Clinical Neurophysiology 142 (2022) 86–93

Contents lists available at ScienceDirect

Clinical Neurophysiology

journal homepage: www.elsevier.com/locate/clinph

Detecting temporal lobe seizures in ultra long-term subcutaneous EEG using algorithm-based data reduction

Line S. Remvig [a,*], Jonas Duun-Henriksen [a], Franz Fürbass [b], Manfred Hartmann [b], Pedro F. Viana [c,d], Anne Mette Kappel Overby [e], Sigge Weisdorf [f], Mark P. Richardson [c], Sándor Beniczky [e,g], Troels W. Kjaer [f,h]

[a] UNEEG Medical A/S, Borupvang 2, DK-3450 Allerød, Denmark
[b] Center for Health & Bioresources, AIT Austrian Institute of Technology GmbH, Giefinggasse 4, 1210 Vienna, Austria
[c] Department of Basic & Clinical Neuroscience, Maurice Wohl Clinical Neuroscience Institute, King's College London, Denmark Hill, London, UK
[d] Faculty of Medicine, University of Lisbon, Av. Prof. Egas Moniz MB, 1649-028 Lisboa, Portugal
[e] The Danish Epilepsy Centre, Filadelfia, Kolonivej 1, 4293 Dianalund, Denmark
[f] Center of Neurophysiology, Department Neurology, Zealand University Hospital, Sygehusvej 10, DK-4000 Roskilde, Denmark
[g] Aarhus University and Aarhus University Hospital, Palle Juul-Jensens Blvd 99, DK-8200 Aarhus, Denmark
[h] Department of Clinical Medicine, University of Copenhagen, Blegdamsvej 3b, DK.2200 Copenhagen, Denmark

ARTICLE INFO

Article history:
Accepted 3 July 2022
Available online 8 August 2022

Keywords:
Epilepsy
Seizure detection
Long-term monitoring
Subcutaneous EEG
Outpatient monitoring

HIGHLIGHTS

- Ultra long-term subcutaneous EEG offers a novel option for the recording of electrographic epileptic seizures in everyday life.
- A semi-automatic seizure detection process is proposed to limit the time spent on review to periods of potential seizure activity.
- The algorithm of the semi-automatic detection process had a sensitivity of 86% and a false detection rate of 2.4 per 24 hours.

ABSTRACT

Objective: Ultra long-term monitoring with subcutaneous EEG (sqEEG) offers objective outpatient recording of electrographic seizures as an alternative to self-reported epileptic seizure diaries. This methodology requires an algorithm-based automatic seizure detection to indicate periods of potential seizure activity to reduce the time spent on visual review. The objective of this study was to evaluate the performance of a sqEEG-based automatic seizure detection algorithm.
Methods: A multicenter cohort of subjects using sqEEG were analyzed, including nine people with epilepsy (PWE) and 12 healthy subjects, recording a total of 965 days. The automatic seizure detections of a deep-neural-network algorithm were compared to annotations from three human experts.
Results: Data reduction ratios were 99.6% in PWE and 99.9% in the control group. The cross-PWE sensitivity was 86% (median 80%, range 69–100% when PWE were evaluated individually), and the corresponding median false detection rate was 2.4 detections per 24 hours (range: 2.0–13.0).
Conclusions: Our findings demonstrated that step one in a sqEEG-based semi-automatic seizure detection/review process can be performed with high sensitivity and clinically applicable specificity.
Significance: Ultra long-term sqEEG bears the potential of improving objective seizure quantification.

1. Introduction

Seizures are the primary symptom of epilepsy. Clinical management aims to avoid seizures and reduce complications. Repetitive seizures present significant challenges to people with refractory epilepsy entailing lifestyle limitations and potential

Abbreviations: sqEEG, subcutaneous encephalography; PWE, people with epilepsy.

* Corresponding author at: UNEEG medical A/S, Borupvang 2, DK-3450 AllerAllerød, Denmark.
E-mail address: lsre@uneeg.com (L.S. Remvig).

https://doi.org/10.1016/j.clinph.2022.07.504
1388-2457/

seizure-related injuries. Accurate seizure documentation is paramount to quantify the burden of people with epilepsy (PWE), but the current standard assessment using seizure diaries based on seizure self-reports by the PWE or by the caregivers have been shown to be unreliable with an average accuracy below 50 percent (Blachut et al. 2017; Elger and Hoppe 2018; Fisher et al. 2012). As seizure quantification constitutes the basis of clinical management of refractory epilepsy and the evaluation of new therapies, more accurate methods are needed. Objective and accurate seizure monitoring using mobile devices during everyday life activities is likely to become a game changer in managing patients with epilepsy.

Currently marketed seizure detection devices typically take advantage of the measurement of motor activity (reflected in video, accelerometry and electromyography) or changes in autonomic measures (reflected in heart rate parameters and electrodermal activity) (Bruno et al. 2020). However, the devices on the market are restricted to detection of convulsive seizures, whereas the detection of focal seizures without major motor components remains a clear gap (Bruno et al. 2020).

EEG shows well-described characteristics associated with epileptic seizures and constitutes a cornerstone in the diagnostic process of any epileptic syndrome. Therefore, multiple seizure detection algorithms have been developed for EEG (Fürbass et al. 2015, 2017; Baumgartner and Koren 2018) – in particular for scalp EEG. While scalp electrodes are not feasible for long-term recordings, minimally invasive subcutaneous EEG (sqEEG) recording devices provide reliable ultra long-term continuous monitoring in patients' habitual environment and during everyday life activities (Duun-Henriksen et al. 2020). Ultra long-term outpatient sqEEG monitoring has been demonstrated to be feasible and well-tolerated for nine PWE with temporal lobe epilepsy (Weisdorf et al. 2019). No serious adverse events were reported and none of the PWE felt constrained in their ability to carry out jobs and leisure activities. In addition, the ability to detect electrographic seizures from sqEEG in real-life settings was demonstrated. Furthermore, the sqEEG signal has been demonstrated to be of high quality and highly stable throughout months of recording (Viana et al. 2021).

Ultra long-term outpatient sqEEG monitoring accommodates needs where traditional routine video-EEG and hospitalization at epilepsy monitoring units are inadequate. This includes recording of rare seizures (e.g. frequency < 1/month) and the ability to capture the temporal fluctuations in seizure patterns present in most PWE. It is well-known that the timing of seizure occurrences is not random – rather both circadian and multi-day cyclic seizure timing patterns exist (Karoly et al. 2018; Baud et al. 2018). Ultra long-term recordings could improve the monitoring of treatment effects by eliminating the effect of high variability of seizure occurrence, when the seizure cycles are known (Goldenholz et al. 2017).

Visual assessment by human experts, to annotate epileptic seizures in ultra long-term sqEEG recordings lasting several months is not feasible as it is extremely time consuming. Therefore, automated seizure detection algorithms are required for clinical implementation, to decrease the huge workload. A hybrid (semi-automatic) approach where an automatic detector marks candidates for electrographic seizures and then an EEG expert validates detections by visual inspection using the sqEEG, could combine the high sensitivity from the algorithm with the high specificity from the expert, yet, decreasing considerably the workload. The UNEEG™ EpiSight Analyzer software (hereafter referred to as "EpiSight") is a visualization tool with a computational algorithm that automatically detects potential electrographic epileptic seizures and reduces the required amount of sqEEG data to be reviewed to these selected events.

Our goal was to evaluate the EpiSight seizure detection algorithm. Our performance assessment aimed to answer two questions of clinical relevance: 1) what are the detection sensitivities and false detections rates? and 2) is data reduced to an amount that is clinically applicable for visual review, in a subsequent step? For ultra long-term sqEEG to become a true game changer in epilepsy management by delivering accurate objective seizure monitoring in real-life settings, the key process of automatic seizure detection/review is of utmost importance. Comparisons will be carried out in order justify the minimally invasive semi-automatic sqEEG-based review process over the self-reported diaries.

2. Methods

2.1. Data

Two separate datasets were included in the current work, both of which were recorded with a preliminary version of the novel, minimally invasive sqEEG solution for ultra long-term outpatient recording (24/7 EEG™ SubQ, UNEEG medical A/S, Allerød, Denmark). The "epilepsy" dataset included nine PWE with temporal lobe epilepsy. All PWE were medically refractory – except from one, who was only recently diagnosed with epilepsy. The sqEEG solution was applied for epileptic seizure monitoring throughout 2–3 months, resulting in a total of 490 days of recorded sqEEG in the epilepsy dataset. All PWE were using manual seizure diaries with a precision limited to date of seizure events. For a detailed account on the study design, data collection procedures and demographics of the PWE, please refer to (Weisdorf et al. 2019) (clinicaltrials.gov NCT02946151). Information on seizure onset zone, semiology, and counts for each PWE is included in Table 1, which holds the main results of the current work. Mostly, non-convulsive seizures were represented, but also a few convulsive seizures (focal to bilateral tonic-clonic seizures).

In addition, a "normal control" dataset including 12 healthy subjects was collected. This constituted a total of 475 real-life recording days of previously unpublished sqEEG data (clinicaltrials.gov NCT02402153). Each study was approved by their regional committee of science ethics and participants provided written informed consent.

The lead of the sqEEG device providing two bipolar EEG channels (3-contact electrode) was implanted unilaterally over the focus on the temporal lobe of the PWE, while it was pointing from behind the ear in a vertical direction towards vertex for the healthy subjects (see Fig. 1). The sqEEG device records data at 207 Hz and has the following filter characteristic: 0.5–48 Hz equiripple FIR bandpass filter with a sidelobe attenuation of 40 dB and passband ripple of < 0.1 dB.

2.2. Validation strategy – the gold standard

The 490 days of sqEEG recording of the epilepsy dataset was reviewed and labelled manually to establish a gold standard of electrographic epileptic seizures to be able to evaluate the performance statistics of the automatic detection algorithm. In a first step, three independent clinical experts from three different institutions reviewed and labelled the dataset with three different labelling approaches, each of which will be described below.

Zealand University Hospital: sqEEG data was reviewed based on 10-min time–frequency epochs as described in detail before (Weisdorf et al. 2019). Whenever the existence of a potential seizure pattern was identified, the sqEEG was reviewed in the time domain to confirm or reject the event of a seizure. Prior to the review process, all available previous scalp EEG recordings and/or reports for each PWE were thoroughly reviewed to establish one

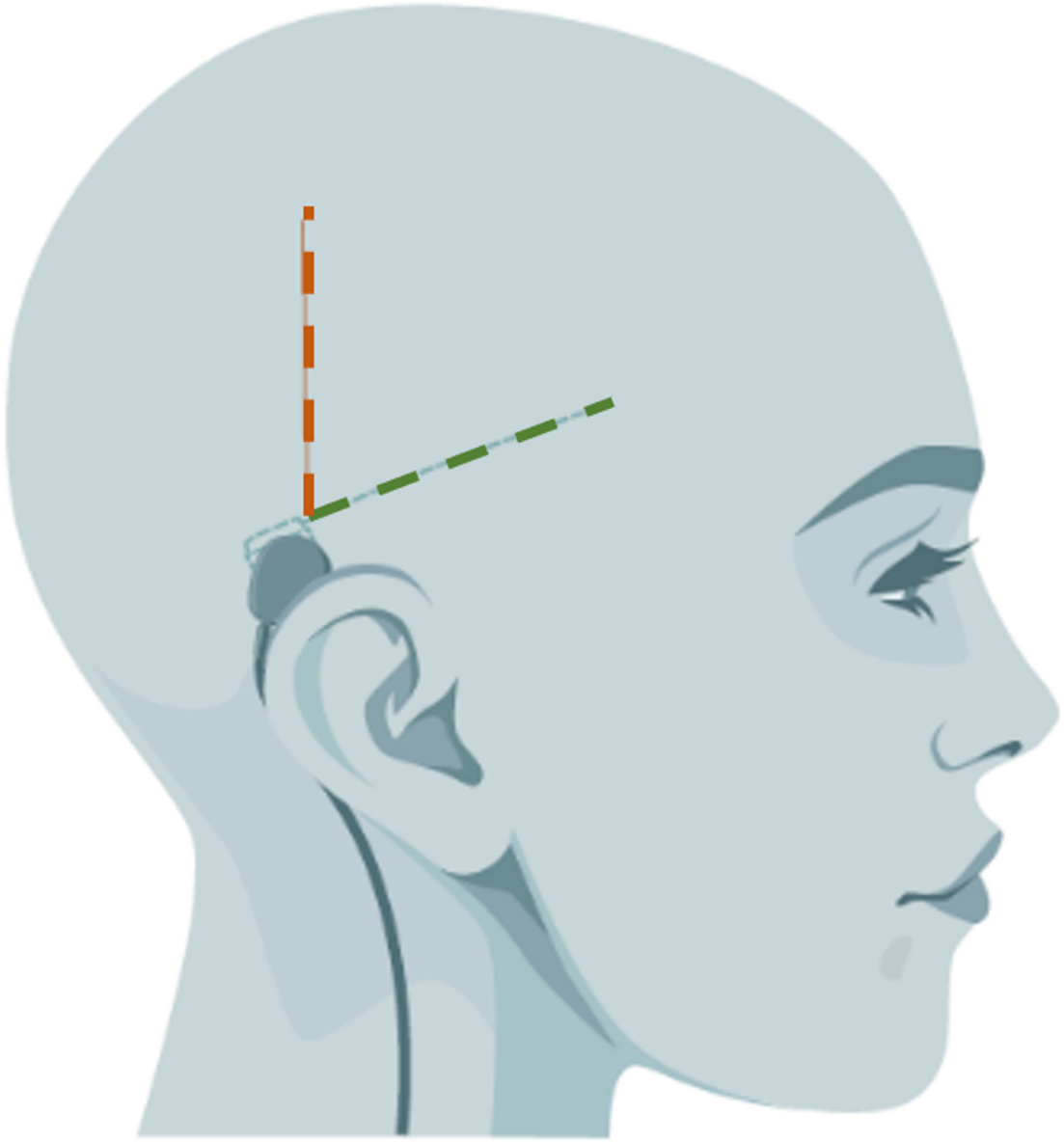

Fig. 1. Illustration of the placement of the implanted recording lead. The dashed green line represents the lead inserted subcutaneously in an almost horizontal direction, recording from the temporal lobe (people with epilepsy). The dashed brown line represents the direction of the recording lead towards vertex (healthy subjects).

or more personal spectrographic seizure signatures. The entire review was performed by a physician with experience in visual EEG review (SW) using Nervus EEG Reader v. 5.95 (Pleasanton, USA). All seizure annotations and events of doubt underwent secondary review by a board-certified clinical neurophysiologist (TWK), who made the final decision.

The Danish Epilepsy Centre, Filadelfia: sqEEG data was reviewed by visual inspection in the time-domain by a certified neurophysiology technician with experience in visual EEG review (AMKO). A previous version of the EpiSight software (v. 1.9.2) was applied for the task, but without enabling the automated seizure detection algorithm for data reduction. All seizure annotations and events of doubt underwent secondary review by a board-certified clinical neurophysiologist (SB), who made the final decision.

King's College London: A previous version of the EpiSight software (v. 1.9.2) was applied as prescribed, reducing the sqEEG data to be reviewed manually for each PWE to the events automatically detected by the seizure detection algorithm (clinically validated detector described in (Fürbass et al. 2015, 2017). The reviewer either discarded or accepted the potential seizure annotations. This was performed by a trained neurologist with experience in EEG review (PFV). This review process constitutes the use case of the UNEEG solution.

In a second step, all three review teams were asked to review all sqEEG patterns, labelled with a seizure annotation by either of the remaining teams, which they had not labelled and/or seen themselves, see Fig. 2. Reviewers were asked to either accept or discard the potential electrographic seizures labelled by a first review team. In this way, all review teams were presented with all potential seizures disregarding the specific review process they performed in the first step. Thereby, the base for a majority decision was established to determine the gold standard of electrographic seizures to which the automatic algorithm could be evaluated against. As three review teams contributed, at least two votes were required to classify an electrographic seizure.

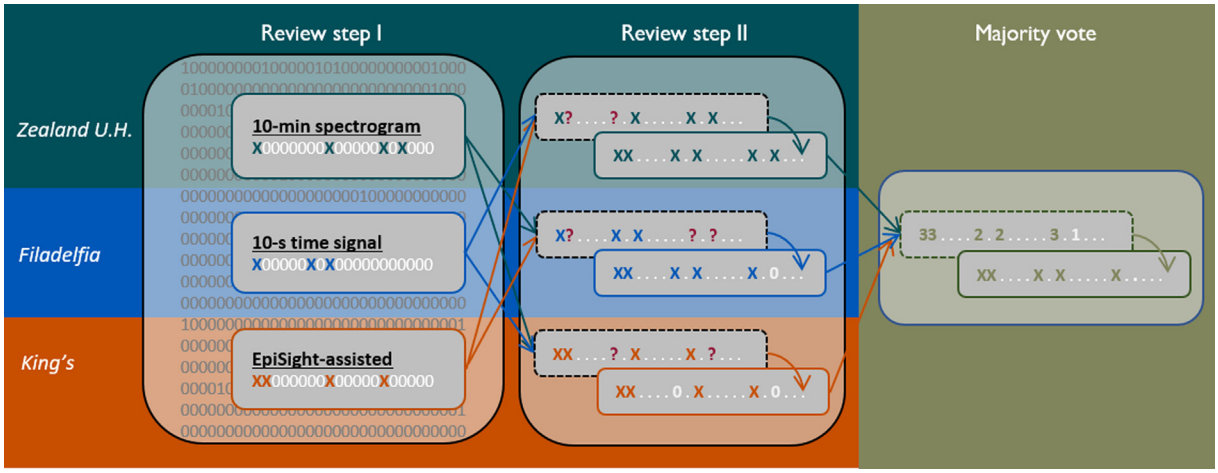

Fig. 2. Illustration of the stepwise review and labelling process. Three independent clinical experts from three different institutions reviewed and labelled data. Review step I: Three different approaches of labelling, each reducing the large data amount to several seizure annotations represented by "X". Each institution holds a color code (green–blue, blue, orange) and the sequences of 0's and X's are aligned in time. The seizure annotations (X's) from each institution are passed to the next step. Review step II: Each reviewer is presented unseen and/or "extra" seizure annotations from the two other reviewers (red "?" in black dashed outlined datasets). Each reviewer either discards (a light gray "0") or accepts the extra seizures (a color coded "X"), resulting in three labelled datasets (solid outlined datasets). Majority vote: the three labelled data sequences are "summed" (green, dashed outlined), and whenever at least two reviewers agree on a seizure annotation, it is considered a true electrographic seizure, representing the gold standard of this study (green X's of solid outlined data sequence).

2.3. The automatic seizure detection algorithm

For automatic seizure detection in EpiSight a deep neural network model was designed, trained, and deployed using the open-source machine learning platform TensorFlow (https://www.tensorflow.org). The model consists of a deep stack of convolutional layers and ResNet blocks extracting high level features, a feature pyramid network combining features with different levels of abstraction and scale (Tan et al. 2020), an onset position estimation network creating proposals for potential seizure onset times, and a multi-layer perceptron as a classifier distinguishing seizure EEG from interictal background EEG. For a detailed description of the model, please refer to (Hartmann et al. 2022).

To train this model a training dataset including scalp EEG from 490 patients of the Temple University Hospital Seizure Detection Corpus (Shah et al. 2018) and ultra long-term sqEEG from ten healthy subjects (unpublished data; clinicaltrials.gov NCT04513743) were used (not the healthy subjects included in the validation dataset). The sqEEG data included nocturnal home monitoring recordings only. The scalp EEGs, recorded at 10–20 electrode positions, were organized in 14 bipolar channel pairs, each of them mimicking "virtual recordings" of three subcutaneously implanted electrodes. For each seizure annotated in the 10–20 scalp EEG recordings the corresponding training goal was set to achieve at least one detection in one of these 14 bipolar channel pairs. High similarity between EEG from subcutaneous and proximate scalp electrodes in PWE has been demonstrated (Weisdorf et al. 2018), justifying the use of scalp recordings to train the sqEEG seizure detection algorithm.

The described automatic seizure detection algorithm of EpiSight is inherently different from that of the previous version of EpiSight (described in (Fürbass et al. 2015, 2017)) which was applied in the data labelling process.

2.4. Statistical analysis

The sqEEG data of both datasets (epilepsy and normal control) were analyzed with the EpiSight automatic seizure detection algorithm. The resulting EpiSight annotations constituted the index test to be evaluated.

For each PWE, the seizure detection sensitivity and the false detection rate (per 24 h) were reported. An EpiSight detection had to be within a 2-min window when compared to the established gold standard of electrographic seizures to be considered a true positive.

When calculating the data reduction ratio (the proportion of data to discard from the semi-automatic visual review process), it was considered that two minutes of sqEEG are needed for reviewers to assess whether a detection is a true electrographic seizure or not. Thereby, for each PWE, the data reduction ratio was determined as $L_{sqEEG} - 2min \times N_{EpiSight}/L_{sqEEG}$, where L_{sqEEG} represents the total sqEEG length [min] and $N_{EpiSight}$ represents the total number of EpiSight annotations.

Cohen's kappa statistic was calculated to determine the degree of agreement between the self-reported diaries and the electrographic seizures as identified in the review process described above. To make the comparison possible, the accuracy in time of the electrographic seizures was reduced to the date of occurrence.

For the normal control dataset, the data reduction ratios and false detection rates were reported.

3. Results

One PWE was excluded from the validation dataset (PWE D) due to an unusual abundance of interictal findings. All three review teams agreed that it was not possible to separate ictal sqEEG from interictal sqEEG and establish a meaningful gold standard of electrographic seizures.

3.1. Automatic seizure detection and data reduction

An overall electrographic seizure detection sensitivity of 86% was demonstrated, evaluating a total of 94 seizures. All seven severe seizures (focal to bilateral clonic-tonic) were detected. Table 1 summarizes and outlines the main results when evaluating each PWE individually. The cross-PWE sensitivity showed a median of 80% (range: 69–100%), with three PWE having 100% sensitivity (43 seizures in total). The demonstrated sensitivity levels were obtained with a median false detection rate of 2.4 detections per 24 hours (range: 2.0–13.0). In clinical practice, that is just below five minutes of non-seizure sqEEG to be reviewed per 24 recording

Table 1
Epilepsy characteristics and EpiSight automatic seizure detection performance measures for each participant. The top part includes the people with epilepsy (PWE), whereas the lower part includes the normal control. Cohen's kappa values, comparing self-reported seizures with gold standard electrographic seizures for each PWE, are included in the rightmost column.

ID	Onset Zone	Semiology	Self-reported seizures	Cohen's kappa diary	EEG data length [h]	EEG data reduced to [h/month]	Data reduction ratio [%]	Electrographic seizures	SE [%]	FDR [per 24 h]
Epilepsy dataset										
A	LT	FAS	0	–	627	2.2	99.6	0	–	2.4
B	LT	FAS	55	0.51	1552	10.9	98.1	25	100	13.0
C	RT	FIAS w/FBTCS	0	0	975	4.8	99.0	9	78	7.3
E	LT	FIAS w/FBTCS	0	0	1147	1.9	99.6	17	100	2.2
F	LT	FAS	21	0.29	1808	8.9	98.7	5	80	9.6
G	LT	Uncertain	13	0.53	1516	2.1	99.7	13	69	2.2
H	LT	FIAS w/FBTCS	1	1.0	1364	1.7	99.7	1	100	2.0
I	LT	FIAS w/FBTCS	133	0.0056	1605	2.0	99.6	24	75	2.3
Median				0.29	1440	2.2	99.6		80	2.4
Normal control dataset										
01	–	–	–	–	1007	0.75	99.9	–	–	0.93
02	–	–	–	–	1010	3.0	99.5	–	–	3.3
03	–	–	–	–	1024	0.53	99.9	–	–	0.63
04	–	–	–	–	988	2.8	99.4	–	–	4.1
05	–	–	–	–	1056	0.80	99.9	–	–	0.89
06	–	–	–	–	968	0.82	99.9	–	–	1.0
07	–	–	–	–	974	0.19	99.9	–	–	0.25
08	–	–	–	–	981	1.1	99.9	–	–	1.5
09	–	–	–	–	888	2.1	99.8	–	–	2.4
10	–	–	–	–	527	0.46	99.7	–	–	0.55
11	–	–	–	–	990	0.46	99.9	–	–	0.65
12	–	–	–	–	1002	0.27	99.9	–	–	0.41
Median				–	989	0.78	99.9	–	–	0.91

Abbreviations: FAS, focal aware seizure; FBTCS, focal to bilateral tonic-clonic seizure; FDR, false detection rate; FIAS, focal impaired awareness seizure; LT, left temporal; RTF, right frontotemporal; RT, right temporal; SE, sensitivity.

hours (when assuming the required two minutes per event). An ictal sqEEG example from each PWE is included in the Supplementary Material.

The algorithm reduced the amount of data to be reviewed with a median of 99.6% (range: 98.1–99.7%) in the epilepsy dataset, evaluating each PWE individually. In absolute hours, the data load was reduced to a median of 2.2 hours of EEG data to be reviewed per month for each PWE (considering the individual user adherence). For the normal control dataset, the EEG data amount was reduced by a median of 99.9% (range: 99.4–99.9%) and had a median false detection rate of 0.91 pr. 24 hours (range: 0.25–4.1).

The three reviewers were asked to report their actual review time, necessary for the evaluation of the whole recording (serving as gold standard in this study). SW reported a total review time of approximately-three months (12 weeks of 40 h work = 480 h), AMKO estimated a total review time of 432 hours, whereas PFV (who performed an EpiSight-assisted semi-automatic review process) reported 26 hours of review time. Thus, the conservative estimation of the review time consumption reduction ratio is (432 h–26 h)/432 h = 94%, when applying the semi-automatic review process (with the previous version of the detection algorithm).

3.2. sqEEG-based annotations vs diary

Fig. 3 visualizes the backbone of the analysis results. For each PWE, all EpiSight annotations (true and false positives) are plotted together with the electrographic seizures, and self-reported diary entries. It is demonstrated that the ability to actively self-report seizures varies across the PWE. Likewise does the agreement with the electrographic seizures and the sensitivity of the automatic algorithm. PWE C and E showed empty diaries (since they were not aware of their seizures), but numerous electrographic seizures, and high seizure detection sensitivities at reasonable levels of false detections. On the other hand, PWE B, F, G, and I were active self-reporters. PWE I reported heavily, PWE B and F reported about twice as many seizure days as electrographic seizures, whereas the self-reports of PWE G sum to the true number of electrographic seizures. Lastly, PWE A and H demonstrated full agreement between diary and electrographic seizures – though data was limited to zero and one electrographic seizure, respectively.

The observations are formalized in Cohen's kappa values of Table 1, which support the general understanding that self-reported seizure diaries are inaccurate, since the agreements between the diary and the electrographic seizures are low (kappa < 0.6 (Landis and Koch 1977)) for all but one PWE (median: 0.29; range: 0–1.0).

4. Discussion

The EpiSight automatic seizure detection algorithm was evaluated on 490 days of outpatient sqEEG data from eight PWE, including 94 electrographic seizures (mainly non-convulsive). An overall seizure detection sensitivity of 86% was demonstrated. False detection rates were found to be a median of 2.4 false detection per 24 hours for PWE (range: 2.0–13.0) and 0.91 per 24 hours (range: 0.25–4.1) for the normal control group. This means that on its own (unsupervised) the fully automated detection does not meet clinical requirements, due to the false alarms. However, it decreases dramatically the dataset, without significant drop in sensitivity, hence making this suitable for a semi-automatic (hybrid) approach, where experts review only the epochs detected by the algorithm.

4.1. The automatic seizure detection algorithm

Available scalp-EEG based seizure detection algorithms applied in a clinical setting provide sensitivities of 75–90% and false detec-

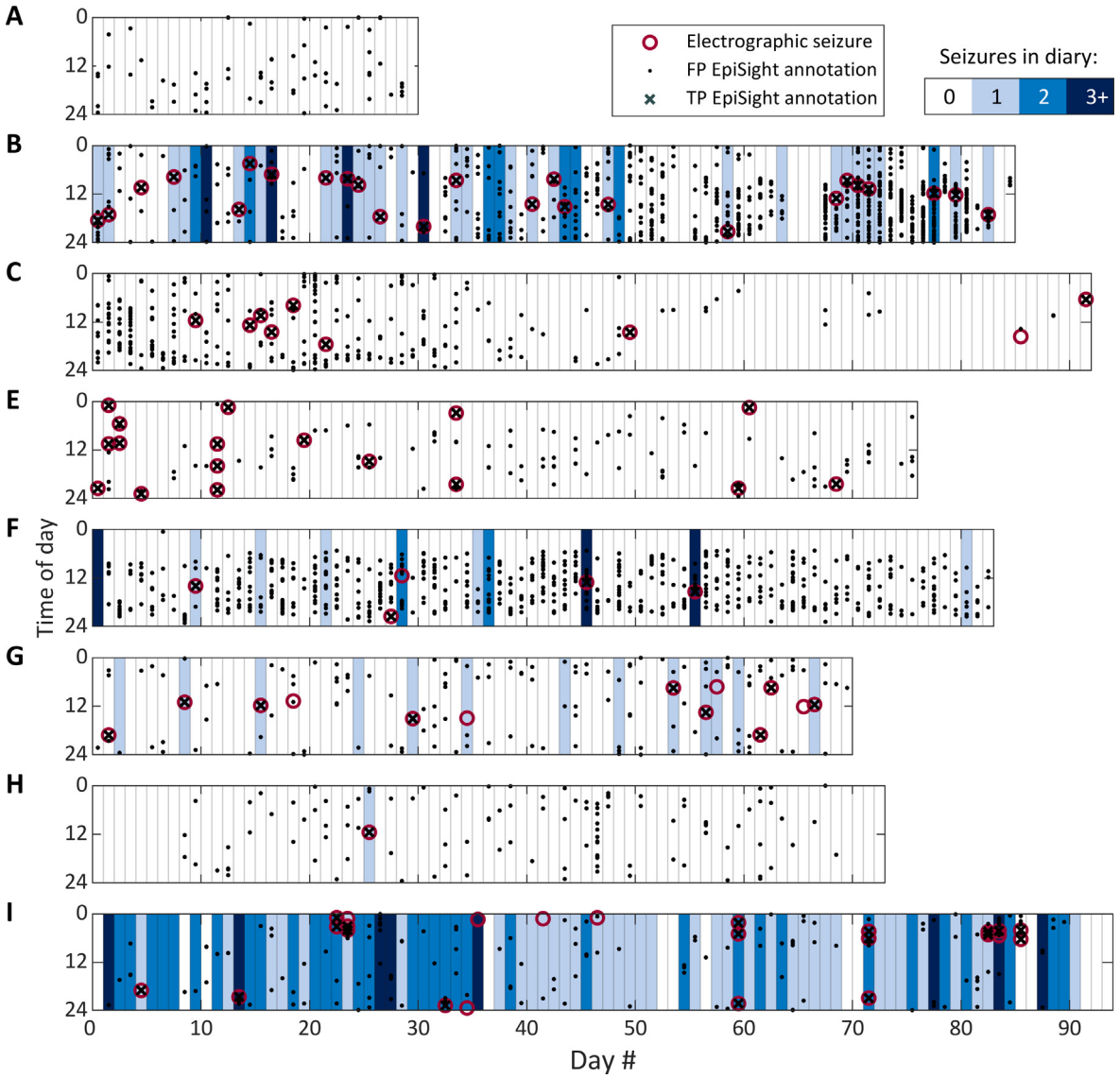

Fig. 3. Chart for all people with epilepsy showing electrographic seizures (red circles), false positive (FP) EpiSight annotations (black dots), true positive (TP) EpiSight annotations (black crosses), and self-reported diary events (shaded lines; light blue = 1 seizure, medium blue = 2 seizures, dark blue = 3 + seizures).

tion rates of 2.4–120 per 24 hours (Baumgartner and Koren 2018). Thus, the performance of this work is indeed comparable – especially considering that the EpiSight algorithm was evaluated on ultra long-term real-life data, which could be expected to generate more false detections. Noteworthy too is that the evaluated database only included seven convulsive seizures (focal to bilateral clonic-tonic), distributed in four different PWE, which are usually the easiest to detect. These were all detected, but most seizures were of other types, challenging the algorithm further.

A walk-through of the false negatives (distributed among half of the PWE) by manual inspection reveals no unique tendency and explanation. However, common for the missing detections of PWE F and I is that these are the seizures of shortest duration. There was no particular distribution in time of occurrence of the false negatives; they occurred in accordance with the circadian distribution of the true positives of the given PWE. A similar manual inspection of a subset of the false detections reveals subject-specific patterns: e.g., chewing artefacts for one PWE, and other high frequent patterns for the other. Homogeneity exists within subjects, but not across subjects. This also holds when considering the distribution of false detections throughout the day: some PWE experience most false detections during daytime (PWE A, E, F), while it is during sleep for another (PWE G). One can imaging taking advantage of such insights of the circadian distribution of both false detections and true seizures in the future, tailoring subject-specific algorithms. PWE G, who demonstrated the lowest sensitivity of this study (9/13 = 69%), could very well benefit from this as the electrographic seizures occurs at times 7–19, whereas the false detections mainly occur outside this time span.

Even without tailoring subject-specific solutions, the general algorithm performance is expected to be improved significantly when more relevant labelled sqEEG real-life data is available for training. For development of the current detection algorithm neither ictal nor interictal sqEEG data were included in the training set as such was not available – only (nocturnal) sqEEG from subjects without epilepsy. A major challenge in the development of seizure detection systems is the difficulty in validating performance, since validation requires a gold standard of true seizures. Thus, when data is available, it still needs to be labelled.

4.2. The data reduction ratio

Comparing the data reduction ratios of the two validation datasets indicates a small difference in the efficiency of a semi-automatic review process for the two groups, since the reduction ratio is higher for the normal control dataset. This difference might be driven by both independent variables: group (healthy vs PWE) and electrode placement (vertical vs optimal to seizure location). The midline implantations in the healthy subjects likely had less muscle artefacts and hence less false detections. In addition, the sqEEG of PWE might include interictal epileptiform discharges and paroxysms, potentially elevating the number of false detections. However, the clinicians might benefit from the review of such events even though they are not classified as electrographic seizures.

The data reduction level of both datasets seems at a fair level for the semi-automatic seizure detection/review process to work in practice. The accepted level of false detections is much higher, when applied in a semi-automatic review context as opposed to an incorporation in a real-time alarm system. In general, when evaluating epileptic seizure detection algorithms, the application must be kept in mind. The sensitivity might be compromised if the purpose is monitoring of *changes* in (non-rare) seizure occurrence rates to clarify multi-day seizure rhythms or responses to changes in AED prescriptions. If the ultimate purpose is replacement of self-reported diaries, the most important is: does the given solution outperform the diary?

4.3. The sqEEG-based annotations vs the diary: adding clinical value?

The Cohen's kappa values reporting on the agreement between the self-reported seizure entries and the recorded electrographic seizures supported the general understanding that diaries are inaccurate – and examples of both underreporting and potential overreporting were demonstrated. However, concerning the electrographic seizures the gold standard is controversial, when evaluating diary performance – especially regarding the potential cases of overreporting (subjects B, F, and I). Given the limited spatial coverage of the brain (and the missing video) it is not clear whether they overreport or the two-channel electrode failed to catch seizure activity at a different location. In addition, there are periods of outpatient monitoring where no sqEEG is recorded – and the existence of clinical seizures outside the monitoring period cannot be excluded. The total adherence was moderate (73%) with high individual variability (45% – 91%) (Weisdorf et al. 2019).

The demonstrated low diary agreement and performance of the sqEEG-based seizure detection algorithm indicate that the new approach is a valuable alternative for the future, potentially outperforming the diary. That said, being (minimally) invasive the sqEEG-based method must add solid clinical value to PWE. This is considered the case for six out of eight PWE, see summary comments in Table 2. Of the six, PWE B could be questioned (kappa 0.51), but shedding light on potential overreporting is relevant even though it seems to have received less attention than underreporting, traditionally. Psychogenic non-epileptic seizures could be considered a kind of overreporting where there is a great need for new diagnostic tools. Another to question is PWE G (kappa 0.53), who self-reported the true number of seizures, but 5 out of 13 were reported on the wrong day. A glance at the overview chart of Fig. 3 indicates that an analysis of change in seizure occurrence rate of PWE G could yield the same results irrespective of diary use or sqEEG. However, if small scale multi-day cycles are of interest, annotation on the correct weekday matters.

A limitation of the current investigation is not being able to conclude on a potential direction of the disagreement between the diary and the electrographic seizures (reflecting diary under- or overreporting) due to the limited sample size. This constitutes a direction for future research, just as more trials are needed to assess the real value of the sqEEG-based seizure detection over the traditional diary.

4.4. The clinical review process of ultra long-term sqEEG

As demonstrated, when comparing the reported actual time consumptions reviewing the ultra long-term sqEEG with three different approaches, a semi-automatic review processes is a prerequisite for sqEEG to become clinical practice. In this case, the semi-automatic process reduced the review time consumption by 94%.

The reviewers found the electrographic seizure patterns of the two-channel sqEEG recognizable as compared to the traditional scalp EEG setup. Occasionally, they missed the information contained within a usual simultaneous video-EEG. However, all reviewers reported adapting very well to the modality at quite a fast pace. In the occasional events of doubt, reviewers expressed that they could have benefitted from examples of typical two-channel sqEEG artefacts (e.g., eye blinking) and seizure paradigms to better separate potential electrographic seizures from artefacts.

4.5. The validation strategy

It is very likely that a confirmation bias was introduced when reviewers evaluated seizure-annotated sqEEG traces in the second review step. The review processes were diverse, and thus direct comparisons were not possible. Nonetheless, the goal was not to compare the review teams and quantify interrater variabilities. Instead, the goal was to establish the best possible labelled dataset

Table 2
Could the ultra long-term sqEEG-based method including semi-automated seizure/review add clinical value for the given people with epilepsy (PWE)? Summing up and comparison with diary.

ID (number of electrographic seizures)	Could the ultra long-term sqEEG-based method, including semi-automatic seizure detection/review, add clinical value for the PWE?
A (0)	No reported seizures, neither in diary nor electrographic. Too short recording time to conclude.
B (25)	Active self-reporter. Low/moderate diary agreement (kappa 0.51), but perfect EpiSight sensitivity. Shed light on potential overreporting in diary. Adds clinical value.
C (9)	Empty diary = no diary agreement. EpiSight sensitivity of 78%. Adds clinical value.
E (17)	Empty diary = no diary agreement. Perfect EpiSight sensitivity. Adds clinical value.
F (5)	Active self-reporter. Low diary agreement (kappa 0.29). EpiSight sensitivity of 80%. Adds clinical value.
G (13)	Active self-reporter. Low/moderate diary agreement (kappa 0.53). Moderate EpiSight sensitivity (69%). Shed light on the mismatch between the reported dairy days and the days with electrographic seizures. Adds clinical value.
H (1)	Only one reported seizure in diary and electrographically. Complete diary agreement and perfect EpiSight sensitivity.
I (24)	Very active self-reporter. Very low diary agreement (kappa 0.0056). EpiSight sensitivity of 75%. Adds clinical value.
Total	Adds clinical value for six out of eight PWE.

and gold standard of electrographic seizures. Applying different review processes supports this. However, a level of agreement was extracted by counting the number of raters to agree in identifying each electrographic seizure as accepted by majority vote (either three or two raters). For two PWE all electrographic seizures were agreed upon by three raters, for four PWE all but one electrographic seizure were accepted by three raters, while for a single PWE, 13 of 24 electrographic seizures were annotated by all three raters. For this latter PWE, the sqEEG included brief interictal rapid discharges (BIRDs) with durations between 4–10 s. The duration of seizures and BIRDs seemed to be the factor of doubt among the raters.

5. Conclusions

Ultra long-term subcutaneous EEG is a novel option for the recording of electrographic seizures in an outpatient setting. However, interaction with a supportive solution of a data reduction algorithm is a prerequisite to make data review possible in clinical practice. The EpiSight automatic seizure detection algorithm was demonstrated to be very efficient, reducing the amount of data to be reviewed by 99.6% in PWE and 99.9% in the normal control group. This with a cross-PWE sensitivity of 86% (median 80%, range 69–100% when PWE were evaluated individually) and a median false detection rate of 2.4 detections per 24 hours (range 2.0–13.0), which is considered a clinical applicable specificity for a semi-automatic review process. In comparison with self-reported seizure diaries, six of eight PWE were considered examples of clinical cases where the objective sqEEG-based seizure monitoring would provide valuable insights to optimize epilepsy treatment strategy.

Conflict of interest

JDH and LSR are employees of UNEEG medical A/S. TWK consults for UNEEG medical A/S. PFV and AMKO received a payment from UNEEG medical A/S for data annotation. SW has received support from UNEEG medical A/S. MPR has been a member of ad hoc advisory boards for UNEEG medical A/S. FF and MH are employees of AIT Austrian Institute of Technology GmbH, which has developed UNEEG™ EpiSight Analyzer.

Acknowledgements

MPR is supported by the Epilepsy Foundation's Epilepsy Innovation Institute My Seizure Gauge Project; the NIHR Biomedical Research Centre at the South London and Maudsley NHS Foundation Trust; the MRC Centre for Neurodevelopmental Disorders (MR/N026063/1).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.clinph.2022.07.504.

References

Baud MO, Kleen JK, Mirro EA, Andrechak JC, King-Stephens D, Chang EF, et al. Multi-Day Rhythms Modulate Seizure Risk in Epilepsy. Nat Commun 2018;9(1):1–10. https://doi.org/10.1038/s41467-017-02577-y.

Baumgartner C, Koren JP. Seizure Detection Using Scalp-EEG. Epilepsia 2018;59:14–22. https://doi.org/10.1111/epi.14052.

Blachut B, Hoppe C, Surges R, Elger C, Helmstaedter C. Subjective Seizure Counts by Epilepsy Clinical Drug Trial Participants Are Not Reliable. Epilepsy Behav 2017;67(February):122–217. https://doi.org/10.1016/J.YEBEH.2016.10.036.

Bruno E, Viana PF, Sperling MR, Richardson MP. Seizure Detection at Home: Do Devices on the Market Match the Needs of People Living with Epilepsy and Their Caregivers? Epilepsia 2020;61(S1). https://doi.org/10.1111/epi.16521.

Duun-Henriksen J, Baud M, Richardson MP, Cook M, Kouvas G, Heasman JM, et al. A New Era in Electroencephalographic Monitoring? Subscalp Devices for Ultra–Long-Term Recordings. Epilepsia 2020;61(9):1805–17.

Elger CE, Hoppe C. Diagnostic Challenges in Epilepsy: Seizure under-Reporting and Seizure Detection. Lancet Neurol 2018;17(3):279–88. https://doi.org/10.1016/S1474-4422(18)30038-3.

Fisher RS, Blum DE, DiVentura B, Vannest J, Hixson JD, Moss R, et al. Seizure Diaries for Clinical Research and Practice: Limitations and Future Prospects. Article *Epilepsy Behav* 2012;24(3):304–10.

Fürbass F, Kampusch S, Kaniusas E, Koren J, Pirker S, Hopfengärtner R, et al. Automatic Multimodal Detection for Long-Term Seizure Documentation in Epilepsy. Clin Neurophysiol 2017;128(8):1466–72.

Fürbass F, Ossenblok P, Hartmann M, Perko H, Skupch AM, Lindinger G, et al. Prospective Multi-Center Study of an Automatic Online Seizure Detection System for Epilepsy Monitoring Units. Clin Neurophysiol 2015;126(6):1124–31.

Goldenholz DM, Goldenholz SR, Moss R, French J, Lowenstein D, Kuzniecky R, et al. Does Accounting for Seizure Frequency Variability Increase Clinical Trial Power? Epilepsy Res 2017;137:145–51.

Hartmann M, Koren JP, Baumgartner C, Duun-Henriksen J, Gritsch G, Kluge T, et al. Seizure Detection with Deep Neural Networks for Review of Two-Channel EEG. Epilepsia 2022. https://doi.org/10.1111/epi.17259.

Karoly PJ, Goldenholz DM, Freestone DR, Moss RE, Grayden DB, Theodore WH, et al. Circadian and Circaseptan Rhythms in Human Epilepsy: A Retrospective Cohort Study. Lancet Neurol 2018;17(11):977–85.

Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics 1977;33(1):159–74. https://doi.org/10.2307/2529310.

Shah V, von Weltin E, Lopez S, McHugh JR, Veloso L, Golmohammadi M, et al. The Temple University Hospital Seizure Detection Corpus. Front Neuroinform 2018;12:1–6. https://doi.org/10.3389/fninf.2018.00083.

Tan M, Pang R, Quoc VL. EfficientDet: Scalable and Efficient Object Detection. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. p. 10778–87. https://doi.org/10.1109/CVPR42600.2020.01079.

Viana PF, Remvig LS, Duun-Henriksen J, Glasstetter M, Dümpelmann M, Nurse ES, et al. Signal Quality and Power Spectrum Analysis of Remote Ultra Long-Term Subcutaneous EEG. Epilepsia 2021;62(8):1820–8.

Weisdorf S, Duun-Henriksen J, Kjeldsen MJ, Poulsen FR, Gangstad SW, Kjær TW. Ultra-long-term Subcutaneous Home Monitoring of Epilepsy—490 Days of EEG from Nine Patients. Epilepsia 2019;60(11):2204–14. https://doi.org/10.1111/epi.16360.

Weisdorf S, Gangstad SW, Duun-Henriksen J, Mosholt KSS, Kjær TW. High Similarity between EEG from Subcutaneous and Proximate Scalp Electrodes in Patients with Temporal Lobe Epilepsy. J Neurophysiol 2018;120(3):1451–60. https://doi.org/10.1152/JN.00320.2018.