richardson, B. H. Brinkmann, Signal quality and patient experience with wearable devices for epilepsy management, Epilepsia 61 (1) (2020 Nov) S25–S35, https://doi.org/10.1111/epi.16527. SupplEpub 2020 Jun 4. PMID: 32497269.

[27] S. Böttcher, S. Vieluf, E. Bruno, B. Joseph, N. Epitashvili, A. Biondi, N. Zabler, M. Glasstetter, M. Dümpelmann, K. Van Laerhoven, M. Nasseri, B.H. Brinkman, M. P. Richardson, A. Schulze-Bonhage, T. Loddenkemper, Data quality evaluation in wearable monitoring, Sci. Rep. 12 (1) (2022 Dec 10) 21412, https://doi.org/10.1038/s41598-022-25949-x. PMID: 36496546; PMCID: PMC9741649.

[28] T.P. Karnowski, D. Aykac, L. Giancardo, Y. Li, T. Nichols, T.K.W. Jr, E. Chaum, Automatic detection of retina disease: robustness to image quality and localization

of anatomy structure, Annu Int. Conf. IEEe Eng. Med. Biol. Soc. 2011 (2011) 5959–5964, https://doi.org/10.1109/IEMBS.2011.6091473. PMID: 22255697.
[29] H. Chen, J. Chen, J. Ding, Data evaluation and enhancement for quality improvement of machine learning, IEEe Trans. Reliab. (2021) 1–17, https://doi.org/10.1109/TR.2021.3070863.
[30] D. Schwabe, K. Becker, M. Seyferth, A. Klaß, T. Schaeffter, The METRIC-framework for assessing data quality for trustworthy AI in medicine: a systematic review, NPJ. Digit. Med. 7 (1) (2024 Aug 3) 203, https://doi.org/10.1038/s41746-024-01196-4. PMID: 39097662; PMCID: PMC11297942.
[31] J. Wiens, S. Saria, M. Sendak, M. Ghassemi, V.X. Liu, F. Doshi-Velez, K. Jung, K. Heller, D. Kale, M. Saeed, P.N. Ossorio, S. Thadaney-Israni, A. Goldenberg, Do no harm: a roadmap for responsible machine learning for health care, Nat. Med. 25 (9) (2019 Sep) 1337–1340, https://doi.org/10.1038/s41591-019-0548-6. Epub 2019 Aug 19. Erratum in: Nat Med. 2019 Oct;25(10):1627. PMID: 31427808.
[32] Fehr J., Citro B., Malpani R., Lippert C., Madai VI. A trustworthy AI reality-check: the lack of transparency of artificial intelligence products in healthcare. Front Digit Health. 2024 Feb 20;6:1267290. doi: 10.3389/fdgth.2024.1267290. PMID: 38455991; PMCID: PMC10919164.
[33] S. Beniczky, P. Ryvlin, Standards for testing and clinical validation of seizure detection devices, Epilepsia 59 (1) (2018 Jun) 9–13, https://doi.org/10.1111/epi.14049. SupplPMID: 29873827.
[34] P. Hamilton, WJ. Tompkins, Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database, IEEE Trans. Biomed. Eng. 33 (12) (1986) 1157–1165, https://doi.org/10.1109/TBME.1986.325695.
[35] E. Escobar-Linero, F. Luna-Perejon, L. Muñoz-Saavedra, J.L. Sevillano, M. Domínguez-Morales, On the feature extraction process in machine learning. an experimental study about guided versus non-guided process in falling detection systems, Eng. Appl. Artif. Intell. 114 (2022) 105170, https://doi.org/10.1016/j.engappai.2022.105170.
[36] P. Meritam Larsen, S. Wüstenhagen, D. Terney, E. Gardella, H. Aurlien, S. Beniczky, Duration of epileptic seizure types: a data-driven approach, Epilepsia 64 (2) (2023 Feb) 469–478, https://doi.org/10.1111/epi.17492. Epub 2023 Jan 3. PMID: 36597206; PMCID: PMC10107943.
[37] M.L. Munoz, A. van Roon, H. Riese, C. Thio, E. Oostenbroek, I. Westrik, E.J. de Geus, R. Gansevoort, J. Lefrandt, I.M. Nolte, H. Snieder, Validity of (Ultra-)short recordings for heart rate variability measurements, PLoS. One 10 (9) (2015 Sep 28) e0138921, https://doi.org/10.1371/journal.pone.0138921. PMID: 26414314; PMCID: PMC4586373.
[38] C. Orphanidou, T. Bonnici, P. Charlton, D. Clifton, D. Vallance, L. Tarassenko, Signal-quality indices for the electrocardiogram and photoplethysmogram: derivation and applications to wireless monitoring, IEEE J. Biomed. Health Inf. 19 (3) (2015 May) 832–838, https://doi.org/10.1109/JBHI.2014.2338351. Epub 2014 Jul 23. PMID: 25069129.
[39] W. Zong, T. Heldt, G.B. Moody, R.G. Mark, An open-source algorithm to detect onset of arterial blood pressure pulses, Comput. Cardiol. 2003 (2003) 259–262, https://doi.org/10.1109/CIC.2003.1291140.
[40] Q. Li, R.G. Mark, GD. Clifford, Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter, Physiol. Meas. 29 (1) (2008) 15–32, https://doi.org/10.1088/0967-3334/29/1/002.
[41] S. Beniczky, S. Wiebe, J. Jeppesen, W.O. Tatum, M. Brazdil, Y. Wang, S.T. Herman, P. Ryvlin, Automated seizure detection using wearable devices: a clinical practice guideline of the international league against epilepsy and the international federation of clinical neurophysiology, Epilepsia 62 (3) (2021 Mar) 632–646, https://doi.org/10.1111/epi.16818. PMID: 33666944.
[42] S. Beniczky, P. Karoly, E. Nurse, P. Ryvlin, M. Cook, Machine learning and wearable devices of the future, Epilepsia 62 (2) (2021 Mar) S116–S124, https://doi.org/10.1111/epi.16555. SupplEpub 2020 Jul 26. PMID: 32712958.
[43] J. Tang, R. El Atrache, S. Yu, U. Asif, M. Jackson, S. Roy, M. Mirmomeni, S. Cantley, T. Sheehan, S. Schubach, C. Ufongene, S. Vieluf, C. Meisel, S. Harrer, T. Loddenkemper, Seizure detection using wearable sensors and machine learning: setting a benchmark, Epilepsia 62 (8) (2021 Aug) 1807–1819, https://doi.org/10.1111/epi.16967. Epub 2021 Jul 15. PMID: 34268728; PMCID: PMC8457135.
[44] S. Yu, R. El Atrache, J. Tang, M. Jackson, A. Makarucha, S. Cantley, T. Sheehan, S. Vieluf, B. Zhang, J.L. Rogers, I. Mareels, S. Harrer, T. Loddenkemper, Artificial intelligence-enhanced epileptic seizure detection by wearables, Epilepsia 64 (12) (2023 Dec) 3213–3226, https://doi.org/10.1111/epi.17774. Epub 2023 Oct 25. PMID: 37715325.
[45] Vaswani A., Shazeer N., Parmar N., et al. Attention is all you need. in: advances in neural information processing systems. Vol 30; 2017.
[46] S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y.X. Wang, X. Yan, Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting, Adv. Neural Inf. Process. Syst. 32 (2019).
[47] Zerveas G., Jayaraman S., Patel D., Bhamidipaty A., Eickhoff C. A transformer-based framework for multivariate time series representation learning. InProceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining 2021 Aug 14 (pp. 2114-2124).
[48] V. Shah, M. Golmohammadi, I. Obeid, J. Picone, Objective evaluation metrics for automatic classification of EEG events, Biomed. Signal. Process.: Innov. Appl. (2021) 223–255.
[49] Lundberg S. A unified approach to interpreting model predictions. arXiv preprint arXiv:1705.07874. 2017.
[50] J. Jeppesen, A. Fuglsang-Frederiksen, P. Johansen, et al., Seizure detection using heart rate variability: a prospective validation study, Epilepsia 61 (Suppl 1) (2020) S41–S46, https://doi.org/10.1111/epi.16511.
[51] D. Cogan, J. Birjandtalab, M. Nourani, et al., Multi-biosignal analysis for epileptic seizure monitoring, Int. J. Neural Syst. 27 (1) (2017) 1650031, https://doi.org/10.1142/S0129065716500313.
[52] M. Hegarty-Craver, B.L. Kroner, A. Bumbut, et al., Cardiac-based detection of seizures in children with epilepsy, Epilepsy Behav. 122 (2021) 108129, https://doi.org/10.1016/j.yebeh.2021.108129.
[53] Jeppesen J., Christensen J., Ahrenfeldt Petersen O., Fenger S., Armand Larsen S., Wüstenhagen S., Wagner S.R., Johansen P., Beniczky S.. Seizure detection using wearable electrocardiogram connected to a smartphone: a phase 3 clinical validation study. EBioMedicine. 2025 Oct;120:105952. doi: 10.1016/j.ebiom.2025.105952. Epub 2025 Sep 29. PMID: 41027311; PMCID: PMC12516532.
[54] K. Vandecasteele, T. De Cooman, J. Dan, E. Cleeren, S. Van Huffel, B. Hunyadi, W. Van Paesschen, Visual seizure annotation and automated seizure detection using behind-the-ear electroencephalographic channels, Epilepsia 61 (4) (2020 Apr) 766–775, https://doi.org/10.1111/epi.16470. Epub 2020 Mar 11. PMID: 32160324; PMCID: PMC7217054.
[55] J. Macea, M. Bhagubai, V. Broux, et al., In-hospital and home-based long-term monitoring of focal epilepsy with a wearable electroencephalographic device: diagnostic yield and user experience, Epilepsia 64 (4) (2023) 937–950, https://doi.org/10.1111/epi.17517.
[56] J. Munch Nielsen, I.C. Zibrandtsen, P. Masulli, T. Lykke Sørensen, T.S. Andersen, T. Wesenberg Kjær, Towards a wearable multi-modal seizure detection system in epilepsy: a pilot study, Clin. Neurophysiol. 136 (2022 Apr) 40–48, https://doi.org/10.1016/j.clinph.2022.01.005. Epub 2022 Jan 25. PMID: 35131637.
[57] F. Onorati, G. Regalia, C. Caborni, et al., Multicenter clinical assessment of improved wearable multimodal convulsive seizure detectors, Epilepsia 58 (11) (2017) 1870–1879, https://doi.org/10.1111/epi.13872.
[58] J. Jeppesen, J. Christensen, P. Johansen, S. Beniczky, Personalized seizure detection using logistic regression machine learning based on wearable ECG-monitoring device, Seizure 107 (2023) 155–161, https://doi.org/10.1016/j.seizure.2023.04.012.