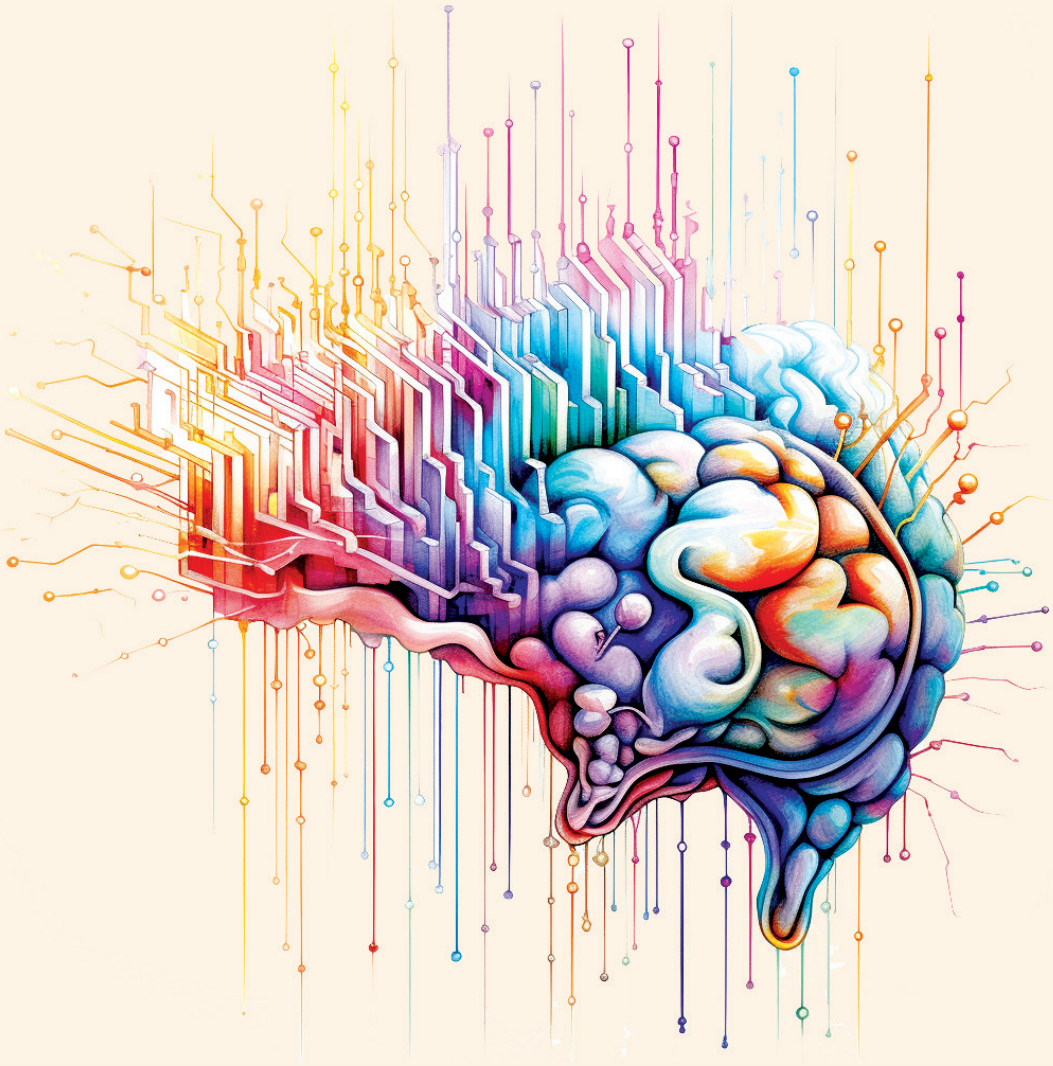
Automated spike and seizure detection: Can we use it?

Using automated spike and seizure detection when reviewing prolonged EEGs at the Epilepsy Monitoring Unit



Elisabeth (Elise) E.M. Reus

Automated spike and seizure detection:

Can we use it?

Using automated spike and seizure detection when reviewing prolonged EEGs at the Epilepsy Monitoring Unit

Elisabeth Emmerentia Maria Reus

Automated spike and seizure detection: Can we use it? ***Using automated spike and seizure detection when reviewing prolonged EEGs at the Epilepsy Monitoring Unit***

ISBN/EAN:	978-90-833743-8-3
Printed by:	proefschriftenprinten.nl
Cover illustration:	Image generated by Midjourney, inc. Midjourney, 2023. Lay-out cover: proefschriftenprinten.nl
Lay-out & Design:	Elise Reus

The author is grateful to Persyst Development Corporation (San Diego, CA, USA), AIT Austrian Institute of Technology, (Vienna, Austria) and BESA Epilepsy (Germany) for allowing the use of trial licenses of their software. None had any influence on the design and results of the research or any other contents of this thesis.

Automated spike and seizure detection: Can we use it?

Using automated spike and seizure detection when reviewing prolonged EEGs at the Epilepsy Monitoring Unit

Automatische piek- en aanvalsdetectie: kunnen we het gebruiken?

Het gebruiken van automatische piek- en aanvalsdetectie bij het uitwerken van langdurige EEG-registraties op de Epilepsy Monitoring Unit

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus Prof.dr.ir. Hester Bijl,
volgens besluit van het college voor promoties
te verdedigen op dinsdag 30 januari 2024
klokke 15.00 uur

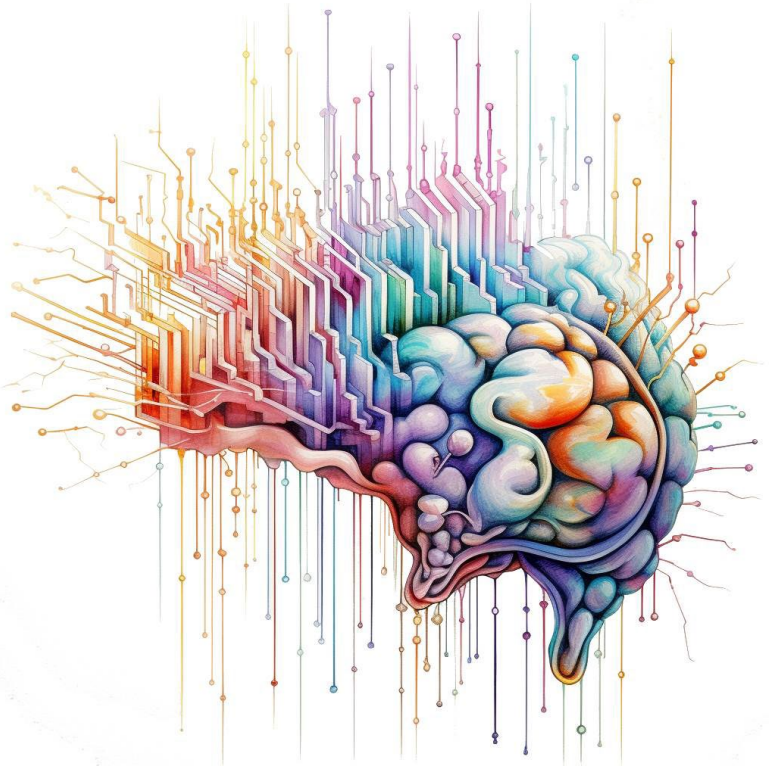
door

Elisabeth Emmerentia Maria Reus

geboren te Hoorn
in 1988

Promotor:	Prof. dr. J.G. van Dijk
Copromotor:	Dr. G.V. Visser (SEIN) Dr. F.M.E. Cox (SEIN)
Thesis committee:	Prof. dr. G.J. Lammers Prof. dr. G.J.M. Zijlmans (SEIN, Utrecht UMC) Prof. dr. N. van Alfen (Radboud UMC) Dr. R.D. Thijs

Contents



CHAPTER 1

General Introduction

Chapter 1 - General introduction

Detecting interictal epileptiform discharges (IEDs) and seizures is helpful in epilepsy diagnosis, classification and monitoring [1,2,3]. Prolonged recordings have been shown to enhance the chances of detecting both seizures and IEDs, thereby leading to higher diagnostic efficiency [4,5].

The Epilepsy Monitoring Unit (EMU) at Stichting Epilepsie Instellingen Nederland (SEIN) conducts prolonged video-EEGs. Technicians visually examine the entire EEG record to identify all relevant interictal and ictal events in order to address the referral question. Selected segments are then reviewed by a clinical neurophysiologist, who provides a final electroclinical diagnosis. This complete visual data analysis is time consuming and expensive because of the need for specialized personnel. The current shortage of staff, especially technicians, causes a high workload and may eventually result in scaling down the number of EEGs. It is therefore necessary to explore time-saving alternatives to review large EEG datasets without compromising quality.

'Sampled visual review' is a time-saving technique that has received limited evaluation. In this approach, only specific segments of the extended EEG are visually analyzed. Research indicates that analyzing the first hour of sleep can predict the presence of IEDs for the entire recording with high accuracy [6]. Another study suggests that sampled reviews are equally effective in determining the final electroclinical diagnosis compared to the traditional method although some events, interictal or ictal, may be overlooked [7].

Another approach is automated EEG analysis, using IED (also referred to as 'spike') and seizure detection software. By detecting spikes and seizures automatically, it can provide reviewers with valuable information and alleviate the burden of visual analysis. The software can detect all possible interictal events and present them to the reviewer, which eliminates the need for a page-by-page inspection of the recording. Despite considerable research efforts and technological advancements in computer science since the 1970s, automated EEG analysis software for routine scalp EEG remains underutilized by clinicians due to issues with usability and lack of confidence [8, 9]. However, in recent years, there has been significant improvement in software performance and detection algorithms, as well as commercially available software packages have emerged.

For an algorithm to be effectively employed in an EMU environment, it should incorporate both seizure detection and IED detection capabilities. Additionally, the algorithm should possess a user-friendly interface to facilitate its utilization by technicians and neurologists who typically lack a technical background. For these reasons, in this thesis, we only use commercial available detection software packages.

Most commercially available software packages use neural networks. These are composed of basic computing units which learn and recognize patterns in ways that are at least

superficially similar to humans. Neural networks do not need specific rules for training, only examples. However, most neural networks methods use supervised learning in which each training signal input into the network is classified [8]. Examples include assessments of power, frequency, segmentation and evolution of rhythmic activity (seizure detection) or amplitude, sharpness, whether the event exceeds that of the background activity and field (IED detection) [10,11]. There are typically several layers between input and output [12]. Furthermore, non-cerebral activity can be identified. This includes electrode artifact, muscle artifact, chewing artifact, and vertical and horizontal eye movement potentials [11].

Objective

The objective of this thesis is to investigate the potential of using automated spike and seizure detection in reviewing prolonged EEGs without compromising the quality of the electroclinical diagnosis. This research includes the potential to combine automated detection with sampled visual review (Figure 1) in order to obtain an impression of the background activity. This helps the reviewer properly assess the automated detections, and to detect potential aspecific (focal or diffuse) dysfunctions or rhythmic delta activity. Visually assessed samples consist of one hour during wakefulness, one hour during sleep, half an hour of wakefulness after wake-up and all clinical events marked by the patient or nurses. After the sampled visual review, the whole record is analyzed using automated spike and seizure detection.

The quality of the diagnosis using sampled visual review in combination with automated spike and seizure detection must be at least equivalent to that achieved through complete visual review, without any loss of accuracy. In other words, the review method using automated detection software must perform at least as well as human experts in detecting IEDs and seizures.

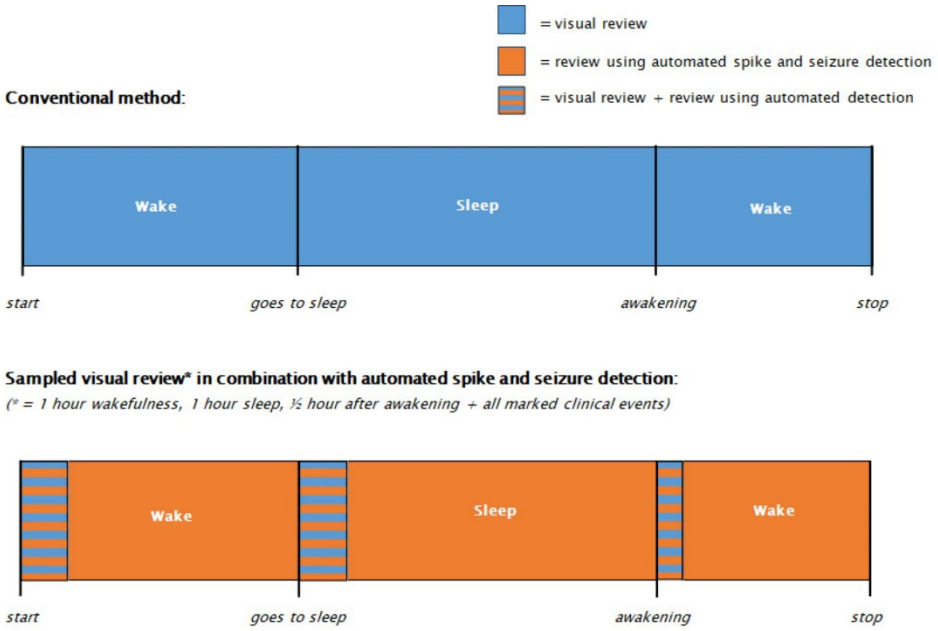


Figure 1. Conventional (above) and proposed (below) EEG review methods of a 24h EEG

Automated spike and seizure packages

In this thesis we used three software packages: Persyst (Persyst Development Corporation, USA), Encevis (AIT Austrian Institute of Technology, Austria) and BESA (BESA Epilepsy, Germany).

Persyst

Persyst has three sensitivity settings for IED detection. In our research we used the default medium setting and the low setting, the setting with the highest specificity. The output is a list of timed IED detections per electrode [11]. Output for seizure detection is also a list of timed detection with corresponding seizure probability. Version P13 has no user-adjustable settings, whereas the latest version (P14) allows a user to adjust the threshold settings for seizure duration and for seizure probability [10].

Encevis

EpiSpike, the IED detection module, has no sensitivity settings. The output is a list of timed IED detections per electrode. The output of the seizure detection module, Episcan, is a list of timed detections with corresponding elektrodes. Encevis is the only package that also uses the ECG channel for seizure detection [13].

BESA epilepsy

The BESA software uses a different method. Instead of showing detections as distinct events, it offers detections as clusters. These clusters are made for every two-hour epochs, using four different parameters focusing on waveform, topography, location and orientation [14]. At least four similar events are required for a cluster to be identified. The remaining detections are placed in a residual section. The software shows the 20 waveforms with the most similarity to the cluster mean to the user, together with the equivalent locations of the events in a head scheme. A human reviewer has to categorize the clusters as epileptiform or not.

The performance of automated detection software is assessed based on two critical factors: sensitivity and specificity (false detection rate). The sensitivity must be high enough to detect all relevant interictal epileptiform discharges and seizures, and the false detection rate must be low enough to avoid requiring significant time to filter through false positives [15].

Aims and outline of the thesis

The first chapter we compared the performance of a commercially available spike detection algorithm to that of human expert consensus when calculating the spike-wave index (**Chapter 2**).

Chapter 3 proposes and studies a new review method using sampled visual review in combination with automated spike and seizure detection (Figure 1). Visually assessed samples consist of one hour during wakefulness, one hour during sleep, half an hour of wakefulness after wake-up and all clinical events marked by the patient or nurses.

Additionally, the thesis researches which of the three software packages performed best in our EMU setting, for spike detection (**Chapter 4**) as well as and seizure detection (**Chapter 5**). IEDs were marked by a consensus of three human experts and were compared with the detections of all three software packages. Seizures were extracted from the original clinical reports and were compared with the detections of all three software packages in combination with the live observations by nursing staff and patients push buttons.

Finally, in the thesis we explored sentiments that can hinder successful implementation of the software (**Chapter 6**) and includes semi-structured interviews with clinical neurophysiology staff and outpatient clinic neurologists regarding their view on automated spike and seizure detection. The interviews identify multiple factors that may hinder or facilitate future implementation, referred to as ‘barriers’ and ‘facilitators’, respectively.

References

1. Fountain NB, Freeman JM. EEG is an essential clinical tool: pro and con. Epilepsia 2006;47:23–5. https://doi.org/10.1111/j.1528-1167.2006.00655.x.

2. Lee C, Lim S, Lien F, Wu T. Duration of electroencephalographic recordings in patients with epilepsy. Seizure 2013;22:438–42. https://doi.org/10.1016/j.seizure.2013.02.016.

3. Thijs RD, Surges R, O'Brien TJ, Sander JW. Epilepsy in adults. Lancet 2019;16;393:689-701. https://doi.org/10.1016/S0140-6736(18)32596-0.

4. Wirrell ED. Prognostic significance of interictal epileptiform discharges in newly diagnosed seizure disorders. J Clin Neurophysiol 2010;27:239-48. https://doi.org/10.1097/WNP.0b013e3181ea4288.

5. Koca G, Morkavukb G, Akkayab E, Karadas O, Leventoglu A, Unay B, et al. Latencies to first interictal epileptiform discharges in different seizure types during video-EEG monitoring. Seizure: European Journal of Epilepsy 2019;69:235–40. https://doi.org/10.1016/j.seizure.2019.05.013.

6. Liu X, Issab NP, Roseb S, Wu S, Sun T, Towle LV, et al. The first-hour-of-the-day sleep EEG reliably identifies interictal epileptiform discharges during long-term video-EEG monitoring. Seizure. 2018;63:48–51. https://doi.org/10.1016/j.seizure.2018.10.015.

7. Badawy RAB, Pillay N, Jetté N, Wiebe S, Federico P. A blinded comparison of continuous versus sampled review of video-EEG monitoring data. Clin Neurophysiol. 2011;122:1086–90. https://doi.org/10.1016/j.clinph.2010.10.048.

8. Halford JJ. Computerized epileptiform transient detection in the scalp electroencephalogram: Obstacles to progress and the example of computerized ECG interpretation. Clin Neurophysiol 2009;120:1909-15. https://doi.org/10.1016/j.clinph.2009.08.007.

9. Rubboli G, Beniczky S, Claus S, Canevini MP, Kahane P, Stefan H, et al. A European survey on current practices in epilepsy monitoring units and implications for patients' safety. Epilepsy & Behav 2015;44:179–84. https://doi.org/10.1016/j.yebeh.2015.02.004.

10. Scheuer ML, Wilson SB, Antony A, Ghearing G, Urban A, Bagić AI. Seizure Detection: Interreader Agreement and Detection Algorithm Assessments Using a Large Dataset. J Clin Neurophysiol 2021;38:439-47. https://doi.org/10.1097/WNP.0000000000000709.

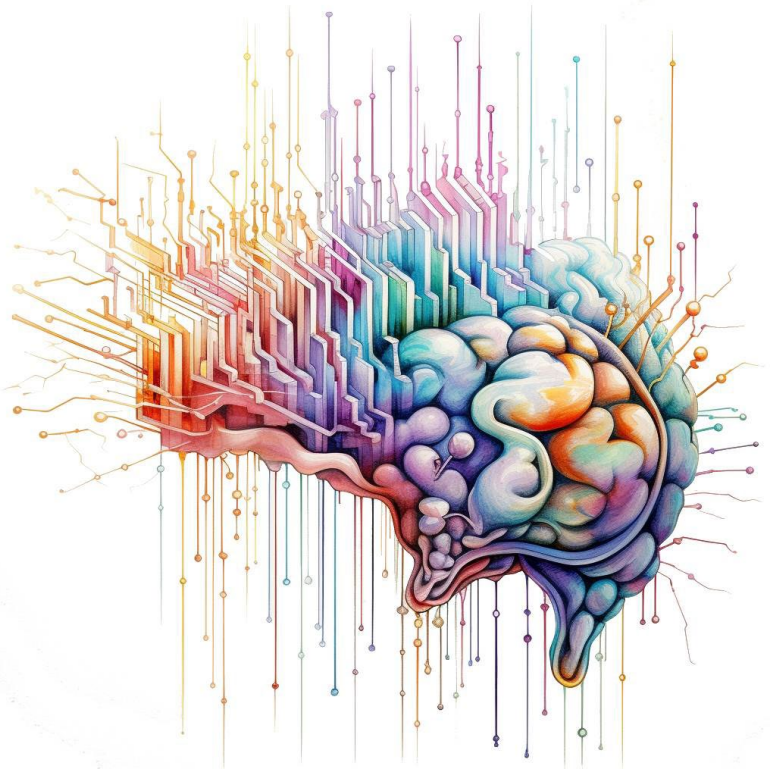
11. Wilson SB, Turner CA, Emerson RG, Scheuer ML. Spike detection II: automatic, perception-based detection and clustering. Clin Neurophysiol 1999;110:404–11. https://doi. org/10.1016/s1388-2457(98)00023-6

12. Tjepkema-Cloostermans MC, De Carvalho RCV, Van Putten MJAM. Deep learning for detection of focal epileptiform discharges from scalp EEG recordings. Clin Neurophys 2018;129:2191-96. https://doi.org/10.1016/j.clinph.2018.06.024.

13. Fürbass F, Ossenblok P, Hartmann M, Perko H, Skupch AM, Lindinger G. Prospective multi-center study of an automatic online seizure detection system for epilepsy monitoring units. Clin Neurophysiol 2015;126:1124-31. https://doi.org/10.1016/j.clinph.2014.09.023.

14. Scherg M, Ille N, Weckesser D, Ebert A, Ostendorf A, Boppel T, et al. Fast evaluation of interictal spikes in long-term EEG by hyper-clustering. Epilepsia 2012;53(7):1196–204. https://doi.org/10.1111/j.1528-1167.2012.03503.x.

15. Lodder SS, van Putten MJAM. A self-adapting system for the automated detection of inter-ictal epileptiform discharges. PLoS One. 2014 Jan 15;9(1):e85180. https://doi.org/10.1371/journal.pone.0085180.



CHAPTER 2

Determining the Spike–Wave Index

using automated detection software

E.E.M. Reus
G.H. Visser
F.M.E. Cox

J Clin Neurophysiol 2021;38:198-201.
https://doi. org/10.1097/WNP.0000000000000672.

Chapter 2 - Determining the Spike–Wave Index using automated detection software

Abstract

Purpose: The spike–wave index (SWI) is a key feature in the diagnosis of electrical status epilepticus during slow-wave sleep. Estimating the SWI manually is time-consuming and is subject to interrater and intrarater variability. Use of automated detection software would save time. Thereby, this software will consistently detect a certain EEG phenomenon as epileptiform and is not influenced by human factors. To determine noninferiority in calculating the SWI, we compared the performance of a commercially available spike detection algorithm (P13 software, Persyst Development Corporation, San Diego, CA) with human expert consensus.

Method: The authors identified all prolonged EEG recordings for the diagnosis or follow-up of electrical status epilepticus during slow-wave sleep carried out from January to December 2018 at an epilepsy tertiary referral center. The SWI during the first 10 minutes of sleep was estimated by consensus of two human experts. This was compared with the SWI calculated by the automated spike detection algorithm using the three available sensitivity settings: "low," "medium," and "high." In the software, these sensitivity settings are denoted as perception values.

Results: Forty-eight EEG recordings from 44 individuals were analyzed. The SWIs estimated by human experts did not differ from the SWIs calculated by the automated spike detection algorithm in the "low" perception mode ($P = 0.67$). The SWIs calculated in the "medium" and "high" perception settings were, however, significantly higher than the human expert estimated SWIs (both $P < 0.001$).

Conclusion: Automated spike detection (P13) is a useful tool in determining SWI, especially when using the "low" sensitivity setting. Using such automated detection tools may save time, especially when reviewing larger epochs.

Introduction

A key feature in the diagnosis of electrical status epilepticus during slow-wave sleep (ESES) is the amount of epileptiform activity occurring during sleep, usually expressed as a "spike–wave index" (SWI) [1]. In 1971, ESES was originally described as an epileptic encephalopathy characterized by sleep-induced activation of epileptiform activity on the EEG [2]. In 1989, the International League Against Epilepsy (ILAE) defined the characteristic EEG pattern in ESES as continuous diffuse spike– waves during slow-wave sleep [3]. This condition mainly affects children and is associated with cognitive decline involving a wide spectrum of developmental and neurocognitive domains [4]. The underlying etiology can be structural or genetic [5]. The ILAE definition of ESES does not include a specific cut-off percentage regarding the amount of epileptiform activity in the EEG. A recent guideline, however, suggested a criterion of at least 50% epileptiform activity during sleep, especially if the clinical symptoms are compatible with an ESES-related syndrome [6]. The same guideline also mentions a cut-off of at least 85% epileptiform activity, mainly to facilitate comparison with existing literature. The methods used to determine the SWI varies, especially regarding the amount of sleep EEG which is analyzed (from 100 seconds to a whole sleep cycle) [6,7]. Automated spike detection algorithms have long been available [8]. They are useful in reviewing EEG recordings by detecting interictal epileptiform discharges, to quantify spike density, and possibly to distinguish different epileptiform morphologies [9]. Experts' confidence in these systems are, however, low [10]. Future users need independent research with this software to gain confidence. An issue in validating such algorithms is the lack of a gold standard in EEG review, mainly because of large interrater and intrarater variability seen in identifying spikes or sharp waves in the same EEG recording [11]. Factors that play a role are, for example, reader style, fatigue, and loss of concentration. The lack of an objective gold standard creates difficulties in assessing whether a detection algorithm is performing well [12]. The Persyst 13 (P13) is one of the available software packages for EEG visualization that has an automated spike and seizure detection feature. The spike detection algorithm is a neural network that attempts to mimic the perception-based marking of human experts (HEs) [13,14]. For users, the precise details of the algorithm and the neural network rules are mostly unknown except for some technical aspects [14]. The algorithm uses different sensitivity settings to present the output; these are denoted as perception values, ranging from zero to one. Ambiguous epileptiform features are assigned nearzero values, and clear epileptiform abnormalities are assigned nearone values [13]. P13 has three different settings: "high," "medium," and "low." The "high" setting has a perception threshold setting of 0.1, the "medium" of 0.4, and the "low" of 0.9. Counting spikes manually is a time-consuming task [15]. Estimating the SWI using automated detection software could save time, as a detection algorithm is able to calculate a SWI in few seconds (after the record is processed). This is independent of the size of the epoch. Thus, the time saved is larger when reviewing longer EEG recordings. Thereby, an automated detection algorithm will consistently detect a certain EEG phenomenon as

epileptiform and is not influenced by individual reader style or other reader factors such as fatigue. A recent report found that the software-calculated SWI using P13 was noninferior to experts' estimates [16]. This report, however, was based on a small number of nonheterogeneous recordings from ESES patients. Thereby, the "high" perception setting was used instead of the "medium" setting, which the Persyst Development Corporation states is the default mode. Furthermore, this report did not provide information about accurate quantification of lower SWIs, which can be useful for follow-up of patients. The algorithm, therefore, needs further validation. In this study, we compared the performance of the P13 algorithm versus HE consensus in a heterogeneous set of recordings, reviewing all three perception value settings ('high', 'medium' or 'low').

Method

All prolonged EEG recordings made with an ESES or follow-up of ESES referral question in children or teenagers (age 0–18 years) between January 1, 2018 and December 31, 2018 were included. Informed consent was not obtained because of the study's retrospective nature. Thereby, only anonymized data, and no video data, were used. This study was approved by the institutional review board. The HEs were a clinical neurophysiologist and a physician assistant each with more than five years of experience in reviewing EEGs. The education of this particular physician assistant contained multiple years of medical training combined with dedicated EEG training, supervised by board-certified clinical neurophysiologists. The HEs were masked to the initial video-EEG monitoring report. The two HEs reviewed the EEGs together and only viewed the first 10 minutes of NREM sleep (starting point at 50% decrease of posterior dominant rhythm, appearance of lateral eye movements or drowsiness, and/or vertex waves). They estimated a SWI for each recording defined as the average percentage of each 1-second epoch containing the sharp component of an epileptiform discharge. Interictal epileptiform discharges were defined as paroxysmal, sharply contoured, wave forms, clearly distinguished from the background activity, had a field, and a duration of less than 200 milliseconds [17]. In the SWI estimation, both generalized and focal discharges were included. Both experts had to agree on the presence of the interictal epileptiform discharge for it to be counted. The SWI was estimated without explicit time constraints, and the EEG traces could be reformatted as in the clinical setting. All EEGs were reviewed with the SystemPLUS Evolution software (Micromed, Veneto, Italy) using standard 10 to 20 International electrode recording and 256 Hz sample frequency. The time of the manual count by the HEs was measured for each EEG record. For the automated spike detection, we used the P13 software (Persyst Development Corporation, San Diego, CA). The SWI was calculated using all three different perception settings. An SWI calculated by the detection software was also defined as the average percentage of each 1-second epoch that contained an epileptiform discharge. Continuous variables were analyzed using the Wilcoxon signed rank

test for nonparametric data using SPSS (IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY).

Results

A total of 48 recordings from 44 patients (24 male) were identified. The mean patient age was 7.8 years (SD 2.4 years; range, 3–11 years). Human experts estimated an SWI in a median time of 4 minutes 54 seconds (range, 30 seconds–14 minutes 37 seconds) per record. According to the HEs, 28 recordings included spikes. The SWIs estimated by the HEs did not significantly differ from the SWI calculations of the algorithm in the 'low' perception settings (Tables 1 and 2).

Table 1. Median SWI (in %) of all recordings (N=48)

	Median SWI in %	Range (%)		Percentile		difference from HE (p-value)
		min	max	Q1	Q3	
HE consensus	18	0	99	0	80	
P13 (low[1])	16	0	96	1	78	p=0.67
P13 (median[1])	28	1	98	6	82	p=0.000
P13 (high[1])	36	3	99	36	83	p=0.000

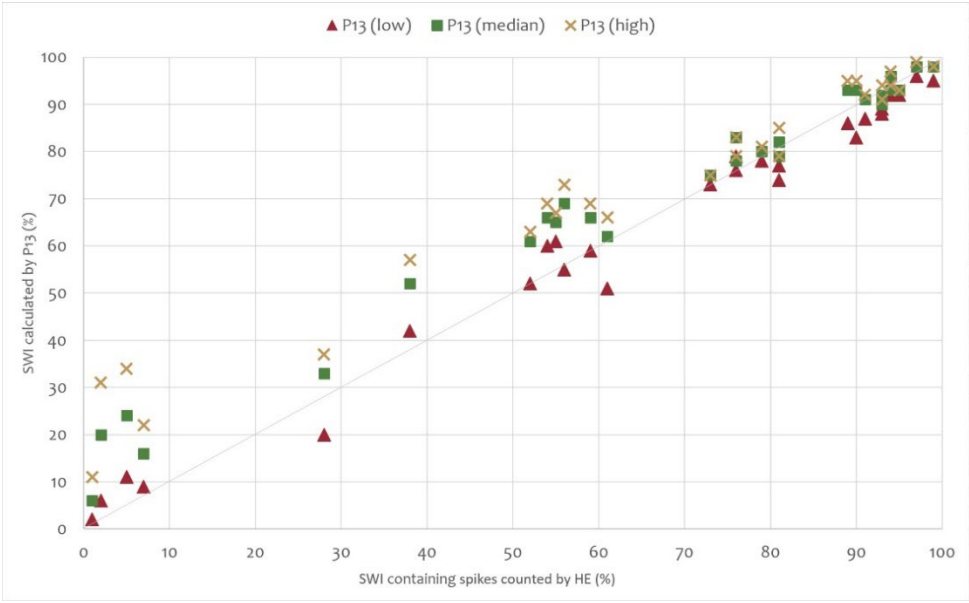
[1] *sensitivity setting, HE = human experts, P13 = Persyst 13 spike detection*

Table 2. Median SWI (in %) of recordings containing spikes (N=28)

	Median SWI in %	Range (%)		Percentile		difference from HE (p-value)
		min	max	Q1	Q3	
HE consensus	76	1	99	52	92	
P13 (low[1])	75	2	96	51	88	p=0.19
P13 (median[1])	79	6	98	61	93	p=0.001
P13 (high[1])	79	11	99	64	94	p=0.000

[1] *sensitivity setting, HE = human experts, P13 = Persyst 13 spike detection*

The SWI estimated by the HEs differed significantly from the 'medium' perception settings and the 'high' perception settings. The SWIs calculated in these modes were higher than the HE-estimated SWIs. The largest difference in calculated SWI within one subject in the 'low' perception setting was 10% (the P13 algorithm calculated 51% vs. 61% for the HEs). The largest difference in calculated SWIs within an individual between the P13 algorithm in the 'medium' setting was 18% and in the 'high' perception setting was 29%. The differences between SWIs calculated by HEs and the three perception settings were, in most cases, smaller for the higher SWIs (especially above 70%) than in the lower SWIs. This is shown in Figure 1.



HE = human experts; P13 = Persyst 13 spike detection; SWI = spike–wave index.

Figure 1. SWI Calculated by HEs and P13.

Based on the SWIs estimated by the HEs, 22 recordings met the ESES criteria of ≥ 50% of 1-second epochs containing spikes (Table 3). All were also identified with an SWI ≥ 50% by the P13 algorithm using the 'low' setting: thus, sensitivity was 100% (confidence interval, 82%–100%) and specificity was also 100% (confidence interval, 52%–100%). In one recording, the algorithm in 'medium' and 'high' settings calculated a SWI ≥ 50%, where the HEs calculated a SWI < 50%: thus, in 'medium' and 'high' perceptions settings, the sensitivity is 100% (confidence interval, 82%–100%) and the specificity is 83% (confidence interval, 36%–99%). No spikes were seen by HEs in 20 recordings. The algorithm, however, detected spikes in most of these recordings. It calculated SWIs ranging from 0% to 6% in the 'low' setting, from 1% to 20% in the 'medium' setting, and from 3% to 33% in the 'high' setting.

Table 3. Number of records meeting ESES-criteria, calculated by HEs and by P13

	P13 (low[1])		P13 (medium and high[1])	
	P13 ≥ 50%	**P13 < 50%**	**P13 ≥ 50%**	**P13 < 50%**
HE >=50%	22	0	22	0
HE >0% and <50%	0	6	1	5

[1] *sensitivity setting, HE = human experts, P13 = Persyst 13 spike detection*

Discussion

We showed that calculating SWI using the spike detection algorithm P13 in the 'low' perception setting is non-inferior to estimating SWI by HEs. The perception setting matters especially in the lower SWIs because the differences between the settings are small in the higher SWIs. We also showed that using the spike detection software may save HEs time in comparison with human estimation. Thereby, the software makes it easy to estimate SWI for larger epochs, such as a first sleep cycle or even a whole night. Another advantage is that the algorithm will always detect the same event as epileptiform and thus eliminate human factors such as reader style or fatigue. There are limitations to our study. We tried to generate a heterogeneous dataset with SWIs in all ranges. There were, however, few recordings with an SWI around the cut-off point of 50%. At group level, the SWI estimated by HEs and the P13 in 'low' perception setting did not differ, but we did see some individual differences between the calculated SWIs. In practice, this can mean the difference in reaching or not reaching the criterion of at least 50% epileptiform activity. However, in ESES-related syndromes, the SWI is only part of the diagnostic criteria, as the clinical symptoms are also taken into account. Thereby, the SWI criterion of 50% is arbitrary. Another issue when testing the reliability of a spike detection algorithm is that the interrater agreement between EEG reviewers is low, so our HE estimated SWI is not the gold standard. We approached this by estimating the SWI in consensus, instead of using a single individual to estimate SWIs (which is current practice in our center). A disadvantage of using this algorithm is that it has false detections, usually sharp physiologic sleep phenomena, especially K-complexes. This was especially noticeable in (near-) normal EEG recordings. The P13 calculated SWI of these normal EEGs is up to 6% in 'low' perception settings and up to 33% in 'high' perception settings and reviewers must always be aware of this especially when reviewing EEGs of children, who often have sharp sleep phenomena. Spike detection software is an useful tool in obtaining SWI and can help reducing the burden of manual estimation. Further validation of the software is needed in larger cohorts, multiple centers, and by multiple HEs.

References

1. Tassinari CA, Rubboli G, Volpi L, Meletti S, D'Orsi G, Franca M, et al. Encephalopathy with electrical status epilepticus during slow sleep or ESES syndrome including the acquired aphasia. Clin Neurophysiol 2000;111:S94–S102.

2. Patry G, Lyagoubi S, Tassinari A. Subclinical "electrical status epilepticus" induced by sleep in children. JAMA Neurol 1971;24:242– 52.

3. Commission on Classification and Terminology of the International League Against Epilepsy. Proposal for revised classification of epilepsies and epileptic syndromes. Epilepsia 1989;30:389–99.

4. Lodderkemper T, Sánchez Fernández I, Peters JM. Continuous spike and waves during sleep and electrical status epilepticus in sleep. J Clin Neurophysiol 2011;28:154–64. https://doi.org/10.1097/WNP.0b013e31821213eb.

5. Sánchez Fernández I, Lodderkemper T, Peters JM. Electrical status epilepticus in sleep: clinical presentation and pathophysiology. Pediatr Neurol 2012;47:390–410. https://doi.org/10.1016/j.pediatrneurol.2012.06.016.

6. Scheltens-de Boer M. Guidelines for EEG in encephalopathy related to ESES/CSWS in children. Epilepsia 2009;50:13–7. https://doi.org/10.1111/j.1528-1167.2009.02211.x.

7. Sánchez Fernández I, Chapman KE, Peters J, Kothare SV, Nordli Jr. DR, Jensen FE, et al. The tower of Babel: survey on concepts and terminology in electrical status epilepticus in sleep (ESES) and continuous spikes and waves during sleep (CSWS) in North America. Epilepsia 2013;54:741–50. https://doi.org/10.1111/epi.12039.

8. Halford JJ. Computerized epileptiform transient detection in the scalp electroencephalogram: obstacles to progress and the example of computerized ECG interpretation. Clin Neurophysiol 2018;120:1909– 15. https://doi.org/10.1016/j.clinph.2009.08.007.

9. Wilson SB, Emerson R. Spike detection: a review and comparison of algorithms. Clin Neurophysiol 2002;113:1873–81.

10. Webber WRS, Lesser RP. Automated spike detection in EEG. Clin Neurophysiol 2017;35:241–2. https://doi.org/10.1016/j.clinph.2016.11.018.

11. Grant AC, Abdel-Baki SG, Weedon J, Arnedo V, Chari G, Koziorynska E, et al. EEG interpretation reliability and interpreter confidence: a large single center study. Epilepsy Behav 2014;32:102–7. https://doi.org/10.1016/j.yebeh.2014.01.011.

12. Webber WR, Litt B, Lesser RP, et al. Automatic EEG spike detection: what should the computer imitate? Electroencephalogr Clin Neurophysiol 1993;87:364–73.

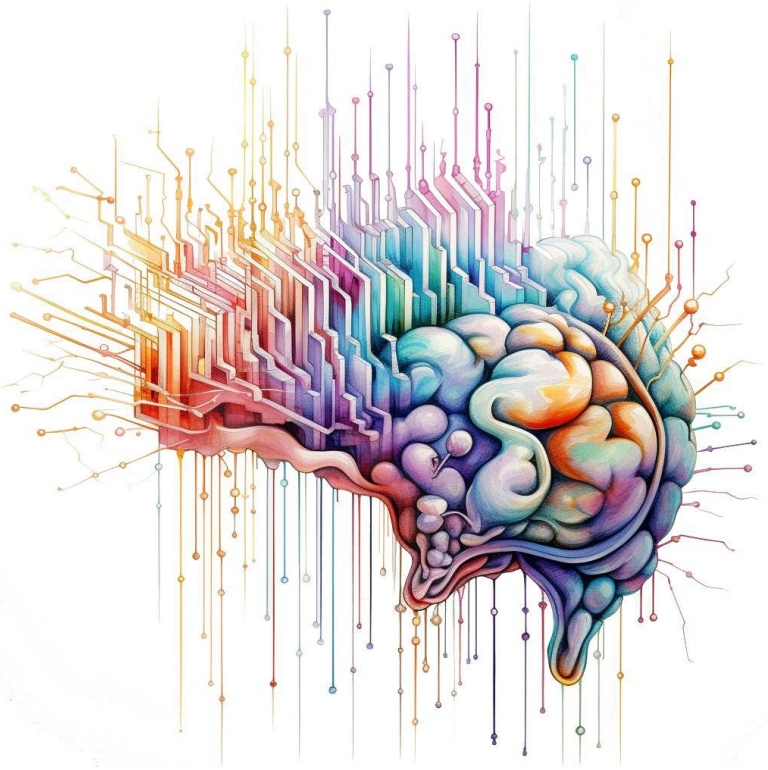
13. Wilson SB, Turner CA, Emerson RG, et al. Spike detection II: automatic, perception-based detection and clustering. Clin Neurophysiol 1999;110:404–11.

14. Scheuer ML, Bagic A, Wilson SB. Spike detection: inter-reader agreement and a statistical Turing test on a large data set. Clin Neurophysiol 2016;128:243–50. https://doi.org/10.1016/j.clinph.2016.11.005.

15. Weber AB, Albert DV, Yin H, Held TP, Patel AD. Diagnosis of electrical status epilepticus during slow-wave sleep with 100 seconds of sleep. J Clin Neurophysiol 2017;34:65–8. https://doi.org/10.1097/WNP.0000000000000307.

16. Joshi CN, Chapman KE, Bear JJ, Wilson SB, Walleigh DJ, Scheurer MM. Semiautomated spike detection software Persyst 13 is noninferior to human readers when calculating the spike-wave index in electrical status epilepticus in sleep. J Clin Neurophysiol 2018;35:370–4. https://doi.org/10.1097/WNP.0000000000000493.

17. Pedley T, Mendiratta A, Walczak T. Chapter 17: seizures and epilepsy. In: Current practice of clinical electroencephalography. Philadelphia: Lippincott Williams & Wilkins, 2003; 512–



CHAPTER 3

Using sampled visual EEG review

in combination with automated detection software

at the EMU

E.E.M. Reus
G.H. Visser
F.M.E. Cox

Seizure. 2020;80:96-9.
https://doi. org/10.1016/j.seizure.2020.06.002

Chapter 3 - Using sampled visual EEG review in combination with automated detection software at the EMU

Abstract

Purpose: Complete visual review of prolonged video-EEG recordings at an EMU (Epilepsy Monitoring Unit) is time consuming and can cause problems in times of paucity of educated personnel. In this study we aimed to show non inferiority for electroclinical diagnosis using sampled review in combination with EEG analysis software (P13 software, Persyst Corporation), in comparison to complete visual review.

Method: Fifty prolonged video-EEG recordings in adults were prospectively evaluated using sampled visual EEG review in combination with automated detection software of the complete EEG record. Visually assessed samples consisted of one hour during wakefulness, one hour during sleep, half an hour of wakefulness after wake-up and all clinical events marked by the individual and/or nurses. The final electro-clinical diagnosis of this new review approach was compared with the electro-clinical diagnosis after complete visual review as presently used.

Results: The electro-clinical diagnosis based on sampled visual review combined with automated detection software did not differ from the diagnosis based on complete visual review. Furthermore, the detection software was able to detect all records containing epileptiform abnormalities and epileptic seizures.

Conclusion: Sampled visual review in combination with automated detection using Persyst 13 is non-inferior to complete visual review for electroclinical diagnosis of prolonged video-EEG at an EMU setting, which makes this approach promising.

Introduction

EEG is an important tool in the management of epilepsy. Interictal and ictal findings can help in epilepsy diagnosis, seizure and syndrome classification, epilepsy monitoring and for identifying surgical candidates [1,2].

At an Epilepsy Monitoring Unit (EMU) prolonged video-EEGs are performed, resulting in large datasets. At our centre, technicians visually review the entire EEG recording, aiming at finding all relevant interictal and ictal events for answering the referral question. Subsequently, a clinical neurophysiologist reviews selections made by the technician and provides a final electro-clinical diagnosis. The complete visual data analysis is time-consuming and costly. In times where there is a paucity of technicians, this can cause problems. It is necessary to look for time saving alternatives to review large EEG datasets, without loss of quality.

One approach for saving time is sampled visual review. This approach has hardly been evaluated. One study suggested that the first hour of sleep reliably predicts the occurrence of interictal epileptiform activity for whole recording [3]. Another study showed that sampled review was non-inferior regarding final electro-clinical diagnosis, although a substantial number of events was missed [4].

Another approach is automated EEG analysis, using detection software. These software packages are widely used in ICU settings, but to our knowledge not commonly used at EMUs. There are several reports on automated detection algorithms, although most focus on the algorithm development rather than clinical validation [5]. The Persyst 13 (P13) spike detector is a commercially available software frequently used in the assessment of automated detection software [6,7,8,9]. Two studies showed P13 was non-inferior to human mark-up when detecting interictal epileptiform abnormalities [6,7]. Two other studies looked at ictal events. The first study found P13 only correctly identified at least one electrographic seizure in 53% of ambulatory recordings [8]. The second showed a previous Persyst version (version 11) detected 76% of electrographic seizure at an EMU setting, but missed most of the myoclonic and focal aware seizures [9].

Our main objective is to determine whether a review approach using a combination of sampled visual review and automated detection software is non-inferior to the conventional method, where the entire EEG is visually reviewed.

Method

Fifty prolonged video-EEG recordings between November 2018 and May 2019 were prospectively included. The only inclusion criterion was a minimum age of 16 years. We excluded presurgical recordings. The analyses of these records were embedded in our usual

clinical workflow, so informed consent was not obtained in accordance to local Ethics guideline.

All recordings were performed at our 8-bedded EMU [10]. A Micromed EEG system (Micromed, Mogliano Veneto, Italy) using the standard 10-20 International electrode placement plus F9/F10 and 256 Hz sample frequency was used. During daytime and early evening, individuals were observed by three nurses, positioned at a nearby video footage observation station; at night, two nurses carried out the task [10, 11]. During the recording individuals were asked to press the button when experiencing a clinical event.

Sampled EEG

The sampled EEG contained the first hour of recording during wakefulness including hyperventilation provocation and intermittent photic stimulation, the first hour of sleep and the first half hour after sleep the next morning, which was added due to the circadian distribution of some generalized epilepsies (especially JME). In addition, EEG and video periods around nurse or patient 'push button' marked events were reviewed.

Automated detection software

The P13 software from Persyst Corporation (P13) was used for the automated spike and seizure detection. Spike detections are clustered per electrode, where maximum amplitude was recorded. It generates 1-second epochs centred around the detected spike, with an average signal per electrode. All single potentially abnormal findings can also be reviewed. Further details can be found in previous literature [12, 13]. Regarding seizure detection, the manual states that the algorithm detects ictal patterns with a minimum duration of 11 seconds. The Persyst software also includes various options for quantitative EEG trends.

Research protocol

The human experts (HEs) were pairs, in varying combinations, of two clinical neurophysiologists and a physician assistant, all with more than five years of video-EEG reviewing experience. To reduce steep learning effects, the HEs had already practiced using the P13 software before the first review for the study.

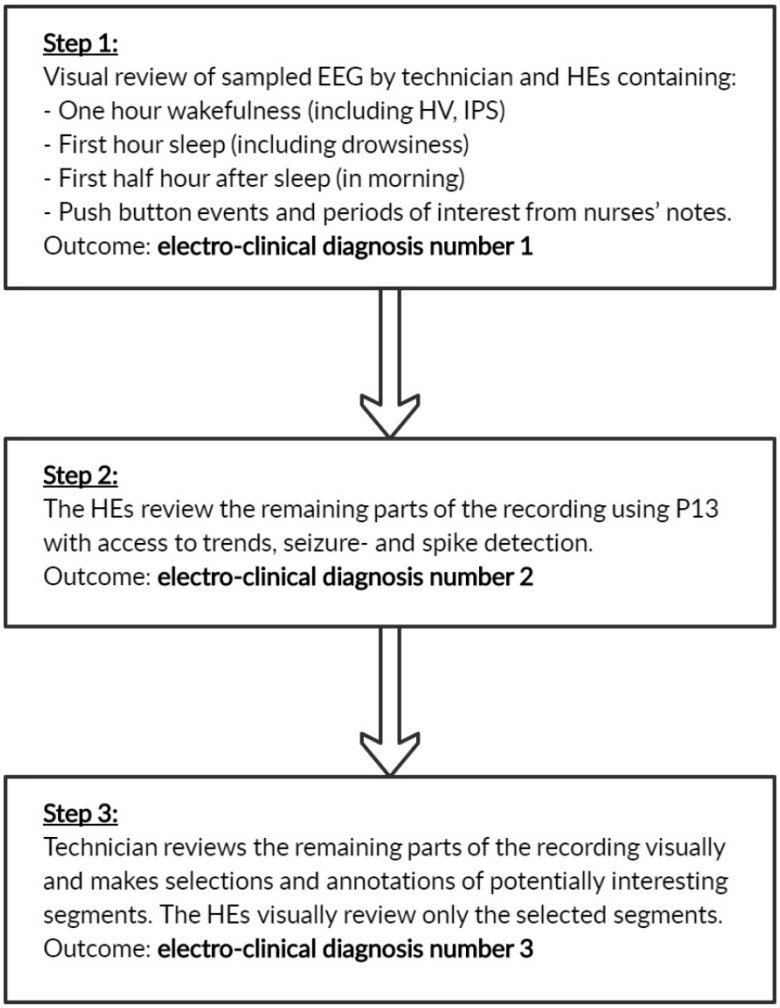
The reviewing process consisted of three steps (Figure 1). As in our normal routine, EEGs were pre-reviewed by EEG technicians. They first made a report of the sampled EEG in step 1, subsequently they reviewed the whole EEG record, and documented additional findings for step 3. For step 1 as well as step 3, video is only reviewed in the period that (1) nurses note a

(possible) event, (2) patients use the push button, (3) technicians see a (possible) ictal EEG rhythm and (4) to distinguish an artefact from cerebral activity.

Technicians were blinded for the results of step 2 during the whole reviewing process. The HEs first reviewed the sampled EEG in step 1, afterwards reviewed the whole EEG using automated software (step 2), and finally reviewed technicians additional findings in step 3. As with the technicians, the video is only reviewed by the HE's in the above-mentioned periods. After each step the HEs formed an electro-clinical diagnosis, using SCORE terminology [14]. Possible epilepsy was used when only a few, or only ambiguous, interictal epileptiform discharges (IEDs) were seen. Probable epilepsy was used when definite interictal epileptiform activity was seen. Definite epilepsy was used when a record contained one or multiple seizures with electro-encephalographic correlate. The electro-clinical diagnosis in step 3 was regarded as the current best available gold standard. Furthermore, HEs described the ictal and interictal findings at each step.

Outcome measures

The primary outcome measure was the difference between the electro-clinical diagnoses in sampled visual review with automated detection (step 2) and the electro-clinical diagnoses in complete visual review (step 3). Secondary outcome measures were the electro-clinical diagnosis for sampled visual review only (step 1), compared to step 3. Furthermore we looked at the occurrence of IEDs, missed seizures by P13 in step 2, false seizure detections by P13 in step 2 and missed seizures by technicians in step 3.



HE = human expert; HV = hyperventilation provocation; IPS = intermittent photic stimulation, P13 = Persyst 13.

Figure 1. Research protocol

Results

A total of 1170 hours of video-EEG from 50 records (18 males and 32 females with median age 32 (range 18 – 73) years) was analysed with a median duration of 23.4 hours (range 17.5 – 44.5 hours). The clinical referral question for the EEGs was presence of interictal epileptiform findings (to support diagnosis or follow-up) in 32 records, classification of epilepsy in 10 records, and event recording in 8 recordings for diagnostic or classification purposes.

Electro-clinical diagnoses

Table 1 shows the various reported electro-clinical diagnoses in each step. In none of the cases the electro-clinical diagnosis in step 2 differed from the electro-clinical diagnosis in step 3.

The electro-clinical diagnosis reported in step 1 differed from step 2 in three cases (6%). In one patient the report changed from aspecific focal dysfunction to probable focal epilepsy, the second changed from normal to possible focal epilepsy, and the third changed from possible to definite focal epilepsy, because a seizure was detected by P13 outside the sampled EEG.

Interictal epileptiform abnormalities

Interictal epileptiform abnormalities were present in 29 records (58%) according to the HEs (step 3). P13 detected all records containing epileptiform activity. One record also contained TIRDA (Temporal Intermittent Rhythmic Delta Activity), which was not detected by the P13 software and also not seen in step 1. No other important interictal findings were missed by P13. There was no difference in localization and frequency of interictal (epileptiform) activity between step 2 and step 3.

Seizures

In 16 records (32%) at least one clinical event occurred (range 1 – 58). In 6 of these records the events were classified as epileptic seizures (5 with focal seizures and 1 with generalized seizures; Table 2). In 5 of these 16 records the events were classified as PNES and in the remaining 5 records as subjective events with uncertain etiology.

P13 alone, detected all records containing seizures. Although in 3 records a part of the seizures were missed. In the record containing 58 absences, P13 missed five of them. Of those five, three absences were shorter than 11 seconds and two were longer than 11 seconds. The other two records contained nocturnal frontal seizures with no or subtle EEG changes (only muscle and movement artefacts).

In 27 records (54%) false seizure detections by P13 occurred, with a total of 81 false detections with a median of two per record (range 1 – 19). Most of these false detections occurred during chewing or rhythmic eye blinking and were easily recognised.

One focal seizure with impaired awareness was missed by the technician in step 3. This seizure was detected by P13 in step 2.

Table 1. Diagnostic significance

		Step 1	Step 2	Step 3
Normal		9	8*	8
No definite abnormality		6	6	6
Focal dysfunction		5	4*	4
Diffuse dysfunction		0	0	0
Focal epilepsy	*Possible*	1	1*	1
	Probable	10	11*	11
	Definite	4	5*	5
Multifocal epilepsy	*Possible*	2	2	2
	Probable	1	1	1
	Definite	0	0	0
Generalized epilepsy	*Possible*	0	0	0
	Probable	5	5	5
	Definite	1	1	1
Epilepsy no classification		1	1	1
PNES		5	5	5
Non epileptic otherwise		0	0	0

*Normal = Normal interictal EEG and no events recorded, PNES = psychogenic non epileptic seizures, *= in one record the electro-clinical diagnosis changed*

Table 2. Detected seizures from the 6 records containing definite epileptic events

			Step 1			Step 2	Step 3
Case	**Total number seizures**	**Seizure type**	**Number occurred in sampled EEG part**	**Number detected by PB**	**Number detected by nurse**	**Number detected by P13**	**Number detected only by technician**
1	58	Absence	13	5	16*	53	1
2	14	FAS	0	1	12	6	1
3	1	FIAS	0	0	1	1	0
4	1	FIAS	0	0	1	1	0
5	5	FAS	2	1	4	1	0
6	1	FIAS	0	0	0	1	0

Part 1 = first hour wakefulness, first hour sleep and first half hour after sleep (in morning),
PB = push button, FAS = focal aware seizure, FIAS = focal seizure with impaired awareness
***** *common practice for the nurse is to stop responding to absences after > 10 seizures.*

Discussion

In this study we showed that the electro-clinical diagnosis after sampled visual EEG review in combination with automated detection software (P13) did not change after successive complete visual review. The advantage of this reviewing approach is that it may substantially save overall reviewing time, especially in our setting with many prolonged EEG recordings.

We also showed that our approach of sampled visual review alone is insufficient, both with respect to interictal findings as for missing seizures.

Relying on automated detection software alone is also not possible. First, sampled visual review remains necessary to get an impression of the background activity and potential (focal or diffuse) dysfunctions or rhythmic delta activity (e.g. TIRDA). We showed that reviewing an hour when awake, an hour when asleep and an hour after wakefulness is enough for this purpose. Second, automated software packages are likely to miss seizures with no or just subtle EEG correlate. This makes the software less suitable for diagnosis or monitoring very short seizures (i.g. myoclonia) and focal aware seizures, which was also shown by previous literature [9]. Therefor additional observations of nurses and markings of patients experiencing clinical events will stay required.

The aim of the study was to show non-inferiority to the human observer. An interesting observation is the seizure missed by the technician, but detected by the P13 software, showing the potential superiority of software above the human observer. This is probably not exceptional, where technicians are requested to review prolonged records at a much faster

speed than real-time risking missing relevant events [15]. Studies should be designed not only to show non-inferiority of detection software, but also to enable the software to "beat" the human observer as gold standard.

Our proposed approach can be used for review of prolonged records at the EMU, since we showed non-inferiority. And although we did not investigated it formally, this approach is very likely to result in time gain, provided that the elaboration of the detection software results does not require extra time. A disadvantage of the P13 is that it has false detections, interictal as well as ictal. Reviewers must be able to filter these out. It would be very helpful if all detected findings are merged in clusters based on similar properties like morphology and localization, that reliably reflect the same functional EEG abnormality. Then groups of artefacts or nonspecific abnormalities could be disregarded, without the need to check all the individual detections. In the P13 software this is currently not possible.

Our study has some limitations. We tried to include a heterogenic group with focal epilepsies, generalized epilepsies and PNES but our sample size was too small to include all different types of epilepsies and seizures. In the 3-step-process the HEs were not blinded to their previous reports. Although this is suboptimal, it prevents the known problems of interrater disagreement [16]. Finally this single centre bias also limits overall generalizability, as monitoring methods, staff expertise, and training vary widely among epilepsy centres. What may be a good set-up in our centre may not be useful or feasible for others. This especially applies for settings with no or minimal nurse observation, such as ambulatory settings.

Further validation of the software is warranted in larger cohorts, multiple centres and by multiple human experts. A progressive approach would be a design combining further validation during implementing of supporting automated software, stepwise reducing the required EEG review time. We think it's feasible to make a step toward using more medical technology in EEG reviewing.

Sampled visual review in combination with automated detection software is a promising time gaining tool in reviewing prolonged video-EEGs of adults at the EMU. It thereby remains warranted that clinical events are continuously observed by trained nurses and patients have the possibility to use a push button when experiencing seizure-like symptoms.

References

1. Wirrell ED. Prognostic significance of interictal epileptiform discharges in newly diagnosed seizure disorders. J Clin Neurophysiol. 2010;27:239-48. https://doi.org/10.1097/WNP.0b013e3181ea4288.

2. Shin HW, Pennell PB, Lee JW, et al. Efficacy of safety signals in the epilepsy monitoring unit (EMU): Should we worry? Epilepsy & Behavior. 2012;23:458–61. https://doi.org/10.1016/j.yebeh.2012.01.018.

3. Liu X, Issab NP, Roseb S, et al. The first-hour-of-the-day sleep EEG reliably identifies interictal epileptiform discharges during long-term video-EEG monitoring. Seizure. 2018;63: 48-51. https://doi.org/10.1016/j.seizure.2018.10.015.

4. Badawy RAB, Pillay N, Jetté N, Wiebe S, Federico P. A blinded comparison of continuous versus sampled review of video-EEG monitoring data. Clin Neurophysiol. 2011;122:1086-90. https://doi.org/10.1016/j.clinph.2010.10.048.

5. Halford JJ. Computerized epileptiform transient detection in the scalp electroencephalogram: Obstacles to progress and the example of computerized ECG interpretation. Clin Neurophysiol. 2009;120:1909-15. https://doi.org/10.1016/j.clinph.2009.08.007.

6. Scheuer ML, Bagic A, Wilson SB. Spike detection: Inter-reader agreement and a statistical Turing test on a large data set. Clin Neurophysiol. 2016;128:243-50. https://doi.org/10.1016/j.clinph.2016.11.005.

7. Halford JJ, Westover MB, LaRoche SM, et al. Interictal epileptiform discharge detection in EEG in different practice settings. J Clin Neurophysiol. 2018;35:375–80. https://doi.org/10.1097/WNP.0000000000000492.

8. González Otárula KA, Milhaeil-Demo Y, et al. Automated seizure detection accuracy for ambulatory EEG recordings. Neurology. 2019;92,e1-7. https://doi.org/10.1212/WNL.0000000000007237.

9. Kamitakia BK, Yumb A, Leea J, et al. Yield of conventional and automated seizure detection methods in the epilepsy monitoring unit. 2019. Seizure. 2019;69:290-5. https://doi.org/10.1016/j.seizure.2019.05.019.

10. Cox FME, Reus EEM, Widman G, Zwemmer JNP, Visser GH. Epilepsy Monitoring Units can be safe places; a prospective study in a large cohort. Epilepsy Behav. 2020;120:102-6. https://doi.org/10.1016/j.yebeh.2019.106718

11. Rommens N, Geertsema E, Jansen Holleboom L, et al. Improving staff response to seizures on the epilepsy monitoring unit with online EEG seizure detection algorithms. Epilepsy Behav. 2018;84:99–104. https://doi.org/10.1016/j.yebeh.2018.04.026.

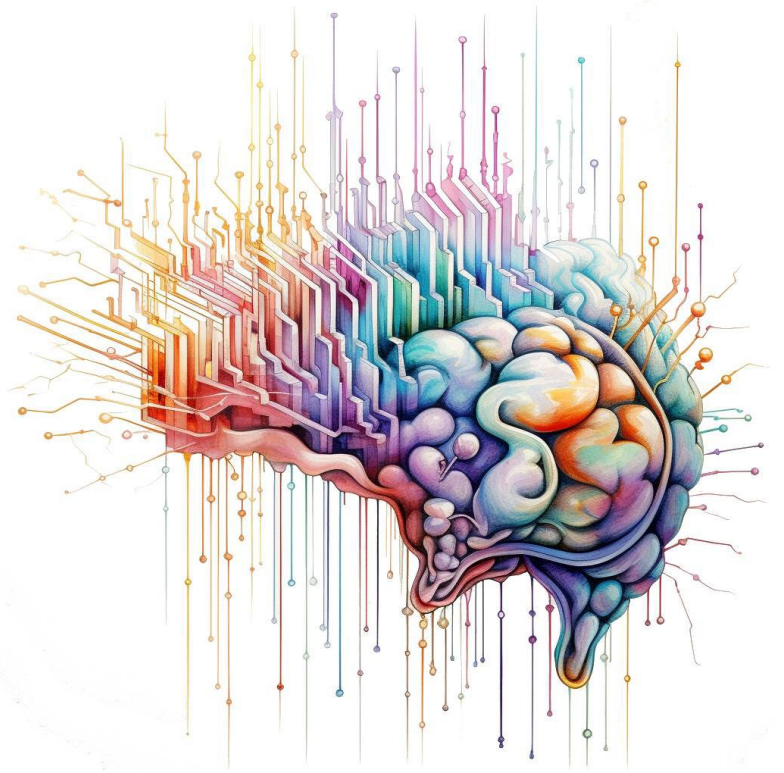
12. Wilson SB, Turner CA, Emerson RG, et al. Spike detection II: automatic, perception-based detection and clustering. Clinical Neurophysiology. 1999;110:404-11. https://doi.org/10.1016/S1388-2457(98)00023-6.

13. Reus EEM, Visser GH, Cox FME. Determining the Spike–Wave Index Using Automated Detection Software. Clin Neurophysiol. 2019; in press. https://doi.org/10.1097/WNP.0000000000000672

14. Beniczky S, Aurlien H, Brøgger, JC, et al. Standardized computer-based organized reporting of EEG: SCORE - Second version. Clin Neurophysiol. 2017;128:2334–46. http://dx.doi.org/10.1016/j.clinph.2017.07.418.

15. Halford JJ, Shiau D, Kern RT, et al. Seizure detection software used to complement the visual screening process for long-Term EEG monitoring. Am J Electroneurodiagnostic Technol. 2010;50:133–47.

16. Bagheri E, Dauwels J, Dean BC, et al. Interictal epileptiform discharge characteristics underlying expert interrater agreement. Clin Neurophysiol. 2017;128:1994–2005. http://dx.doi.org/10.1016/j.clinph.2017.06.252.



CHAPTER 4

Automated spike detection:

which software package?

E.E.M. Reus
F.M.E. Cox
J.G. van Dijk
G.H. Visser

Seizure 2022;95: 33–7.
https://doi.org/10.1016/j.seizure.2021.12.012

Chapter 4 – Automated spike detection: which software package?

Abstract

Purpose: We assessed three commercial automated spike detection software packages (Persyst, Encevis and BESA) to see which had the best performance.

Method: Thirty prolonged EEG records from people aged at least 16 years were collected and 30-minute representative epochs were selected. Interictal epileptiform discharges (IEDs) were marked by three human experts and by all three software packages. For each 30-minutes selection and for each 10-second epoch we measured whether or not IEDs had occurred. We defined the gold standard as the combined detections of the experts. Kappa scores, sensitivity and specificity were estimated for each software package.

Results: Sensitivity for Persyst in the default setting was 95% for 30-minute selections and 82% for 10-second epochs. Sensitivity for Encevis was 86% (30-minute selections) and 61% (10-second epochs). The specificity for both packages was 88% for 30-minute selections and 96%-99% for the 10-second epochs. Interrater agreement between Persyst and Encevis and the experts was similar than between experts (0.67-0.83 versus 0.63-0.67). Sensitivity for BESA was 40% and specificity 100%. Interrater agreement (0.25) was low.

Conclusion: IED detection by the Persyst automated software is better than the Encevis and BESA packages, and similar to human review, when reviewing 30-minute selections and 10-second epochs. This findings may help prospective users choose a software package.

Introduction

Detecting interictal epileptiform discharges (IEDs) is helpful in epilepsy diagnosis [1,2]. Prolonged EEG recordings improve the chances of finding interictal activity, yielding higher diagnostic efficiency, but also needing more review time [3,4]. The typical procedure for reviewing a prolonged EEG record involves human scrutiny of the complete record, which is time-consuming.

Automated detection software might decrease review time while ensuring a high yield, but expert confidence in using such software is low [5]. This may partly be due to presumed high false-positive and false-negative rates, resulting from variable IED morphology and similarity to normal EEG activity or artefacts [5,6]. Automated detection software may serve as a screening tool to reduce the need for comprehensive visual review, provided it is sufficiently reliable.

In a pilot study using only the Persyst software, we showed that the diagnostic yield was sufficient to be used as a substitute for a complete visual review of prolonged recordings [7]. The present study compares three different commercial software packages on a different heterogeneous dataset containing IEDs.

Method

EEG data

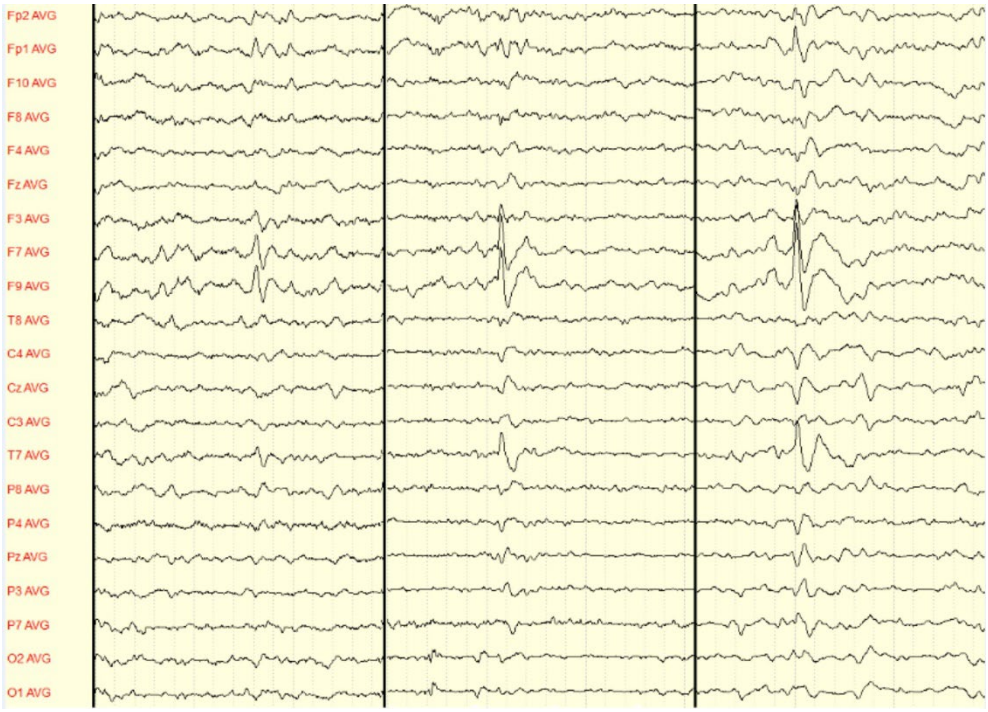
Two human experts not participating in the marking process retrospectively screened recordings from 108 people, made at our EMU (Epilepsy Monitoring Unit) between August and September 2019); 43 were from people younger than 16 years and therefore excluded. A study set of 30 records was created from the remaining 65 records using an online randomization tool [8]. A 30-minute section of the EEG during wakefulness, deemed representative of the entire wake EEG, was selected for analysis.

EEG data were recorded using the Micromed EEG system (Micromed, Mogliano Veneto, Italy), using the standard 10-20 international electrode recording and additional F9/F10 positions sampled at 256 Hz. The local medical ethics committee approved this study. As we recorded the EEGs exclusively in clinical care, the need for informed consent was waived according to Dutch rules.

Human detections

Three human experts independently marked all IEDs in the EEGs. Two were clinical neurophysiologists and one was a physician assistant. All had more than five years' EEG reviewing experience at a dedicated epilepsy center. The experts had no prior knowledge of

nor access to the clinical information or the EEG report. Each expert reviewed all thirty 30-minute selections using the Micromed software; they could use all available montages. To identify IEDs, the criteria of the International Federation of Clinical Neurophysiology were used [9]. The experts were instructed to annotate the most negative point of each IED and to classify them as possible or definite IED (Figure 1).



On the left side is a possible IED. On the middle and right side are definite IEDs. IEDs are shown in an average reference montage, but reviewers could use any montage. IED = Interictal Epileptiform Discharge

Figure 1. Examples of IEDs

Automated IED detection

We used three software packages: Persyst (Persyst Development Corporation, USA), Encevis (AIT Austrian Institute of Technology, Austria) and BESA (BESA Epilepsy, Germany).

Persyst

Persyst version 14 has three sensitivity settings for IED detection. We used the default medium setting and the low setting. The low setting emerged as optimal in previous research [7,10,11]. The output is a list of timed IED detections per electrode.

Encevis

We used the Epispike module in the Encevis software, version 1.9.2. There are no sensitivity settings. The output is a list of timed IED detections per electrode [12].

BESA Epilepsy

The BESA software (version 2.0) uses a different method. Instead of showing detections as distinct events, it offer the detections as clusters. These clusters are made for every two-hour epochs using four different parameters focusing on waveform, topography, location and orientation [13]. At least four similar events are required for a cluster to be identified. The remaining detections are placed in a residual section. The software shows the 20 waveforms with the most similarity to the cluster mean, together with the equivalent locations of the events in a head scheme. A human reviewer has to categorize the clusters as epileptiform or not.

For this research, we used a 30-minute selection. For the software to make proper clusters, 45-minute epochs before and after the 30-minute selection were also reviewed by the BESA software. BESA Epilepsy also has no sensitivity settings.

Minimal requirements hardware and process time

Persyst and Encevis have similar minimal requirements: i5 processor (1.6 GHz) or better and a memory of 4 GB RAM or better. Both packages review spikes and seizures together and require one minute of processing time for every six minutes of EEG, so each 30-minute selection takes around five minutes to process. In our experience, EEGs with many detections need more review time. The Persyst software can review four EEGs at the same time without losing processing speed. BESA Epilepsy review seizures and spikes separately. Spike detection also needs around one minute for every six minutes of EEG. Minimal requirements

are different: 1 GHz processor, 1 GB RAM and a graphics card OpenGL 1.1 with at least 16 MB RAM.

Analysis

Whether or not IEDs had occurred was assessed per 30-minute selection and per 10-second epoch. A 10-second duration was used because electroencephalographers usually detect IEDs by visual inspection of an 'EEG page' of 10 or 15 seconds [5]. Interrater agreement between the three human experts was calculated using Fleiss' kappa scores.

A gold standard was created using the 'expert scores' of the three experts: an EEG period (either a 30-minute selection or a 10-second epoch), was considered 'definitely epileptiform' if at least two of the three experts had marked it as containing at least one definite IED. A selection or epoch was considered 'possibly epileptiform' if only one expert had marked it as containing one or more definite IEDs, or if at least two experts had marked it as containing at least one possible IED. A selection or epoch was considered 'not epileptiform' when none of the experts had detected definite IEDs, and a maximum of one expert had marked at least one possible IED.

Sensitivity and specificity for all three software packages were estimated for the complete 30-minute selection and all separate 10-second epochs. Interrater agreement was also estimated between the gold standard and the software packages. Additionally sensitivity, specificity and interrater agreement were estimated between the gold standard and the original clinical report. This was only possible for the 30-minute selections.

We want to use the detection software as a screening tool, so possible epileptiform selections and epochs were considered as epileptiform when calculating sensitivity, specificity and interrater agreement. All kappa scores were estimated using SPSS (IBM SPSS Statistics for Windows, Version 26.0.).

Results

Database characteristics

The study set consisted of records from 13 men and 17 women. The median age was 39 years (range 18 – 76 years).

According to the clinical reports, 20 EEGs contained IEDs (Table 1), of which six were generalized, and 14 were focal. Ten of these were temporal (unilateral or bilateral) and four were extratemporal.

Performance

Table 1. Presence of definite or possible IEDs per human expert

30-minute selection	HE 1	HE 2	HE 3	'expert scores'	Clinical report
# definite IEDs	19	18	21	20	20
# possible IEDs	3	3	2	2	0
# no IEDs	8	9	7	8	10
10-second epoch	**HE 1**	**HE 2**	**HE 3**	**'expert scores'**	
# definite IEDs	392	318	453	419	
# possible IEDs	190	113	297	194	
# no IEDs	4818	4969	4650	4787	

 # = number containing, IEDs = interictal epileptiform discharges, HE = human expert.

30-minute selection

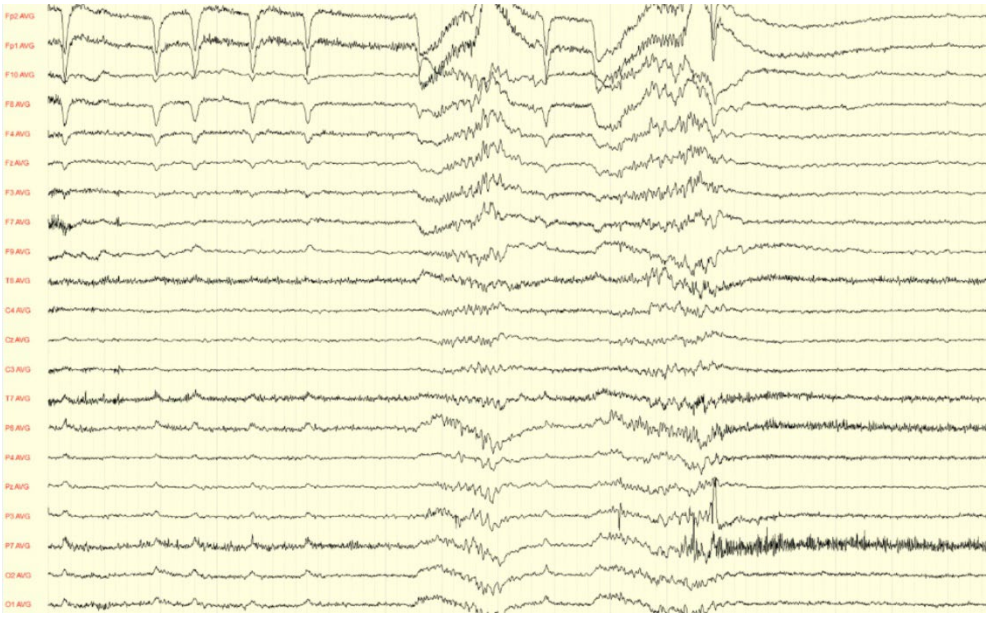
The kappa between the human experts was 0.69 (CI 0.53 – 0.86; Table 1). Sensitivity, specificity and kappa between the gold standard and software results are shown in Table 2.

Table 2. IED detections per 30-minute selection

	HE +	HE ±	HE -	Sensitivity*	Specificity*	Kappa*
Persyst + (M)	19	2	1	95% (CI 75-100%)	88% (CI 47-99%)	0.83 (CI 0.60-1.00)
Persyst – (M)	1	0	7			
Persyst + (L)	17	1	0	82% (CI 59-94%)	100% (CI 60-100%)	0.71 (CI 0.45-0.96)
Persyst – (L)	3	1	8			
Encevis +	17	2	1	86% (CI 64-96%)	88% (CI 47-99%)	0.68 (CI 0.40-0.97)
Encevis -	3	0	7			
BESA +	8	0	0	36% (CI 18-59%)	100% (CI 60-100%)	0.23 (CI 0.05-0.42)
BESA -	12	2	8			
Clinical report +	8	2	0	91% (CI 69-98%)	100% (CI 60-100%)	0.80 (CI 0.54-1.00)
Clinical report -	1	1	8			

HE = human experts, + = IEDs detected, ± = only possible IEDs detected, - = no IEDs detected, M = medium setting, L = low setting, IEDs = interictal epileptiform discharges.
** possible epileptiform selections and epochs were considered as epileptiform when calculating sensitivity, specificity and kappa score.*

All three software packages missed epileptiform discharges in one particular record containing multiple IEDs, which could be interpreted as eye closure sensitivity (Figure 2). Encevis missed all IEDs in a further two records containing one and four focal IEDs. The BESA software missed 12 records containing IEDs; in four records, no accurate detection was made. In eight records, IEDs were detected but only presented in the residuals.



The example is shown in an average reference montage.

Figure 2. Example of the eye closure sensitivity all three software packages missed.

10-second epochs

The kappa between the experts was 0.63 (CI 0.50-0.76; Table 1). Sensitivity, specificity and kappas are shown in Table 3.

Table 3. IED detections per 10-second epoch

	HE +	HE ±	HE -	Sensitivity*	Specificity*	Kappa*
Persyst +(M)	374	128	215	82% (CI 79-85%)	96% (CI 95-96%)	0.72 (CI 0.69-0.75)
Persyst – (M)	45	66	4572			
Persyst + (L)	320	70	35	64% (CI 60-67%)	99% (CI 99-96%)	0.73 (CI 0.70-0.76)
Persyst – (L)	99	124	4752			
Encevis +	341	34	45	61% (CI 57-65%)	99% (CI 99-99%)	0.67 (CI 0.63-0.70)
Encevis -	98	180	4702			
BESA +	259	30	56	50% (CI 46-54%)	99% (CI 98-99%)	0.57 (CI 0.53-0.61)
BESA -	150	174	4731			

HE = human experts, + = IEDs detected, - = no IEDs detected, M = medium setting, L = low setting, IEDs = interictal epileptiform discharges
** possible epileptiform selections and epochs were considered as epileptiform when calculating sensitivity, specificity and kappa score.*

Discussion

Persyst in the medium (default) setting achieved the highest sensitivity, together with a reasonable specificity, which is appropriate when using the software as a screening tool. BESA had missed the most IEDs, mainly due to the cluster system, in which some of the missed IEDs were initially detected but categorized in the residual section where they are easily missed. Persyst, in the low setting, and BESA had the highest specificity.

These specificities are higher than reported in previous studies [14,15,16,17], probably due to the fact that we did not review single IEDs, but measured whether 10-second epochs and 30-minute selections contained IEDs.

Kappa's for detection of IEDs between human experts were 0.69 for 30-minute and 0.63 for 10-second periods. Similar interrater kappas have been reported in the literature [18,19]. Comparing the gold standard and the software packages, kappa ranged between 0.23 (BESA) and 0.83 (Persyst). The Persyst and Encevis software packages show similar interrater agreements when compared to the agreement between human experts and the original clinical report.

Together with our previous findings [7], our results suggest that automated spike detection can perform almost as well as human review when reviewing EEGs. Detections made by the software must always be checked and verified by experts, especially when using the software as a screening tool without complete visual EEG assessment; this requires high sensitivity, usually associated with lower specificity.

The study has some limitations. Our results cannot be applied to paediatric or sleep EEGs. To minimize the problem of the low interrater agreement for detecting IEDs [19] and to address the most relevant clinical question, i.e. whether or not a record contains IEDs, we did not count the number of IEDs but focused on whether or not IEDS were detected per selection or epoch. The exact number of IEDs counted by human experts is likely to differ between experts and even within the same expert, and some waveforms will remain a matter of discussion. An objective quantification by software might be a better index than human expert counting to study its clinical relevance. This approach also ensures each record is taken equally into account when calculating performance. We also included epochs with possible IEDs, whereas most studies use a simple spike or no-spike characterization when experts label IEDs in a record [14,18]. This resulted in medium to high kappa scores. Lastly, we reviewed 30-minute selections in this work, whereas prolonged EEGs can last for hours or days.

Future work must focus on reviewing the entire prolonged EEG, preferably together with automated seizure detection, as in our previous study [7] and an additional study [20], to investigate whether automated software can (partially) replace visual review of the EEG. Implementing automated software is also challenging; we know that experts' confidence in this software is low, and using the software requires a different approach than conventional human review.

IED detection by automated software from the Persyst performs better than Encevis and BESA and is similar to human review, when reviewing 30-minute selections and 10-second epochs. This finding may help prospective users choose a software package.

References

1. Fountain NB, Freeman JM. EEG is an essential clinical tool: pro and con. Epilepsia. 2006;47:23–5. https://doi.org/10.1111/j.1528-1167.2006.00655.x.

2. Chih-hong Lee a, Siew-Na Lim a, Frank Lien b, Tony Wua,* Duration of electroencephalographic recordings in patients with epilepsy. Seizure 22 (2013) 438–442. http://dx.doi.org/10.1016/j.seizure.2013.02.016.

3. Wirrell ED. Prognostic significance of interictal epileptiform discharges in newly diagnosed seizure disorders. J Clin Neurophysiol. 2010;27:239-48. https://doi.org/10.1097/WNP.0b013e3181ea4288.

4. Guray Koca,⁎, Gulin Morkavukb, Efdal Akkayab, Omer Karadasa, Alev Leventoglub, Bulent Unayc, Zeki Gokcild. Latencies to first interictal epileptiform discharges in different seizure types during video-EEG monitoring. Seizure: European Journal of Epilepsy 69 (2019) 235–240. https://doi.org/10.1016/j.seizure.2019.05.013.

5. Halford JJ. Computerized epileptiform transient detection in the scalp electroencephalogram: Obstacles to progress and the example of computerized ECG interpretation. Clin Neurophysiol. 2009;120:1909-15. https://doi.org/10.1016/j.clinph.2009.08.007.

6. Lodder SS, van Putten MJ. A self-adapting system for the automated detection of inter-ictal epileptiform discharges. PLoS One. 2014;9:e851-80. https://doi.org/10.1371/journal.pone.0085180.

7. Reus EEM, Visser GH, Cox FME. Using sampled visual EEG review in combination with automated detection software at the EMU. Seizure. 2020;80:96-9. https://doi.org/10.1016/j.seizure.2020.06.002.

8. Urbaniak GC, Plous S. Research Randomizer (Version 4.0). https://www.randomizer.org/; 2013. [assessed 9 September 2021].

9. Kane N, Acharya J, Beniczky S, Caboclo L, Finnigan S, Kaplan PW et al. A revised glossary of terms most commonly used by clinical electroencephalographers and updated proposal for the report format of the EEG findings. Revision 2017. Clin Neurophysiol Prac. 2017;2:170-85. http://dx.doi.org/10.1016/j.cnp.2017.07.002.

10. Wilson SB, Turner CA, Emerson RG, et al. Spike detection II: automatic, perception-based detection and clustering. Clin Neurophysiol. 1999;110:404-11. https://doi.org/10.1016/s1388-2457(98)00023-6.

11. Reus EEM, Visser GH, Cox FME. Determining the Spike–Wave Index Using Automated Detection Software. J Clin Neurophysiol. 2021;38:198–201. https://doi.org/10.1097/WNP.0000000000000672.

12. Hartmann MM, Furbass F, Perko H, Skupch A, Lackmayer K, Baumgartner C, et al. EpiScan: online seizure detection for epilepsy monitoring units. 2011 Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE. 2011:6096–9. https://doi.org/10.1109/IEMBS.

13. Scherg M, Ille N, Weckesser D, Ebert A, Ostendorf A, Boppel T et al. Fast evaluation of interictal spikes in long-term EEG by hyper-clustering. Epilepsia. 2012;53(7):1196–1204. https://doi.org/10.1111/j.1528-1167.2012.03503.x.

14. Scheuer ML, Bagic A, Wilson SB. Spike detection: Inter-reader agreement and a statistical Turing test on a large data set. Clin Neurophysiol. 2016;128:243-50. https://doi.org/10.1016/j.clinph.2016.11.005.

15. Jing J, Sun H, Kim JA, Herlopian A, Karakis J, Ng M, et al. Development of Expert-Level Automated Detection of Epileptiform Discharges During Electroencephalogram Interpretation. JAMA Neurol. 2020;77:103-8. https://doi.org/10.1001/jamaneurol.2019.3485.

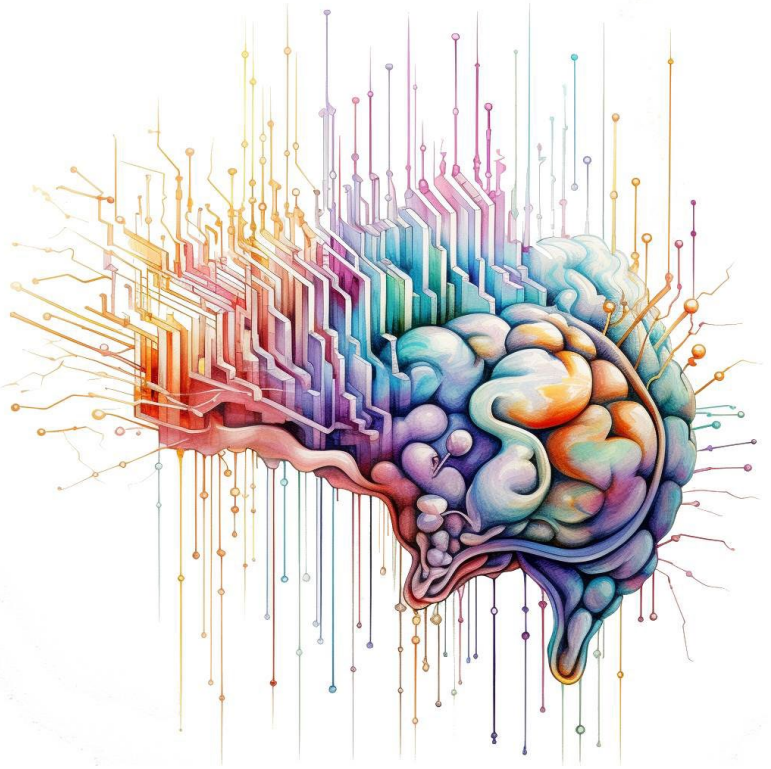
16. Halford JJ, Brandon Westover M, LaRoche SM, Macken MP, Kutluay E, Edwards JC et al. Interictal Epileptiform Discharge Detection in EEG in Different Practice Settings. J Clin Neurophysiol 2018;35: 375–380. https://doi.org/10.1097/WNP.0000000000000492.

17. Fürbass F, Kural MA, Gritsch G, Hartmann M, Kluge T, Beniczky S. An artificial intelligence-based EEG algorithm for detection of epileptiform EEG discharges: Validation against the diagnostic gold standard. Clin Neurophysiol. 2020;131:1174-79. https://doi.org/10.1016/j.clinph.2020.02.032.

18. Bagheri E, Dauwels J, Dean BC, Waters CG, Westover BM, Halford, JJ. Interictal epileptform discharge characteristics underlying expert interrater agreement. Clin Neurophysiol. 2017;128:1994–2005. https://doi.org/10.1016/j.clinph.2017.06.252.

19. Grant AC, Abdel-Baki SG, Weedon J, Arnedo V, Chari G et al. EEG interpretation reliability and interpreter confidence: A large single-center study. Epilepsy Behav. 2014;32:102-7. https://doi.org/10.1016/j.yebeh.2014.01.011.

20. Reus EEM, Visser GH, van Dijk JG, Cox FME. Automated seizure detection in an EMU setting: Are software packages ready for implementation? Seizure 2022;96:13-7. https://doi.org/10.1016/j.seizure.2022.01.009.



CHAPTER 5

Automated seizure detection in an EMU setting: are software packages ready for implementation?

E.E.M. Reus
G.H. Visser
J.G. van Dijk
F.M.E. Cox

Seizure 2022;96:13-7.
https://doi.org/10.1016/j.seizure.2022.01.009.

Chapter 5 - Automated seizure detection in an EMU setting: are software packages ready for implementation?

Abstract

Purpose: We assessed whether automated detection software, combined with live observation, enabled reliable seizure detection using three commercial software packages: Persyst, Encevis and BESA.

Method: Two hundred and eighty-six prolonged EEG records of individuals aged 16 – 86 years, collected between August 2019 and January 2020, were retrospectively processed using all three packages. The reference standard included all seizures mentioned in the clinical report supplemented with true detections made by the software and not previously detected by technicians. Sensitivity was measured for offline review by technicians and software seizure detection, both in combination with live monitoring in an EMU setting, for all three software packages at record and seizure level.

Results: The database contained 249 seizures in 64 records. The sensitivity of seizure detection was 98% for Encevis and Persyst, and 95% for BESA, when a positive results was defined as detection at least one of the seizures occurring within an individual record. When positivity was defined as recognition of all seizures, sensitivity was 93% for Persyst, 88% for Encevis and 84% for BESA. Technicians' review had a sensitivity of 100% at record level and 98% at seizure level. The median false positive rate per record was 1.7 for Persyst, 2.4 for BESA and 5.5 for Encevis per 24 hours.

Conclusion: Automated seizure detection software does not perform as well as technicians do. However, it can be used in an EMU setting when the user is aware of its weaknesses. This assessment gives future users helpful insight into these strengths and weaknesses. The Persyst software performs best.

Introduction

Seizure recording using video-EEG plays an essential role in diagnosing epilepsy, seizure classification and identification of candidates for epilepsy surgery [1,2]. Prolonged EEG recordings improve the chances of finding ictal activity [3]. Longer recordings, however, result in more review time and a cost increase.

The typical procedure of recording a prolonged EEG in an Epilepsy Monitoring Unit (EMU), involves continuous observation of individuals by trained nurses and staff, as well as alerts by patients who press an alarm at the onset of a perceived seizure [4]. This procedure detects around two-thirds of all seizures [5], and the remaining one third is detected by technicians who later review the entire EEG record offline. Automated detection software may serve as a screening tool to reduce the need for complete visual reviewing of the recording and save time, provided it is sufficiently reliable.

A pilot study showed that automated detection software in combination with sampled visual review could be used as a reliable substitute for a complete visual review of prolonged video-EEGs concerning IEDs (interictal epileptiform discharges) [6]. The number of seizures in that study was however too low to validate the performance of the automated seizure detection, not yet allowing its use as a substitute for visual review.

To approach the real life use of seizure detection software, we compared the seizure detection performance using online human observation (in live setting by trained nurses) in combination with both conventional review by technicians and software seizure detection using three commercially available software packages.

Method

EEG data

We retrospectively collected 286 anonymous prolonged video-EEG records (> 4 hours) from 283 individuals aged at least 16 years between Augustus 2019 and January 2020. EEG data were recorded using the Micromed EEG system (Micromed, Mogliano Veneto, Italy), using the standard 10-20 international electrode recording and additional F9/F10 positions sampled at 256 Hz. We recorded EEGs exclusively in the context of clinical care, so, according to Dutch rules, individual informed consent was not required. The local medical ethics committee approved this study.

Automated detection software packages

We used three commercially available software packages: Persyst (Persyst Development Corporation, USA) version 14, Encevis (AIT Austrian Institute of Technology, Austria) detection), version 1.9.2. and BESA (BESA Epilepsy, Germany) version 2.0. Encevis is the only package that also uses the ECG channel for seizure detection. For all three packages, the output is a list of timed seizure detections. Additionally, BESA also presents lateralization information (i.e. left, right). We used only the seizure detection features of the software, ignoring other tools.

Reference standard

The EEG data of all seizures mentioned in the original EEG report were reviewed in a consensus procedure by at least one technician and one epileptologist. Seizures were categorized according to when the seizure was first detected: in the live setting, i.e. through nurses' observations and individuals' alarm buttons, or offline through technicians' review. All detections were recorded into a sheet. EEG outside seizure selections was not reviewed.

The same records were analyzed with the three automated detection software, and all detections made by one or more software packages were compared with the seizures mentioned in the original EEG report. Detections were classified as congruent if the software detection fell within a time window of 30 seconds before the onset or after the end of the seizure, and incongruent otherwise.

All incongruent detections were reviewed by a trained human expert with more than five years' experience in reviewing EEGs. An actual seizure detection was defined as repetitive epileptiform EEG discharges of >2 Hz or a characteristic pattern with a quasi-rhythmic spatio-temporal evolution (i.e. a gradual change in frequency, amplitude, morphology or location), usually lasting ten or more seconds [7]. A second human expert then double-checked this. Software detections that did not meet the criteria were considered false detections. Two or more false software detections within 60 seconds were counted as a single false detection.

The reference standard included all seizures mentioned in the clinical report supplemented with true detections made by the software and not previously detected by technicians. The durations of all ictal EEG patterns were identified, and seizure classification was determined using the latest ILAE seizure classification [8]. Only the first ten seizures per EEG record were included to reduce sampling bias. We regarded records in which any seizure was detected as positive for epilepsy, regardless of whether all seizures in that record had been identified.

Analysis

In an EMU setting most seizures are detected in the live setting, so our primary outcome measure was the sensitivity for live seizure detection in combination with both offline review

by technician and software seizure detection for all three software packages. Differences in performance between the technicians and all three software packages were analyzed using the McNemar test for non-parametric data using SPSS (IBM Corp. Released 2019. IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp.) We made a distinction between seizures with no or short (< 10 seconds) ictal pattern and seizures with a seizure pattern duration of at least 10 seconds. We also measured sensitivity using the software seizure detection alone, including the seizures detected online. Finally we estimated the false positive rate per 24 hours.

Results

Database characteristics

We included 286 prolonged EEG records from 283 people (135 male, 148 female) with a median age of 36 years (range 16 – 86 years) and a summed recording time of over 8771 hours. The median duration was 20 hours and 40 minutes (range 4 hours and 3 minutes to 97 hours and 56 minutes).

There were 336 seizures in 64 records (range 1 – 39 seizures per record). From the eight records with more than ten seizures we included the first ten, remaining 249 seizures for further analysis.

Performance per record

Of the 64 records containing seizures, 56 were recognized as containing seizures in the live setting. In the later offline review, technicians detected seizures in an additional eight records. The software packages did not identify one record which contained one generalized myoclonic event. BESA missed two further records, one having a focal seizure and one containing an electroencephalographic seizure. See Figure 1. Hence, sensitivity for the combination of live observation and offline review by technician was 100% (CI 93-100%). Sensitivity for the combination of live observation and automated detection using Persyst and Encevis was 98% (CI 90–100%) and 95% (CI 86–99%) when using BESA. There was no statistically significant difference in performance between the reference standard and either of the software packages (for all $P > 0.05$) nor between the reference standard and the technician ($P > 0.05$).

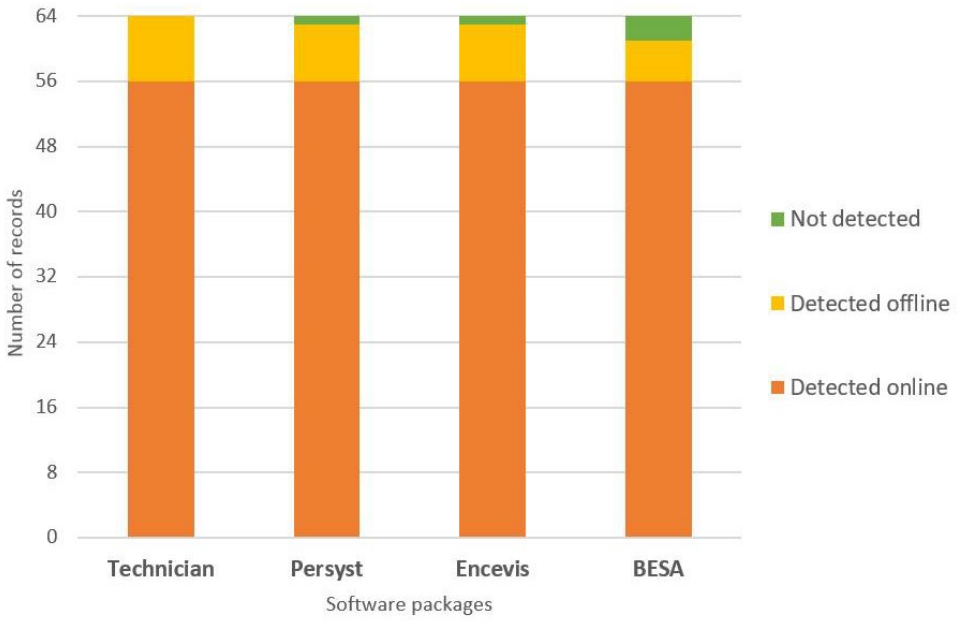


Figure 1. performance software packages on record level.

Table 1. Seizures detected in live setting or by technicians

Seizure type	Total #	# first detected in live setting (online)	# first detected by technician (off line)	# not detected (live or by technician)
Generalized tonic clonic	1	1	0	0
Focal to bilateral tonic clonic	13	13	0	0
Focal impaired awareness	80	69	11	0
Focal aware/unknown - motor	59	49	8	2
Focal aware – non motor	26	26	0	0
Absence	11	4	7	0
Generalized myoclonic	21	15	6	0
Generalized tonic	14	3	11	0
Focal electro-encephalographic	24	4	17	3
Total #	**249**	**184**	**60**	**5**
Duration ictal EEG ≥ 10 seconds	146	95	46	5
Duration ictal EEG < 10 seconds	103	89	14	0

 # = number of seizures

Performance per seizure

Of the 249 seizures, 184 were recognized in the live setting. Table 1 shows the recorded seizure types, the duration of ictal patterns, and the technicians' performance. Sensitivity for the combination of live monitoring and offline review by a technician was 98% (CI 95–99%). The five undetected seizures were all focal in nature.

Table 2. Seizures not detected online

Seizure type	Total # not detected online	# detected by Persyst	# detected by Encevis	# detected by BESA
Generalized tonic clonic	0	0	0	0
Focal to bilateral tonic clonic	0	0	0	0
Focal impaired awareness	11	11	10	7
Focal aware/unknown - motor	10	8	7	5
Focal aware – non motor	0	0	0	0
Absence	7	6	6	5
Generalized myoclonic	6	0	0	0
Generalized tonic	11	5	2	1
Focal electro-encephalographic	20	18	10	6
Total #	**65**	**48**	**35**	**24**
Duration ictal EEG ≥ 10 seconds	51	48	34	24
Duration ictal EEG < 10 seconds	16	0	1	0

= number of seizures

Table 2 and Figure 2 show the performance of the three software packages. Sensitivity for the combination of live monitoring and seizure detection by Persyst was 93% (CI 89–96%), by Encevis 88% (CI 83–92%), and by BESA 84% (CI 78–88%). The differences in performance between the technician and the software packages were significantly different (Persyst, P = 0.02; Encevis P < 0.001; BESA P < 0.001). The undetected seizures are shown in table 2; they mostly concerned generalized myoclonic, generalized tonic seizures. Whether focal seizures remained undetected depended on the software package (Table 2). Closer inspection of the focal seizures showed subtle events, with slowly evolving ictal patterns with low amplitudes and frequencies (See Supplementary data). The generalized myoclonic seizures had short (one to two seconds) ictal patterns and occurred in people with a (suspected) generalized myoclonic epilepsy. The missed tonic seizures had somewhat longer (two to five seconds)

ictal patterns and occurred only in individuals with mental impairment and a history of tonic seizures.

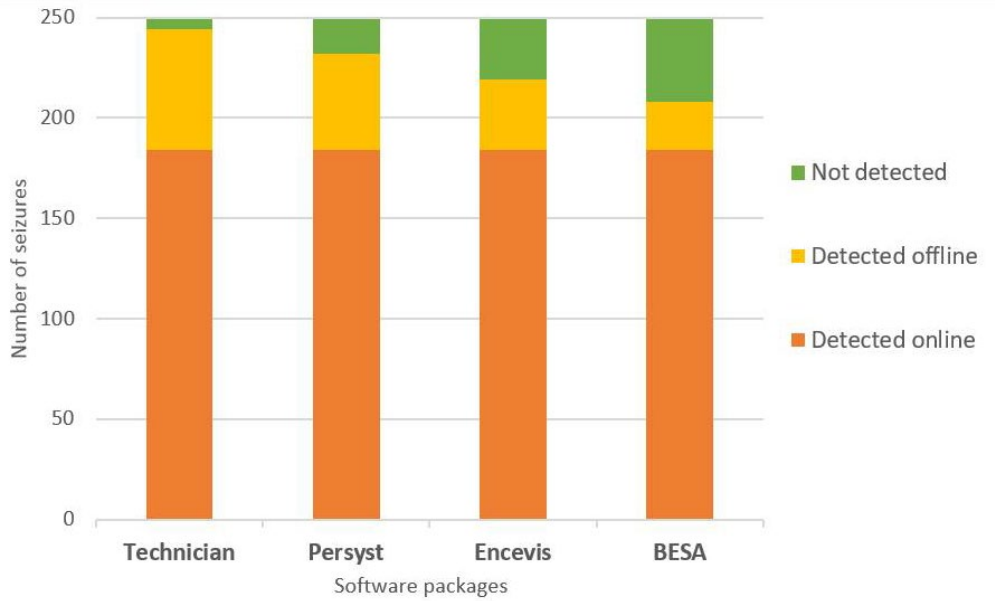


Figure 2. performance software packages on seizure level.

Table 3. Total number of seizures

	Total #	# detected by Persyst	# detected by Encevis	# detected by BESA
Total #	**249**	**139**	**129**	**107**
Duration ictal EEG ≥ 10 seconds	146	133	121	101
Duration ictal EEG < 10 seconds	103	6	8	6

= number of seizures

Sensitivity regarding all 249 seizures was 56% (CI 49-62%) for Persyst, 52% (CI 45 – 58%) for Encevis and 43% (CI 37 – 49%) for BESA. Sensitivity regarding seizures with an EEG pattern lasting 10 seconds or longer was 91% (CI 87–94%) for Persyst, 83% (CI 77-87%) for Encevis detected and 69% (CI 63-75%) for BESA detected (Table 3).

False positive detections

False positive rate for Persyst is 1.7 per 24 hours, for Encevis 5.5 per 24 hours and for BESA 2.4 per 24 hours. Most of the false positives were chewing artifacts, non-seizure related tachycardia (Encevis), muscle artifacts, movement artifacts or interictal activity.

Discussion

Seizure detection by a combination of live monitoring and automated software had a sensitivity of 95% (BESA) and 98% (Encevis, Persyst) when aiming to detect at least one of the seizures occurring within an individual record and sensitivity of 84% (BESA), 88% (Encevis) and 93% (Persyst) when aiming to detect all seizures. Technicians' review had a sensitivity of 100% on record level and 98% on seizure level. Hence, Persyst detected the highest number of seizures, and BESA the lowest. The software packages performed better on seizures with 10 seconds or longer duration. We found a false positive rate of 1.7 and 2.4 per 24 hours when using Persyst and BESA, which we considered acceptable. This false positive rate is lower than reported in previous literature, using an older version of Persyst [9]. A validation study of the currently used version (P14) reported false positive rate comparable to present study [10]. Encevis showed a considerably higher false positive rate.

Earlier studies found that detection algorithms had a sensitivity for epileptic seizures between 73% and 96% [11,12]. A recent study comparing the same three software packages reported a sensitivity of 76.6% for BESA, 77.8% for Encevis and 81.6% for Persyst on a database containing largely focal seizures [9]. In our study we approach how the software would really be used in an EMU setting, by reviewing the combination of live human observation and offline review, comparing the performance of the technicians versus the software. Furthermore our database also contains generalized seizures, such as myoclonia and tonic seizures. Sensitivities for these seizure types, and for focal aware seizure, are low. This is due to the fact that they usually have no or short EEG correlates [13]. The highest sensitivities are reported for (focal to) generalized tonic-clonic seizures and focal seizures with impaired awareness [13].

Both the present study and previous reports suggest that detection software does not perform as well as technicians. We believe, however, that detection software can be of use provided the user is aware of its weaknesses. Patients can usually detect myoclonus and focal aware seizures themselves, and report them via the push button [14,15]. This does not apply to generalized tonic seizures; our data show a large proportion of those were undetected in the live setting. This seizure type usually occurs in people with mental impairment with a history of tonic seizures. We suggest EEGs of this population should be thoroughly visually reviewed to avoid missing significant events. Our previously proposed method with a targeted sampled review, including a period after waking in people with suspected JME, can

also increase seizure detection [6]. Thus, the ictal patterns seen in myoclonic seizures, which are usually are too short to be detected by a seizure detector, will be detected by a spike detector. To a lesser extent, slowly evolving seizures with low amplitudes and frequencies can also be missed by the software. Previous literature shows that the use of quantitative EEG spectrograms can increase the detection of these seizures [16]. Automated detection software, however, also detected five additional seizures, which were initially missed in the offline review by the technician. Finally, in the design of this study we used automated seizure detection as a screening tool. Detections made by the software must always be checked and verified by experts.

Our study has some limitations. It is a single centre study and results may differ in other settings. We also only used EEG recordings from teenagers and adults, so our results do not apply to paediatric EEGs. We only focused on seizure detection. In an additional, yet unpublished, study we also compare the performance of the spike detection features of these software packages [17]. We used a pragmatic approach for the reference standard. However, ideally the EEG records should be reviewed in their totality and by two epileptologists. Furthermore, the online usability of these detection software packages should be investigated, as they might possibly be beneficial for patient safety and ictal testing. Lastly, it would be insightful to look at experts' confidence of this software.

Automated seizure detection software does not perform as well as technicians do. However, it can be used in an EMU setting when the user is aware of its weaknesses. The software is most sensitive to focal seizures with impaired awareness and tonic clonic seizures and least sensitive to generalized tonic and generalized myoclonic seizures.
The use of such detection software can potentially save time. This assessment may give future users helpful insight into the strengths and weaknesses of this software and help prospective users choose a software package. The Persyst software has the best performance.

References

1. Shih JJ, Fountain NB, Herman ST, Bagic A, Lado F, Arnold S et al. Indications and methodology for video-electroencephalographic studies in the epilepsy monitoring unit. Epilepsia 2018;59:27–36. https://doi.org/10.1111/epi.13938.

2. Fountain NB, Freeman JM. EEG is an essential clinical tool: pro and con. Epilepsia 2006;47:23–5. https://doi.org/10.1111/j.1528-1167.2006.00655.x.

3. Friedman DE, Hirsch LJ. How long does it take to make an accurate diagnosis in an epilepsy monitoring unit? Clin Neurophysiol 2009;26:213–7. https://doi.org/10.1097/WNP.0b013e3181b2f2da.

4. Cox FME, Reus EEM, Widman G, Zwemmer JNP, Visser GH. Epilepsy Monitoring Units can be safe places; a prospective study in a large cohort. Epilepsy Behav 2020;102:106718. https://doi.org/10.1016/j.yebeh.2019.106718.

5. Rommens N, Geertsema E, Jansen Holleboom L, Cox FME, Visser GH. Improving staff response to seizures on the epilepsy monitoring unit with online EEG seizure detection algorithms. Epilepsy Behav 2018;84:99–104. https://doi.org/10.1016/j.yebeh.2018.04.026.

6. Reus EEM, Visser GH, Cox FME. Using sampled visual EEG review in combination with automated detection software at the EMU. Seizure 2020;80:96-9. https://doi.org/10.1016/j.seizure.2020.06.002.

7. Kane N, Acharya J, Beniczky S, Caboclo L, Finnigan S, Kaplan PW et al. A revised glossary of terms most commonly used by clinical electroencephalographers and updated proposal for the report format of the EEG findings. Revision 2017. Clin Neurophysiol Prac 2017;2:170-85. http://dx.doi.org/10.1016/j.cnp.2017.07.002.

8. Fisher RS, Cross JH, French JA, Higurashi N, Hirsch E, Jansen FE, et al. Operational classification of seizure types by the international league against epilepsy: position paper of the ILAE commission for classification and terminology. Epilepsia 2017;58:522-30. https://doi.org/10.1111/epi.13670.

9. Koren J, Hafner S, Feigl M, Baumgartner C., Systematic analysis and comparison of commercial seizure-detection software. Epilepsia 2021;62:426-38. https://doi.org/10.1111/epi.16812

10. Scheuer ML, Wilson SB, Antony A, Ghearing G, Urban A, Bagić AI. Seizure Detection: Interreader Agreement and Detection Algorithm Assessments Using a Large Dataset. J Clin Neurophysiol 2020;38:439-47. https://doi.org/10.1097/WNP.0000000000000709.

11. Baumgartner C, Koren JP. Seizure detection using scalp-EEG. Epilepsia 2018a;59:14-22. https://doi.org/10.1111/epi.14052.

12. Baumgartner C, Koren JP. Rothmayer M. Automatic Computer-Based Detection of Epileptic Seizures. Front Neurol 2018b; 9:639. https://doi.org/10.3389/fneur.2018.00639.

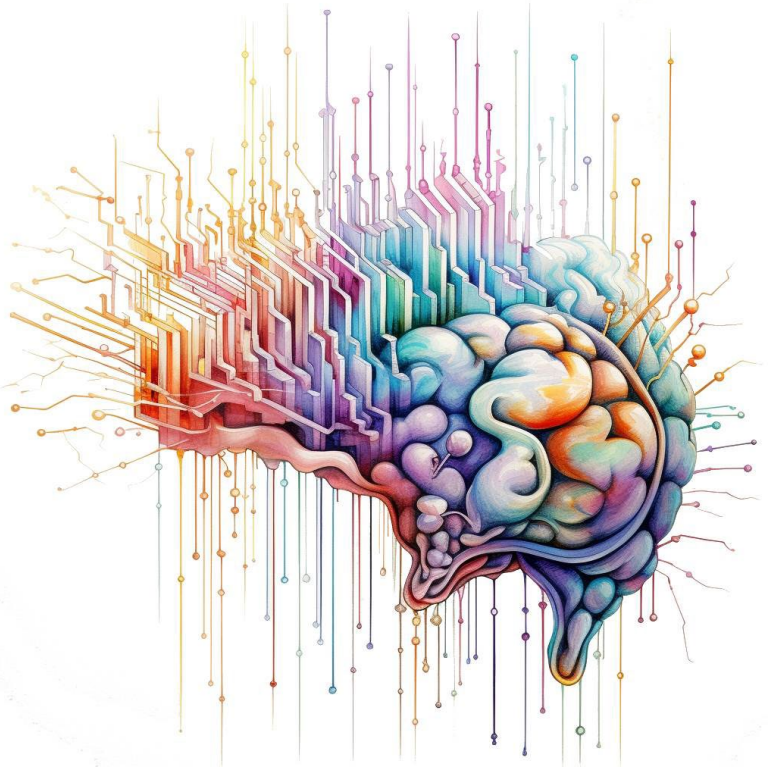
13. Kamitaki BK, Yum A, Lee J, Rishty S, Sivaraaman K, Esfahanizadeh A, Mani R, Wong S. Yield of conventional and automated seizure detection methods in the epilepsy monitoring unit. Seizure 2019; 69:290-295. https://doi.org/10.1016/j.seizure.2019.05.019

14. Hoppe C, Poepel A, Elger CE. Epilepsy: accuracy of patient seizure counts. Arch Neurol 2007; 64:1595–9. https://doi.org/10.1001/archneur.64.11.1595.

15. Atkinson M, Hari K, Schaefer K, Shah A. Improving safety outcomes in the epilepsy monitoring unit. Seizure 2012;21:124–7. https://doi.org/10.1016/j.seizure.2011.10.004.

16. Goenka A, Boro A, Yozawitz, E. Comparative sensitivity of quantitative EEG (QEEG) spectrograms for detecting seizure subtypes. Seizure 2018;55:70-5. https://doi.org/10.1016/j.seizure.2018.01.008.

17. Reus EEM, Cox FME, Dijk van JG, Visser GH. Automated spike detection: which software package? Seizure 2022;95: 33–7. https://doi.org/10.1016/j.seizure.2021.12.012



CHAPTER 6

Automated spike and seizure detection: are we ready for implementation?

E.E.M. Reus
G.H. Visser
M.P.J. Sommers-Spijkerman
J.G. van Dijk
F.M.E. Cox

Seizure 2023;108:66-71.
https://doi.org/10.1016/j.seizure.2023.04.010.

Chapter 6 – Automated spike and seizure detection: are we ready for implementation?

Abstract

Purpose: Automated detection of spikes and seizures has been a subject of research for several decades now. There have been important advances, yet automated detection in EMU (Epilepsy Monitoring Unit) settings has not been accepted as standard practice. We intend to implement this software at our EMU and so carried out a qualitative study to identify factors that hinder ('barriers') and facilitate ('enablers') implementation.

Method: Twenty-two semi-structured interviews were conducted with 14 technicians and neurologists involved in recording and reporting EEGs and eight neurologists who receive EEG reports in the outpatient department. The study was reported according to the Consolidated Criteria for Reporting Qualitative Studies (COREQ).

Results: We identified 14 barriers and 14 enablers for future implementation. Most barriers were reported by technicians. The most prominent barrier was lack of trust in the software, especially regarding seizure detection and false positive results. Additionally, technicians feared losing their EEG review skills or their jobs. Most commonly reported enablers included potential efficiency in the EEG workflow, the opportunity for quantification of EEG findings and the willingness to try the software.

Conclusion: This study provides insight into the perspectives of users and offers recommendations for implementing automated spike and seizure detection in EMUs.

Introduction

Machine learning has increasingly been used and been the subject of research in health care with the aim of improving efficiency [1]. Fields of interest in epilepsy include analysis of imaging and clinical data, epilepsy source localization, prediction of medical and surgical outcomes, and automated EEG-based detection [2]. The latter has been the scope of research for several decades [3,4,5], with some remarkable achievements [6,7]. The research focused on development and testing of new detection algorithms; validation studies of various commercially available software packages were published, often with promising results [8,9,10,11,12,13]. Despite these publications and advances automated EEG-based detection in EMU (Epilepsy Monitoring Unit) settings has not been accepted as standard practice.

Implementing changes in health care practice is often challenging. Successful implementation largely depends on acceptance by professionals; that is, the extent to which they believe that a given innovation is agreeable or satisfactory, and a willingness to try an innovation [14].

In the current qualitative study we surveyed thoughts, attitudes, experiences and needs of both producers and recipients of EEG reports regarding automated EEG-based detection, using semi-structured interviews. We aimed to identify factors that hinder ('barriers') and facilitate ('enablers') future implementation. This information may help guide successful implementation of such software.

Method

We conducted semi-structured interviews using a phenomenological approach. Method and results were reported according to the Consolidated Criteria for Reporting Qualitative Studies (COREQ) [15].

Current EEG (review) process

The study was performed in Stichting Epilepsie Instellingen Nederland (SEIN), a tertiary referral center with two clinical locations (Heemstede and Zwolle) and an outpatient clinic network. Each location has an Epilepsy Monitoring Unit (EMU) where we perform prolonged EEG recordings [16].

Participants

We applied consecutive sampling for the selection of participants: all technicians, neurologists and physician assistants working at both clinical neurophysiology departments

and all neurologists at the outpatient clinics received an email invitation to participate in the study. Potential participants were invited for interview by email. Prior to the invitation, they attended a presentation about the use of automated detection for reviewing prolonged EEGs, including previous research on automated detection [9,10,11] and our proposed method of using automated detection in combination with sampled review [17], see Figure 1.

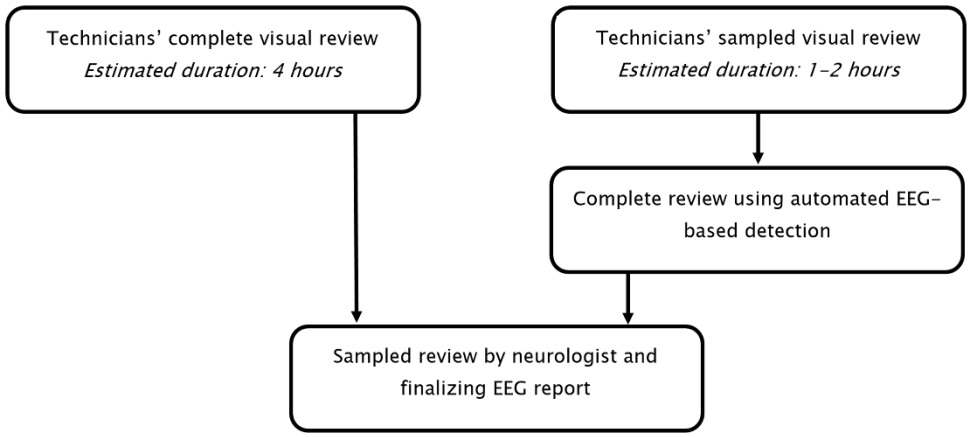


Figure 1. Current EEG workflow (left) and possible future EEG workflow (right).

Nineteen of 36 clinical neurophysiology staff members and twelve of 29 outpatient clinic neurologists who we contacted were willing to participate in an interview. Based on the reached data saturation, we included nine technicians (participants TC1 to TC9), five medical staff members, consisting of neurologists and physician assistants (participants MC1 to MC5), and eight outpatient clinic neurologists (participants MO1 to MO8). No participant dropped out. Seventeen of the 22 participants were familiar with the interviewer, and all participants knew that the researcher was involved in research on automated EEG detection.

Data collection

Semi-structured interviews were conducted by ER, a female physician assistant and research fellow working at the department of clinical neurophysiology. ER was trained to perform semi-structured interviews. The training included discussion of content and practicalities regarding qualitative interviewing, as well as practice interviews including reflection and feedback afterwards.

The preliminary interview questions underwent pilot testing with two non-participating colleagues prior to the commencement of the study. Feedback was solicited from the participants at the end of the pilot interviews and subsequently integrated into the final interview protocol. The same set of questions was used for all participants, and all participants were interviewed once. Topics regarded experience with, knowledge of, and trust in automated EEG-based detection. In addition, technicians and neurologists at the clinical neurophysiology departments were interviewed about their current EEG review method, as well as willingness to work with automated EEG-based detection and requirements necessary for that. Most such questions were open ended. Only the interviewer and the participant were present during the interview, which took place within the institution or online.

Interviews were semi-structured with open-ended questions to facilitate participant led, free-flowing conversation. Participants could raise new subjects. Interviews lasted between 14 and 42 min. After each interview, the interviewer asked whether the participant was satisfied with all the answers or wanted to add anything. We continued to invite participants until data saturation was reached; that is, no new information was gathered and no new themes or subjects had emerged in the last three interviews. The range of work experience of the participants was 3 to 46 years (median 12 years).

Data analysis

Interviews were audio recorded and transcribed in full. No field notes were made. Participants did not receive interview transcripts. Software package NVivo was used to analyze interview transcripts (QSR International Pty Ltd. Version 12, 2020). ER coded the interviews using an inductive thematic analysis [18]. All coded interview transcripts were reviewed for a second time and were discussed within the research team until consensus was reached on all themes. Quotes were selected to illustrate the final themes.

Ethical approval

This study was approved by the institutional review board of SEIN. All participants gave their written informed consent prior to the interview.

Results

Clinical neurophysiology: technicians, physician assistants and neurologist

We identified 13 barriers and 13 enablers, see Table 1 and 2.

Table 1. Barriers regarding use of automated detection software.

Barriers	Technicians	Neurologists and PAs clin neurophys	Neurologists outpatient department
Satisfaction with quality of current review process	X	X	X
Colleagues' presumed unwillingness to try	X		
Technicians need a lot of training and guidance in order to use the software		X	
Teaching EEG review to new students is suboptimal with new review method	X		
Fear losing ability to review long periods of EEG	X		
Previous experience with automated seizure detection was disappointing	X		
Can only be used as supplement, not as (partial) replacement	X		
Only works in EEGs with normal background	X		
Too many false positives	X		
Software sometimes malfunctions	X	X	
Software doesn't perform as well as human experts do	X		
Fear of missing (subtle) seizures or other important information	X	X	X
Fear of missing non-specific EEG abnormalities			X
Fear of losing job	X		

X = subject was mentioned by at least one participant, clin neurophys = clinical neurophysiology department, PAs = physician assistants.

Table 2. Enablers regarding use of automated detection software.

Enablers	Technicians	Neurologists and PAs clin neurophys	Neurologists outpatient department
Current EEG reviewing process is time-consuming	X	X	
Need for more efficient workflow	X	X	X
Software can potentially make workflow more efficient	X	X	X
Possibility of growing future trust in the software	X		
Positive attitude among most participants	X	X	
The need to adopt machine learning in modern diagnostics	X	X	X
Presentations and discussions about the subject increased willingness use	X		
Trust in neurologists and PAs of Clin Neurophys to use the software only when it performs properly			X
Willingness to try themselves	X	X	X
Opportunity for technicians to learn new skills		X	
Trend analyses have added value	X	X	
Helpful in identifying subtle EEG changes over time	X		
Can quantify (interictal and ictal) events	X	X	X
Prefer review by automated EEG-based detection when it means the workflow is faster			X

X = subject was mentioned by at least one participant, clin neurophys = clinical neurophysiology department, PAs = physician assistants.

Current review process

Technicians were satisfied with the quality of the current review process, which they felt ensured that no important information was missed: "*I don't think that we miss important information*" (TC1). The medical staff agreed: "*I have the impression that EEGs are read very carefully*" (MC1). Some technicians mentioned the labor intensiveness of the review process: "*It is a lot of work. We record quite a few hours of EEG*" (TC2). Neurologists and physician assistants also stated that reviewing EEGs took a technician a long time: "*It is a time-consuming process, especially for technicians …They are really working on it for many hours*" (MC4).

Necessity

The need for a more efficient workflow was recognized by most technicians and all medical staff: "*Somehow it has to be made more efficient and faster*" (MC1); "*Health care is only getting more expensive*" (MC4). Respondents felt automated EEG-based detection could play a role in this: "*You don't suddenly have more EEG technicians. So you have to make an efficiency move in a different way*" (MC4).

Technicians, physician assistants and neurologists thought automated detection could help quantify epileptiform discharges: "*If you receive a score which says this abnormality occurs 600 times, that quantifies it more than an estimation by words*" (TC6); "*An algorithm can quantify the spikes for us… this will save the technician time*" (MC3). Additionally, automated detection could help detect gradual changes over time; "*you don't always immediately see the EEG is gradually slowing in, for instance, presurgical patients*." (TC4).

Previous experience

Most participants reported no prior use of automated EEG detection. However, some previously worked with trend analyses, automated seizure detection or automated spike detection software. Users were positive regarding trend analyses: "*It is useful to be able to see it objectively*" (TC3), but negative toward automated seizure detection: "*I found the results disappointing*" (TC2). Previous experience with spike detection had two aspects: users were satisfied with the interface: "*very nice averages of spikes*" (TC6), but more skeptical about the rate of false positive results: "*especially with muscle artifacts*" (TC6).

Willingness

Most technicians felt that future use of automated EEG-based detection was unavoidable: "*We can't get around it anymore, so we have to deal with it*" (TC3); "*So much data is quantified, we can't stay behind with the EEG*" (TC4). All technicians stated they were willing

to try using automated EEG-based detection. Some participants proposed initially using the detection software as an additional review method, along with complete visual review: "*by doing it simultaneously for a while. Just to experience the software*" (TC5); "*I think we need to do both at first. That is the investment we need to make to find out if it is working or not*" (TC4). Some technicians compared the change to using automated EEG-based detection for review with the transition from analog to digital EEG: "*In the beginning the digital EEG was also like 'oh, help', and now you're so used to that*" (TC4).

Some technicians thought some colleagues would be hesitant to use the software: "*I would want to start today, but I think maybe some technicians will need more time to get used to this and gain some confidence*" (TC1).

All neurologists and physician assistants stated willingness to try the software: "*Medicine continues to develop and this is a form of development*" (MC2).

Trust in performance

Most technicians did not trust the software to perform as well as they did. However, most stated that trust could grow with experience with the software: "*I think that trust in these machines has to grow, trust that it performs well. I think that it will take a long time before you can say: 'we will let the machine review the last couple of hours'*" (TC2); "*we need to gain trust in the system*" (TC4).

Some technicians said that not all seizures would be detected without visual review of the EEG: "*We know patients do not always report their seizures adequately, and also that nurses can miss seizures, so…*"(TC7). They felt that the software could not take over this part of EEG review. Regarding spike detection, technicians were mostly worried about false positives: "*I think that it will take a lot of time. You can get confused or insecure about software detection, because the software detects a spike*" (TC2). Some technicians thought automated EEG-based detection could only work in EEGs with normal background activity: "*Detection is a lot simpler when you have a normal background pattern with low amplitude because spikes then distinguish themselves from background really clearly*" (TC5).

Most medical staff stated trust in the software. Others said they did not know yet, because they had not used it before: "*I don't have an exact image of how sensitive automated detection is*" (MC2).

Fears

One medical staff member noticed much distrust regarding implementation, in particular from technicians: "*they also saw a danger to their own job. If everything is going to be automated, where would that leave them?*" (MC4). Some technicians also mentioned a fear of

losing their job: "*At the beginning, I indeed was skeptical, I thought I might lose my job. But when I heard more and we talked about it some more with colleagues, you begin to think this might be useful after all*" (TC3); "*The idea that you are kind of unnecessary, well, that is a difficult step*" (TC2).

Some technicians feared that teaching EEG review to trainee technicians would be less than optimal when using automated EEG-based detection, or that they themselves might lose the ability to review long periods of EEG, when only reviewing shorter parts of the EEG. They stated the need to see the raw EEG to keep or to gain experience: "*You must continue reviewing longer periods of EEG, you can't learn if something is abnormal or not based on half an EEG page*" (TC3).

In addition, technicians feared loss of quality: "*fear of missing something that you might have detected yourself*" (TC2).

User needs

Technicians stated that they do not need much time to start working with the software, just proper instructions and clear guidelines: "*Which part do I need to review visually? … And what do we do with the information we get from the software*?" (TC4). Some of the medical staff felt in contrast that: "*They [the technicians] need a lot of training and guidance in doing so*" (MC4).

Future use of the software

Almost all participants said that automated EEG-based detection could at least have an assisting role. Some were surprised that this kind of detection software was not already used in clinical practice: "*It already surprised me when I started working as a technician. And if we continue to review EEGs only visually for the next ten years. Well.. that sounds really old-fashioned*" (TC2); "*We need to enter the 21st century*" (MC4).

Some technicians thought automated EEG detection would never take over the review workload: "*I can't imagine that visual review by a technician will ever disappear*." (TC5). Others thought this might happen in the future, but not in the near future: "*I have been doing this for 5 years now, and for all these 5 years, the review process stayed the same. So I won't be surprised if we are still doing the same thing in 5 years' time*" (TC6); "*It is going to be a long time before you can really say: 'well, let's have the last few hours checked by a machine instead of a human expert'*" (TC2).

Most medical staff thought the visual review could in part be replaced by automated EEG-based detection: "*I think we can have some of the work done by the computer rather than just*

by manpower" (MC3), which would make the review process more efficient. Additionally, they saw an opportunity for the technician's job to evolve: "*Then technicians will get some task shifting. Getting some different work instead of, well, scrutinizing those EEGs, which is also a waste of their qualifications*" (MC5).

Both technicians and medical staff agreed that automated detection must be reviewed by human experts: "*You want to know if the detections are true and not for instance horizontal movements*" (TC6). You also need a back-up in case of software malfunction: "*I can imagine that such a program sometimes malfunctions*" (MC1).

Outpatient clinic neurologists

We identified 3 barriers and 7 enablers, see Table 1.

Quality current EEG process (recording, review and report)

All participants were satisfied with the quality of the current EEG review and report: "*The current EEG report is fine*" (MO2). They felt that clinical neurophysiology staff were doing a good job: "*they work meticulously and they know exactly what to look for*" (MO6); "*I always get an answer to my referral question*" (MO5). Some neurologists would like to see epileptiform abnormalities quantified: "*I like it myself if there is a kind of quantification of abnormalities, and if you have a previous EEG you can compare*" (MO3). Others said they found this information less relevant.

Necessity

The need for more efficiency was recognized by all participants, pointing to increasing EEG data and decreasing availability of personnel: "*I can see that reviewing EEGs takes a lot of time. And given the aging population … we can't expect that this amount of work can be done by humans alone*" (MO1). Respondents also felt that waiting times were too long: "*my only complaint are the long waiting times*" (MO5). Most neurologists supported review with automated detection if this meant shorter waiting times: "*If it helps speed things along I would rather have the software review the EEG*" (MO2).

Trust in performance

Some neurologists were hesitant to trust the automated software to review EEGs, and doubted it provided the same quality as human reviewers: "*If there is a chance that you miss something relevant. You don't want to miss that*" (MO4); "*I wonder, subtle ictal EEG changes,*

does the software detect that? I have my doubts" (MO6); "*Does it also detect slow activity?*" (MO4).

Others were fully confident that automated EEG-based detection software would only be implemented when it worked properly: "*I trust the opinion of the clinical neurophysiology neurologists*" (MO3); "*We [outpatient clinic neurologists] know that it is carefully looked after*" (MO1).

Future use of the software

All outpatient clinic neurologists thought automated detection software would be used within the next 5 years: "I think that will be the next step" (MO2); "That would be great, that we indeed are confident" (MO6); "I think that much more will be automated in the future" (MO1). One respondent thought it would only be used on a small scale: "I think it will be used for specific purposes" (MO4). Respondents mentioned that a control system must be built in, to ensure no important information would be missed: "Provided it is properly checked. I think you should check that randomly" (M07); "As long as there is a human check" (MO8). Finally, most neurologists stated they did not want to have a say in deciding whether or not automated EEG-based software would be implemented, but wanted to be informed: "It would be great if we were kept informed" (MO8).

Discussion

Nearly all participants expressed a need for a more efficient workflow and believed that automated EEG-based detection could play a role in this. They stressed that this kind of detection software could adapt EEG review to growing healthcare costs and personnel shortage. Furthermore, the EEG report producers group felt trends analyses have additional value, and that they are were willing to try the software.

We also noted significant challenges. The most prominent barrier was trust in the software, especially regarding automated seizure detection. Both producers and recipients of EEG reports feared the software would miss seizures or other important information. Most additional barriers were reported by technicians. Most believed that automated EEG-based detection could only be used as a supplement, mainly useful to quantify EEG spike detections, and not as (partial) replacement. Additionally, they fear a large quantity of false positives. Some technicians feared they would lose their ability to review long periods of EEG or that they might lose their jobs. A few participants doubted whether all technicians would be willing to try the software.

Limitations

This study was conducted with employees of two different EMUs from the same tertiary epilepsy center. The results may not be applicable elsewhere. However, our findings may serve as a baseline to consider challenges when implementing automated EEG-based detection software.

We chose to inform participants about automated detection software before the interviews. This approach may have introduced a bias, but ensured participants were well-informed. Additionally, there might be a selection bias in that respondents volunteered to participate, leaving open the possibility that nonrespondents felt differently.

The interviewer was familiar with the work and knew the respondents, which may have affected responses in an unknown direction.

Practice recommendations

Qualitative methods are a valuable tool in implementation research because they help to answer complex questions such as how and why efforts to implement best practices may succeed or fail [19]. We evaluated potential factors influencing the future implementation of automated EEG-based software.

Based on the results, we suggest the following recommendations regarding implementation. We learned that trust in the software needs to be gained, especially regarding the ability to detect seizures. Merely reading papers stating that automated EEG-based detection can be used safely does not inspire sufficient trust. Users need to acquire first-hand experience regarding the performance of automated EEG-based detection and must therefore be given time to do so. We propose reviewing EEGs both visually and with automated detection software. Furthermore, we suggest applying the software selectively, as we previously showed that the software did not detect all seizure types adequately, nor was it equally useful for all groups of patients. For example, reliability was limited in pediatric EEGs and short tonic seizures [10]. Hence, EEGs in these categories are better reviewed by the conventional methods, implying the need for triage.

Some technicians mentioned a fear of losing their ability to review long periods of EEG or even losing their job. We recommend that untrue fears be addressed as such. This can be achieved by providing sufficient information. We previously proposed a method where we use sampled visual review combined with automated EEG-based detection in a selection of EEGs [17]. With such a hybrid approach, technicians would still review EEGS, just to a lesser extent and for shorter periods. Furthermore, technicians need to be given the opportunity to learn new skills. This can be, for example, extracting more information from the EEG using trend analysis or improving the skill of reviewing difficult pediatric EEGs. Finally, outpatient clinic neurologists must also be kept informed regarding changes in the EEG review workflow using (educational) meetings.

Both information providing and training can be achieved by frequent educational meetings and feedback [20,21]. Outcome improves with, for example, shorter meetings, better attendance, shorter follow-up or interactive teaching methods [21].

The advantage of automated spike and seizure detection is improved efficiency. It would be useful to measure savings in time and money, after implementation, as would users' thoughts, attitudes, experiences and needs. Furthermore, the output of these detection software packages can also be used for other purposes, for instance averaging interictal epileptiform discharges for source localization and determining seizure onset zones [22,23]. Additionally, it would be informative to share experience with other EMUs.

This research gives an insight into (future) users' perspectives. Thereby we provide practice recommendations regarding implementation.

References

1. Jiang F, et al. Artificial intelligence in healthcare: past, present and future. Stroke and Vascular Neurology. 2017;2(4):230-43. doi: 10.1136/svn-2017-000101.

2. Abbasi B, Goldenholz DM. Machine learning applications in epilepsy. Epilepsia. 2019;60:2037–47. doi: 10.1111/epi.16333.

3. Binnie CD, Batchelor BG, Bowring PA, Darby CE, Herbert L, Lloyd DS et al. Computer-assisted interpretation of clinical EEGs. Electroencephalogr Clin Neurophysiol. 1978;44(5):575-85. doi: 10.1016/0013-4694(78)90125-6.

4. Gotman J, Ives JR, Gloor P. Automatic recognition of inter-ictal epileptic activity in prolonged EEG recordings Electroencephalogr Clin Neurophysiol. 1979;46(5):510-20. doi: 10.1016/0013-4694(79)90004-x.

5. Gotman J. Automatic recognition of epileptic seizures in the EEG. Electroencephalogr Clin Neurophysiol. 1982;54(5):530-40. doi: 10.1016/0013-4694(82)90038-4.

6. Baumgartner C, Koren JP. Seizure detection using scalp-EEG. Epilepsia. 2018;59 (S1):14-22. doi: 10.1111/epi.14052.

7. Supriya S, Siuly S, Wang H, Zhang Y. Automated epilepsy detection techniques from electroencephalogram signals: a review study. Health Inf Sci Syst. 2020;8(1):33. doi: 10.1007/s13755-020-00129-1.

8. Ganguly TM, Ellis CA, Tu D, Shinohara RT, Davis KA, Litt B, Pathmanathan J. Seizure Detection in Continuous Inpatient EEG: A Comparison of Human vs Automated Review. Neurology. 2022;98(22):e2224-e2232. doi: 10.1212/WNL.0000000000200267.

9. Reus EEM, Cox FME, Dijk van JG, Visser GH. Automated spike detection: which software package? Seizure. 2022a;95:33-7. doi: 10.1016/j.seizure.2021.12.012.

10. Reus EEM, Visser GH, Dijk van JG, Cox FME. Automated seizure detection in an EMU setting: are software packages ready for implementation? Seizure. 2022b;96:13-7. doi: 10.1016/j.seizure.2022.01.009.

11. Koren J, Hafner S, Feigl M, Baumgartner C., Systematic analysis and comparison of commercial seizure-detection software. Epilepsia 2021;62:426-38. doi: 10.1111/epi.16812.

12. Kamitaki BK, Yum A, Lee J, Rishty S, Sivaraaman K, Esfahanizadeh A, Mani R, Wong S. Yield of conventional and automated seizure detection methods in the epilepsy monitoring unit. Seizure. 2019 Jul;69:290-95. doi: 10.1016/j.seizure.2019.05.019.

13. Hopfengärtner R, Kasper BS, Graf W, Gollwitzer S, Kreiselmeyer G, Stefan H, et al. Automatic seizure detection in long-term scalp EEG using an adaptive thresholding technique: a validation study for clinical routine. Clin Neurophysiol. 2014 Jul;125(7):1346-52. doi: 10.1016/j.clinph.2013.12.104.

14. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, Griffey R, Hensley M. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health. 2011;38(2):65-76. doi: 10.1007/s10488-010-0319-7.

15. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care. 2007;19(6):349-357.

16. Cox FME, Reus EEM, Widman G, Zwemmer JNP, Visser GH. Epilepsy Monitoring Units can be safe places; a prospective study in a large cohort. Epilepsy Behav. 2020;102:106718. doi: 10.1016/j.yebeh.2019.106718.

17. Reus EEM, Visser GH, Cox FME. Using sampled visual EEG review in combination with automated detection software at the EMU. Seizure. 2020;80:96-9. doi: 10.1016/j.seizure.2020.06.002.

18. Flick, U. An introduction to qualitative research, 4th edition. Sage. 2009.

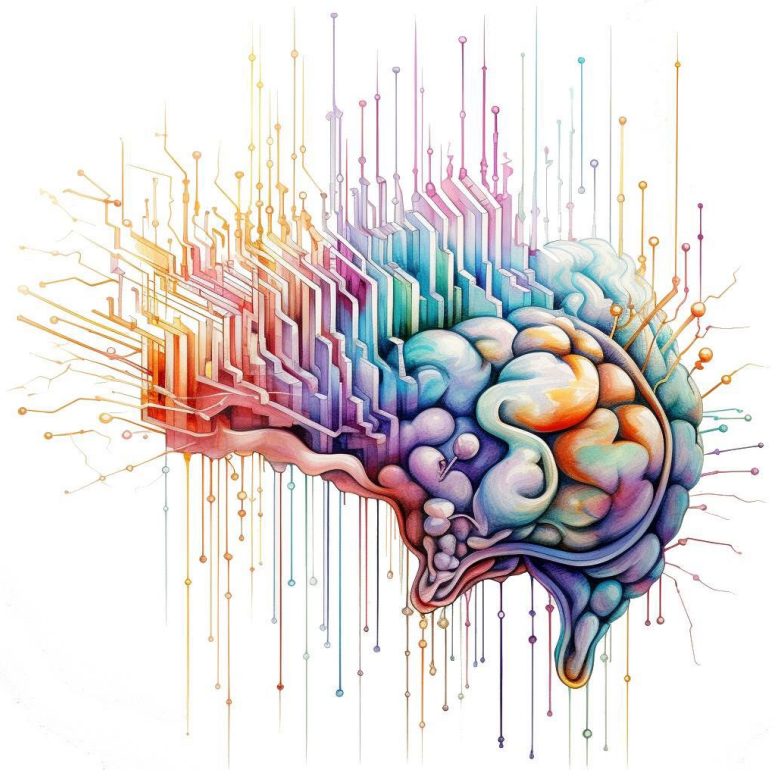
19. Hamilton AB, Finley EP. Qualitative methods in implementation research: An introduction. Psychiatry Res. 2019 Oct;280:112516. doi: 10.1016/j.psychres.2019.112516.

20. Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Audit and feedback: effects on professional practice and health care outcomes. Cochrane Database Syst Rev. 2006;2:CD000259. doi: 10.1002/14651858.CD000259.pub2. Update in: Cochrane Database Syst Rev. 2012;6:CD000259.

21. Forsetlund L, O'Brien MA, Forsén L, Reinar LM, Okwen MP, Horsley T, Rose CJ. Continuing education meetings and workshops: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev. 2021 Sep 15;9(9):CD003030. doi: 10.1002/14651858.CD003030.pub3.

22. Van Mierlo P, Strobbe G, Keereman V, Birot G, Gadeyne S, Gschwind M, et al. Automated long-term EEG analysis to localize the epileptogenic zone. Epilepsia Open 2017;30:322-33. doi: 10.1002/epi4.12066.

23. Spinelli L, Baroumand AG, Vulliemoz S, Momjian S, Strobbe G, van Mierlo P, et al. Semiautomatic interictal electric source localization based on long-term electroencephalographic monitoring: A prospective study. Epilepsia 2022. Epub ahead of print. doi: 10.1111/epi.17460.



CHAPTER 7

General discussion

Chapter 7 - General discussion

This thesis aimed to investigate whether automated spike and seizure detection can be used to review prolonged EEGs without compromising diagnostic quality. This research included the potential to combine automated detection with sampled visual review. In this chapter, we will provide a summary of our findings and discuss future implementation and potential applications of our research.

Most important findings and discussion

Automated spike detection (Chapter 4)

The software performed as well as human experts in detecting IEDs (interictal epileptiform discharges), making it a useful screening tool for finding these abnormalities in parts of the EEG not reviewed visually. When the software is set to its default (medium) setting, Persyst and Encevis achieved the highest sensitivities (82-95% for Persyst and 61-86% for Encevis). Both software packages had kappa scores similar to those of human experts, meaning they can identify most IEDs with a reasonable level of accuracy. These specificities are higher than reported in previous studies [1,2,3,4], probably due to the fact that we did not review single IEDs but measured whether 10-second epochs and 30-minute selections contained IEDs. This high sensitivity is necessary when using the software as a screening tool, although it may also generate false positives. Specificity for Persyst and Encevis was 88%. Consequently, it is imperative that the findings generated by the software are visually verified by human experts. This implies that the most significant time savings are in normal or nearly normal prolonged EEGs, as well as in EEGs displaying unifocal abnormalities, since there would be minimal or no findings necessitating verification.

Additionally, objective quantification of detected IEDs is another useful and reliable feature, as we showed in the calculation of the spike-wave index.

Automated seizure detection (Chapter 5)

The Persyst software performed best, resulting in a sensitivity of 95% (BESA) and 98% (Encevis, Persyst) when aiming to detect at least one of the seizures occurring within an individual record; sensitivity was 84% (BESA), 88% (Encevis) and 93% (Persyst) when aiming to detect all seizures. A recent study comparing the same three software packages reported a sensitivity of 77% for BESA, 78% for Encevis and 82% for Persyst [5]. Other earlier studies found that detection algorithms had a sensitivity for epileptic seizures between 73% and 96% [5,6,7,8]. Our study reported higher sensitivities than other earlier studies. This is due to the fact that we used another study design where we simulated how the software would be used

in a realistic EMU setting. We did so firstly by reviewing the combination of live human observation and offline review, comparing the performance of the clinical physiologists versus the software and secondly by using a broad scale of seizure types including generalized seizures, such as myoclonia and tonic seizures. However, these results show that even the best performing automated seizure detection algorithm does not detect all seizures when using automated seizure detection alone.

Sampled visual review in combination with automated detection software (Chapter 3)

A new review method, using sampled visual EEG review alongside automated detection software, yielded comparable electro-clinical diagnoses to those obtained through complete visual review. The notable benefit of this review approach lies in its potential reduction of overall reviewing time, particularly in our setting where numerous prolonged EEGs are conducted.

We found relying on automated detection software alone is not possible (Chapter 5). First, sampled visual review remains necessary to form an impression of the background activity and potential (focal or diffuse) dysfunctions or rhythmic delta activity. We showed that reviewing a one hour period while awake, another hour while asleep and an hour after wakefulness was enough for this purpose. Second, automated software packages were likely to miss short seizures or seizures with subtle or no EEG changes, such as myoclonic, short tonic and focal aware seizures. However, we found that patients could usually detect myoclonus and focal aware seizures themselves and report these events via the push button. This finding is supported by previous literature [9,10]. Furthermore, our proposed method using a targeted sampled review, including a period after waking, can also increase detection of myoclonic seizures that usually occur in this period.

This conclusion does not extend to generalized tonic seizures; our data show a large proportion of those remained undetected in the live setting. We suggest that EEGs, performed to investigate tonic seizures, should be thoroughly visually reviewed to avoid missing significant events.

Furthermore, all three automated detection software packages used in our research include quantitative EEG spectrograms. These are mathematically compressed assessments of raw EEG. These spectrograms fell outside the scope of our current research. However, previous research showed the use of quantitative EEG spectrograms can increase the detection of seizures [11].

The highest sensitivities were found for (focal to) generalized tonic clonic seizures and focal seizures with impaired awareness. The latter are regularly missed in the live setting by nurses and patients but are often detected by automated seizure detection [12].

Interviews regarding implementation of automated detection software (Chapter 6)

We identified several barriers and facilitators regarding implementation of automated spike and seizure detection software. Most important points of attention are: technicians and medical staff need to gain trust in the software, especially in its seizure detection ability, and they should be given the opportunity to experience the software's performance themselves. Some technicians feared losing their ability to review EEGs or even lose their job; such these fears need to be addressed by providing sufficient information. Although human review of the software-detected epileptiform discharges and seizures will always be necessary, technicians' review time will be reduced, and their job will change in nature but will not disappear. Technicians need to be given the opportunity to learn new skills, such as extracting more information from the EEG using trend analysis or improving their skill of reviewing more difficult paediatric EEGs. Lastly, the majority of neurologists in outpatient clinics perceive the utilization of automated detection methods as a beneficial advancement. However, they express the importance of being informed about any modifications in the EEG review workflow.

Current implementation process

In 2022, SEIN obtained a license for the Persyst software package, which includes features for spike detection, seizure detection, and trend analyses. The spike detection module offers three sensitivity settings, namely low, medium (default), and high. The software detects spikes and clusters them per electrode based on maximum amplitude. For each detected spike, the software generates 1-second epochs centered around the event, with an average signal per electrode. It also allows for the review of individual potentially abnormal findings.

With respect to seizure detection, the manual specifies that the algorithm can detect ictal patterns with a minimum duration of 11 seconds (version 13) or 8 seconds (version 14). The output of this module is a list of timed seizure detections.

Additionally, the software offers various options for quantitative EEG spectograms, based on parameters such as amplitude, frequency, rhythmicity, and degree of electrographic asymmetry [11].

We utilized the findings of our qualitative study to optimize the implementation process, and as a result, the detection software is now employed for the following purposes.

1) Calculating Spike-wave index.
   Useful in calculating a first as well as a follow-up record, for instance after a prednisolone treatment.

2) Reviewing multiple prolonged EEGs (20 to 48 hours) per week in teenagers and adults (> 16 years) using sampled review in combination with automated EEG-based detection.

   We use this for the following indications:

   a. Psychogenic non-epileptic seizures.
      Rule out any (inter)ictal epileptiform discharges.

   b. Temporal lobe epilepsies.
      Detect seizures and interictal epileptiform discharges.

   c. Interictal discharges.
      Detect and quantify interictal epileptiform discharges.

   d. Generalized epilepsy such as JME.
      Detect seizures and interictal epileptiform discharges.

The greatest advantage proved to be increased efficiency in the EEG workflow. Given the current shortage of personnel and lengthy waiting lists, this improvement could potentially enable the examination of a greater number of EEG hours.

Future perspectives

We plan to expand the use of automated detection software to review an increased number of prolonged EEGs, including those of presurgical patients. In this patient group, EEG records can last up to 5 days, making their analysis very labour intensive. In this patient group it could be particularly useful to measure the ratio of epileptiform discharges in the left versus the right hemisphere to determining the seizure onset zone. We are also looking to expand the patient group to children under 16 years old. For instance, absence epilepsy appears particularly suitable for reviewing using automated detection software.

Additionally, we aim to optimize the use of trend analyses provided by the software to improve the detection of seizures and other relevant EEG findings [11]. Furthermore, it can be helpful to also use other modalities for seizure detection, such as surface electromyography, accelerometry, video detection systems and mattress sensors for seizures with motor manifestations. And heart rate and oxygen saturation for ictal changes in physiologic parameters [13].

Furthermore, the output of these detection software packages can also be used for other purposes. For instance, labeling spikes for scientific purposes, averaging or source localization may prove useful [14,15].

Finally, it is possible that automated detection may also be implemented in the online setting. This can improve patient safety, create live automated documentation of seizures, and perhaps even computer-based neurological and neuropsychological testing during and after seizures [16].

Acknowledgements

We are grateful to Persyst Development Corporation (San Diego, CA, USA), AIT Austrian Institute of Technology, (Vienna, Austria) and BESA Epilepsy (Germany) for allowing the use of trial licenses of their software.

References

1. Scheuer ML, Bagic A, Wilson SB. Spike detection: inter-reader agreement and a statistical Turing test on a large data set. Clin Neurophysiol 2016;128:243–50. https://doi.org/10.1016/j.clinph.2016.11.005.

2. Jing J, Sun H, Kim JA, Herlopian A, Karakis J, Ng M, et al. Development of Expert-Level Automated Detection of Epileptiform Discharges During Electroencephalogram Interpretation. JAMA Neurol 2020;77:103–8. https://doi.org/10.1001/jamaneurol.2019.3485.

3. Halford JJ, Brandon Westover M, LaRoche SM, Macken MP, Kutluay E, Edwards JC, et al. Interictal Epileptiform Discharge Detection in EEG in Different Practice Settings. J Clin Neurophysiol 2018;35:375–80. https://doi.org/10.1097/WNP.0000000000000492.

4. Fürbass F, Kural MA, Gritsch G, Hartmann M, Kluge T, Beniczky S. An artificial intelligence-based EEG algorithm for detection of epileptiform EEG discharges: validation against the diagnostic gold standard. Clin Neurophysiol 2020;131:1174–9. https://doi.org/10.1016/j.clinph.2020.02.032.

5. Koren J, Hafner S, Feigl M, Baumgartner C. Systematic analysis and comparison of commercial seizure-detection software. Epilepsia 2021;62:426–38. https://doi.org/10.1111/epi.16812.

6. Baumgartner C, Koren JP. Seizure detection using scalp-EEG. Epilepsia 2018;59:14–22. https://doi.org/10.1111/epi.14052. a.

7. Baumgartner C, Koren JP, Rothmayer M. Automatic computer-based detection of epileptic seizures. Front Neurol 2018;9:639. https://doi.org/10.3389/fneur.2018.00639. b.

8. Kamitaki BK, Yum A, Lee J, Rishty S, Sivaraaman K, Esfahanizadeh A, Mani R, Wong S. Yield of conventional and automated seizure detection methods in the epilepsy monitoring unit. Seizure 2019;69:290–5. https://doi.org/10.1016/j.seizure.2019.05.019.

9. Hoppe C, Poepel A, Elger CE. Epilepsy: accuracy of patient seizure counts. Arch Neurol 2007;64:1595–9. https://doi.org/10.1001/archneur.64.11.1595.

10. Atkinson M, Hari K, Schaefer K, Shah A. Improving safety outcomes in the epilepsy monitoring unit. Seizure 2012;21:124–7. https://doi.org/10.1016/j.seizure.2011.10.004.

11. Goenka A, Boro A, Yozawitz E. Comparative sensitivity of quantitative EEG (QEEG) spectrograms for detecting seizure subtypes. 2018. Seizure, 55, 70-5. https://doi.org/10.1016/j.seizure.2018.01.008.

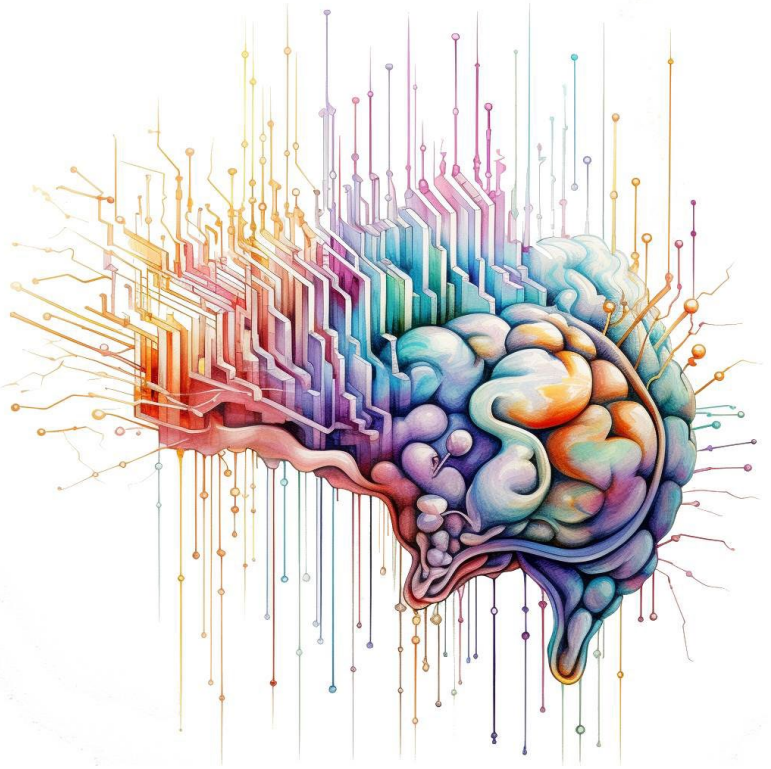
12. Rommens N, Geertsema E, Jansen Holleboom L, Cox FME, Visser GH. Improving staff response to seizures on the epilepsy monitoring unit with online EEG seizure detection algorithms. Epilepsy Behav 2018;84:99–104. https://doi.org/10.1016/j.yebeh.2018.04.026.

13. Beniczky S, Jeppesen J. Non-electroencephalography-based seizure detection. Curr Opin Neurol. 2019 Apr;32(2):198-204. doi: 10.1097/WCO.0000000000000658.

14. Van Mierlo P, Strobbe G, Keereman V, Birot G, Gadeyne S, Gschwind M, et al. Automated long-term EEG analysis to localize the epileptogenic zone. Epilepsia Open 2017;30:322-33. doi: 10.1002/epi4.12066.

15. Spinelli L, Baroumand AG, Vulliemoz S, Momjian S, Strobbe G, van Mierlo P, et al. Semiautomatic interictal electric source localization based on long-term electroencephalographic monitoring: A prospective study. Epilepsia 2022. Epub ahead of print. doi: 10.1111/epi.17460.

16. Touloumes G, Morse E, Chen WC, Gober L, Dente J, Lilenbaum R, et al. Human bedside evaluation versus automatic responsiveness testing in epilepsy (ARTiE). Epilepsia (2016) 57:e28–32. doi: 10.1111/epi.13262.



CHAPTER 8

Summary

Chapter 8 - Summary

The aim of this thesis was to investigate whether automated spike and seizure detection can be used to review prolonged EEGs without compromising the quality of the results. This research included the potential to combine automated detection with sampled visual review.

Chapter 2 - The spike–wave index (SWI)

The thesis started with a study on calculation of the spike-wave index (SWI), which quantifies the amount of epileptiform activity occurring during sleep. This is a key feature in the diagnosis of electrical status epilepticus during slow-wave sleep (ESES) [1].

The research compared the performance of a commercially available spike detection algorithm with human expert consensus. The results showed that SWIs estimated by human experts did not differ from the SWIs calculated by the automated spike detection algorithm in its 'low' (most strict) sensitivity setting ($P = 0.67$).

Chapter 3 - Sampled visual EEG review in combination with automated detection software

In this study we compared electroclinical diagnosis using sampled visual review in combination with automated spike and seizure detection software to complete visual review. The sampled EEG comprised the first hour of recording during wakefulness, including hyperventilation provocation and intermittent photic stimulation, the first hour of sleep, and the first half-hour after sleep the next morning, which was included due to the circadian distribution of some generalized epilepsies (especially JME). Additionally, EEG and video periods around events marked by nurse or patient were reviewed.

Our findings indicate that the electroclinical diagnosis based on sampled visual review combined with automated detection software did not differ from the diagnosis based on complete visual review. This means the detection software was able to identify all records containing epileptiform abnormalities and epileptic seizures.

Chapter 4 - Automated spike detection: Which software package?

In this chapter, we evaluated the performance of three commercial automated spike detection software packages (Persyst, Encevis, and BESA) to determine which performed the best at our EMU setting. A heterogeneous dataset was annotated by three human experts and by all three software packages. We defined the gold standard as the combined detections of the experts and compared this to each software package. For each 30-minute selection and for each 10-second epoch, we measured whether or not interictal epileptiform activity occurred.

The sensitivity of Persyst in the default (medium) setting was 95% for 30-minute selections and 82% for 10-second epochs. The sensitivity of Encevis was 86% for 30-minute selections and 61% for 10-second epochs. The specificity for both packages was 88% for 30-minute selections and 96%-99% for the 10-second epochs. The inter-rater agreement between Persyst and Encevis and the experts was similar to that between the experts (0.67–0.83 versus 0.63–0.67). The sensitivity for BESA was 40% and specificity 100%. The inter-rater agreement (0.25) was low.

These results indicate that the automated spike detection software in Persyst performs better than Encevis and BESA packages and is comparable to human review when reviewing 30-minute selections and 10-second epochs for the detection of interictal epileptiform activity.

Chapter 5 - Automated seizure detection in an EMU setting: Are software packages ready for implementation?

We evaluated the reliability of automated seizure detection software in an EMU setting, using three commercial packages: Persyst, Encevis, and BESA. To assess the practical use of these software packages, we compared their seizure detection performance to that of online human observation (performed by trained nurses) in combination with both conventional review by technicians and automated seizure detection using the three software packages.

Seizure detection sensitivity was 98% for both Encevis and Persyst, and 95% for BESA, when a positive result was defined as detection of at least one seizure occurring within an individual record. When positivity was defined as recognition of all seizures, sensitivity was 93% for Persyst, 88% for Encevis, and 84% for BESA. Technicians' review had a sensitivity of 100% at the record level and 98% at the seizure level. The median false positive rate per 24 hours was 1.7 for Persyst, 2.4 for BESA, and 5.5 for Encevis. Sensitivity was especially high for focal seizures with impaired awareness and (focal to bilateral) tonic-clonic seizures. Sensitivity was low for myoclonia, short tonic seizures and focal aware seizures. Furthermore, the automated detection software detected five additional seizures, which were initially missed in the offline review by the technician.

In conclusion, automated seizure detection software does not perform as well as human technicians. However, it can be useful in an EMU setting as long as the user is aware of its limitations. Among the software packages tested, Persyst performed best.

Chapter 6 - Automated spike and seizure detection: Are we ready?

The objective of this study is to identify the factors that impede ('barriers') and facilitate ('facilitators') the implementation of automated EEG-based detection in clinical practice. The

implementation of changes in healthcare practices is a challenging task, and its success largely depends on the acceptability of the innovation by healthcare professionals and stakeholders, and their willingness to try it [2].

Semi-structured interviews were conducted with 22 participants, including nine technicians, five neurologists and physician assistants from a clinical neurophysiology department and eight outpatient clinic neurologists. The interviews were recorded and transcribed for analysis.

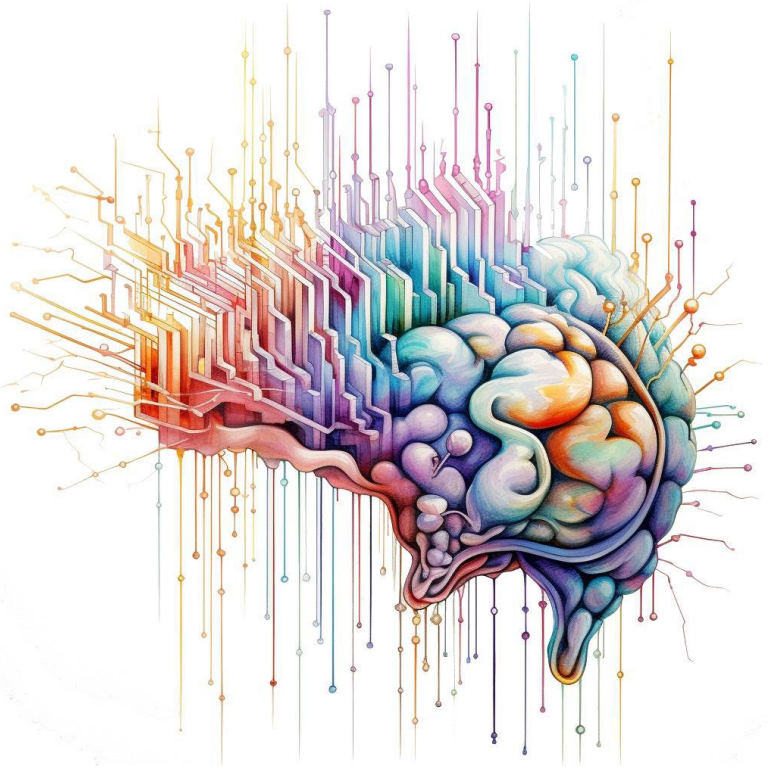
We identified 14 barriers and 14 enablers for future implementation. Most barriers were reported by technicians. The most prominent barrier was lack of trust in the software, especially regarding seizure detection and false positive results. Additionally, technicians feared losing their EEG review skills or even their jobs. Most commonly reported enablers included potential efficiency in the EEG workflow, the opportunity for quantification of EEG findings and the willingness to try the software.

This study provides insight into the perspectives of users and offers recommendations for implementing automated spike and seizure detection in EMUs.

References

1. Tassinari CA, Rubboli G, Volpi L, et al. Encephalopathy with electrical status epilepticus during slow sleep or ESES syndrome including the acquired aphasia. Clin Neurophysiol 2000;111:S94–S102. https://doi.org/10.1016/s1388-2457(00)00408-9.

2. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, Griffey R, Hensley M. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health. 2011 Mar;38(2):65-76. https://doi.org/10.1007/s10488-010-0319-7.



APPENDICES

Samenvatting

List of publications

Acknowledgements

Curriculum Vitae

Nederlandse samenvatting

Het doel van deze thesis was om te onderzoeken of geautomatiseerde piek- en aanvalsdetectie kan worden gebruikt om langdurige EEGs te beoordelen zonder de kwaliteit hiervan in gevaar te brengen. Daarbij omvatte deze thesis ook een nieuwe methode waarbij automatische detectie werd gecombineerd met gesamplede visuele beoordeling.

Hoofdstuk 2 - de Spike-wave index (SWI)

Dit promotietraject begon met een onderzoek naar de berekening van de spike-wave index (SWI), die de hoeveelheid epileptiforme activiteit tijdens de slaap kwantificeert. Dit is een belangrijk onderdeel van de diagnose electrical status epilepticus during slow-wave sleep (ESES) [1]. Het onderzoek vergeleek de prestaties van een commercieel beschikbaar piekdetectie algoritme met de consensus van menselijke experts. De resultaten toonden aan dat SWI geschat door menselijke experts niet verschilden van de SWI berekend door de automatische piekdetectie in de 'low' (meest strenge) sensitiviteit setting (P = 0,67).

Hoofdstuk 3 - Gesamplede visuele EEG-beoordeling in combinatie met beoordeling door automatische piek- en aanvalsdetectie

In deze studie vergeleken wij de electroklinische diagnose gebaseerd op gesamplede visuele beoordeling in combinatie met automatische piek- en aanvalsdetectie met volledige visuele beoordeling. Het gesamplede EEG omvatte het eerste uur van de registratie tijdens waak, inclusief hyperventilatieprovocatie en intermitterende fotostimulatie, het eerste uur van de slaap, en de eerste half uur na de slaap de volgende ochtend. Het laatste werd opgenomen vanwege de circadiane distributie van sommige gegeneraliseerde epilepsiesyndromen (vooral JME). Daarnaast werd zowel de video als EEG visueel bekeken rondom de door verpleging of patiënt gemelde gebeurtenissen.
De electroklinische diagnose op basis van gesamplede visuele beoordeling in combinatie met geautomatiseerde detectiesoftware verschilde niet van de diagnose op basis van volledige visuele beoordeling. Dit betekent dat de detectiesoftware in staat was om alle registraties te identificeren die epileptiforme afwijkingen of epileptische aanvallen bevatten.

Hoofdstuk 4 - Automatische piekdetectie: welk softwarepakket?

In dit hoofdstuk hebben we de prestaties van drie commerciële automatische piekdetectiesoftwarepakketten (Persyst, Encevis en BESA) geëvalueerd om te bepalen welke het beste presteerde in onze EMU-setting. Interictale epileptiforme afwijkingen in een heterogene dataset werd geannoteerd door drie menselijke experts en door alle drie de softwarepakketten. We definieerden de gouden standaard als de gecombineerde detecties van de experts en vergeleken dit met de detecties van elk softwarepakket. Voor elke selectie van 30 minuten en voor elke 10-seconden epoch hebben we gemeten of interictale epileptiforme activiteit al dan niet voorkwam. De sensitiviteit van Persyst in de standaard ('medium') instelling was 95% voor selecties van 30 minuten en 82% voor 10-seconden-epochs. De sensitiviteit van Encevis was 86% voor selecties van 30 minuten en 61% voor 10-seconden-epochs. De specificiteit voor beide pakketten was 88% voor selecties van 30 minuten en 96%-99% voor de 10-seconden-epochs. De interbeoordelaarsovereenkomst tussen Persyst en Encevis en de experts was vergelijkbaar met die tussen de experts (0,67-0,83 versus 0,63-0,67). De sensitiviteit voor BESA was 40% en de specificiteit 100%. De interbeoordelaarsovereenkomst (0,25) was laag.
Deze resultaten geven aan dat de automatische piekdetectiesoftware van Persyst beter presteert dan de Encevis- en BESA-pakketten en vergelijkbaar is met menselijke beoordeling van interictale epileptiforme activiteit in selecties van 30 minuten en 10-seconden-epochs.

Hoofdstuk 5 - Geautomatiseerde detectie van epileptische aanvallen in een EMU omgeving: zijn softwarepakketten klaar voor implementatie?

We hebben drie commerciële automatische aanvalsdetectie softwarepakketten gevalideerd: Persyst, Encevis en BESA. Om de praktijk zo goed mogelijk te benaderen, hebben we de detecties van epileptische aanvallen door live observatie (verpleegkundigen) in combinatie met zowel conventionele beoordeling door laboranten als in combinatie met geautomatiseerde aanvalsdetectie softwarepakketten bekeken. De gevoeligheid van de aanvalsdetectie was 98% voor zowel Encevis als Persyst en 95% voor BESA, waarbij een positief resultaat werd gedefinieerd als detectie van ten minste één aanval die binnen een individueel registratie plaatsvond. Wanneer een positief resultaat werd gedefinieerd als herkenning van alle aanvallen binnen een registratie, was de gevoeligheid 93% voor Persyst, 88% voor Encevis en 84% voor BESA. De beoordeling door laboranten had een

gevoeligheid van 100% op registratieniveau en 98% op aanvalsniveau. De mediane fout-positieve ratio per 24 uur was 1,7 voor Persyst, 2,4 voor BESA en 5,5 voor Encevis. De sensitiviteit van automatische detectiesoftwarepakketten was vooral hoog voor focale aanvallen met verminderde gewaarwording en (focaal naar bilateraal) tonisch-clonische aanvallen. De sensitiviteit was laag voor myoclonieën, korte tonische aanvallen en focale aanvallen met intacte gewaarwording. Daarnaast detecteerde de automatische detectiesoftware vijf extra aanvallen die aanvankelijk gemist waren door de laborant.

Concluderend presteert automatische aanvalsdetectiesoftware niet zo goed als laboranten. Echter kan automatische aanvalsdetectie toch nuttig zijn bij het beoordelen van langdurige EEGs zolang de gebruiker zich bewust is van de beperkingen. Onder de geteste softwarepakketten presteerde Persyst het beste.

Hoofdstuk 6 - Geautomatiseerde piek- en aanvalsdetectie: zijn we er klaar voor?

Implementatie van innovaties in de gezondheidszorg is een uitdagende taak en het succes ervan hangt grotendeels af van de acceptatie door zorgprofessionals en hun bereidheid om het te proberen [2]. Het doel van deze studie is om factoren te identificeren die de implementatie van automatische piek- en aanvalsdetectie in de klinische praktijk belemmeren (‘barrières’) en bevorderen (‘facilitators’). Er werden semigestructureerde interviews gehouden met 22 deelnemers, waaronder negen KNF-laboranten, vijf KNF neurologen en physician assistants werkzaam op een KNF-afdeling en acht poli-neurologen. De interviews werden opgenomen en getranscribeerd voor analyse. We hebben 14 barrières en 14 facilitators geïdentificeerd voor toekomstige implementatie. De meeste barrières werden gemeld door laboranten. Zij noemden het gebrek aan vertrouwen in de software het vaakst, met name met betrekking tot de detectie van aanvallen en vals positieve resultaten. Daarnaast waren de laboranten bang om hun vaardigheden voor het beoordelen van EEG's of zelfs hun baan te verliezen. De meest genoemde faciliterende factoren waren de mogelijke efficiëntie in de EEG workflow, de mogelijkheid om EEG-bevindingen te kwantificeren en hun bereidheid om de software te proberen. Deze studie biedt inzicht in de perspectieven van gebruikers en biedt aanbevelingen voor de implementatie van geautomatiseerde piek- en aanvalsdetectie in de Epilepsy Monitoring Unit (EMU).

Referenties

1. Tassinari CA, Rubboli G, Volpi L, et al. Encephalopathy with electrical status epilepticus during slow sleep or ESES syndrome including the acquired aphasia. Clin Neurophysiol 2000;111:S94–S102. https://doi.org/10.1016/s1388-2457(00)00408-9.

2. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, Griffey R, Hensley M. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health. 2011 Mar;38(2):65-76. https://doi.org/10.1007/s10488-010-0319-7.

List of publications

This thesis

Reus EEM, Visser GH, Cox FME. Determining the Spike-Wave Index Using Automated Detection Software. *J Clin Neurophysiol*. 2021;38(3):198-201. doi:10.1097/WNP.0000000000000672.

Reus EEM, Visser GH, Cox FME. Using sampled visual EEG review in combination with automated detection software at the EMU. Seizure. 2020;80:96-9. doi: 10.1016/j.seizure.2020.06.002.

Reus EEM, Cox FME, Dijk van JG, Visser GH. Automated spike detection: which software package? Seizure. 2022a;95:33-7. doi: 10.1016/j.seizure.2021.12.012.

Reus EEM, Visser GH, Dijk van JG, Cox FME. Automated seizure detection in an EMU setting: are software packages ready for implementation? Seizure. 2022b;96:13-7. doi: 10.1016/j.seizure.2022.01.009.

Reus EEM, Visser GH, Sommers-Spijkerman MPJ, van Dijk JG, Cox FME. Automated spike and seizure detection: Are we ready for implementation? Seizure. 2023;108:66-71. doi:10.1016/j.seizure.2023.04.010.

Other

Rubboli G, Beniczky S, Claus S, Canevini MP, Kahane P, Stefan H, Boas WVE, Velis DN, **Reus EEM**, Gil-Nagel A, Steinhoff BJ, Trinka E, Ryvlin P. A European survey on current practices in epilepsy monitoring units and implications for patients' safety. Epilepsy Behav. 2015;44,179-84. doi:10.1016/j.yebeh.2015.02.004.

Bennis FC, Geertsema EE, Velis DN, **Reus EEM**, Visser GH. The use of single bipolar scalp derivation for the detection of ictal events during long-term EEG monitoring. *Epileptic Disord*. 2017;19(3):307-314. doi:10.1684/epd.2017.0926

Cox FME, **Reus EEM**, Visser GH. Timing of first event in inpatient long-term video-EEG monitoring for diagnostic purposes. *Epilepsy Res*. 2017;129:91-94. doi:10.1016/j.eplepsyres.2016.12.00

Cox FME, **Reus EEM**, Widman G, Zwemmer JNP, Visser GH. Epilepsy Monitoring Units can be safe places; a prospective study in a large cohort. Epilepsy Behav. 2020;102:106718. doi: 10.1016/j.yebeh.2019.106718.

Reus EEM, Cox FME. Juveniele myoclonusepilepsie: vergeet praxisinductie niet. Tijdschr neurolneurochir. 2023;124,4:165–8.

Curriculum Vitae

Elise Reus was born on July 21, 1988, in Hoorn, the Netherlands. In 2006 she obtained her 'VWO' diploma at the Oscar Romero in Hoorn. She then pursued a BSc Psychology and MSc Psychology with a specialization in clinical neuropsychology at the University of Amsterdam, graduating in 2012.

In 2012 she started her career at the clinical neurophysiology department of 'Stichting Epilepsie Instellingen Nederland' (SEIN). In the following year, in 2013, she commenced the master Physician Assistant at Hogeschool Rotterdam, successfully earning her degree in 2016.

Her interest with automated spike and seizure detection was piqued in 2018/2019, leading her to initiate a formal PhD trajectory in 2020. Alongside her research endeavors, Elise actively serves as a physician assistant for the clinical neurophysiology department at SEIN, a role she will continue post-dissertation.

Elise lives in Haarlem together with Mark and their two children Lieuwe and Doris.

Curriculum Vitae - Nederlands

Elise Reus werd geboren op 21 juli 1988 in Hoorn, Nederland. In 2006 behaalde ze haar VWO-diploma aan het Oscar Romero in Hoorn. Vervolgens volgde ze een BSc Psychologie en MSc Psychologie met een specialisatie in klinische neuropsychologie aan de Universiteit van Amsterdam, waar ze in 2012 afstudeerde.

In 2012 begon ze haar carrière op de afdeling klinische neurofysiologie van de 'Stichting Epilepsie Instellingen Nederland' (SEIN). Het daaropvolgende jaar, in 2013, startte ze de master Physician Assistant aan de Hogeschool Rotterdam, waar ze in 2016 succesvol haar diploma behaalde.

Haar interesse voor geautomatiseerde spike- en aanvalsdetectie werd gewekt in 2018/2019, wat haar ertoe bracht om in 2020 een formele PhD-traject te starten. Naast haar onderzoeksinspanningen vervult Elise actief de rol van physician assistant op de afdeling klinische neurofysiologie bij SEIN, een rol die ze na haar proefschrift zal blijven vervullen.

Elise woont in Haarlem samen met Mark en hun twee kinderen Lieuwe en Doris.

Dankwoord (acknowledgements)

Om te beginnen wil ik mijn promotor prof. dr. J.G. van Dijk bedanken. Beste Gert, dank voor zowel de mogelijkheid om mijn promotieonderzoek aan de Universiteit Leiden te kunnen uitvoeren als voor jouw tijd en de waardevolle en altijd duidelijke input die jij mij hebt gegeven.

Daarnaast wil ik mijn twee copromotoren, dr G.V. Visser en dr F.M.E. Cox, bedanken: dank voor mogelijkheid om aan dit traject te kunnen beginnen en voor de begeleiding gedurende het traject. Zonder jullie steun en inbreng had ik dit nooit tot een goed einde kunnen brengen, hopelijk stopt onze (wetenschappelijke) samenwerking hier niet.

Gerhard, ik kan me herinneren dat je zei: 'ik moet je toch ergens mee bezig houden' en je had gelijk, ik ben er een flinke tijd zoet mee geweest. Fieke, je bent altijd erg snel met jouw feedback op mijn stukken en vaak zitten we op één lijn met onze manier van denken, echter ik waardeer het ook altijd erg dat je je uitspreekt wanneer je een andere mening hebt.

I also want to thank prof. dr. J.W. Sander. Dear Ley, I appreciate your constructive feedback on my writing skills. Your editing significantly improved readability (and also significantly shorted the text).

Beste Jack, Boriana, Eva, Robert en KNF-laboranten: dank voor jullie vertrouwen om ons het gebruik van automatische piek- en aanvalsdetectiesoftware te laten uitzoeken. En ook dank voor jullie aanpassingsvermogen en inspanning die de (toekomstige) veranderingen in werkwijze van jullie heeft gevraagd en vraagt. Eva en Boriana, mijn paranimfen: wat ben ik blij dat jullie naast me staan op de bijzondere dag van mijn promotie! Holger en Robert, dank voor alle technische ondersteuning bij het installeren en werken met de software. En dank aan de laborantengroep voor jullie meedenken, hulp en steun aan dit onderzoek.

Ook dank aan Evelyn en alle andere collega's op de AIOS-kamer voor de wandelingen in de lunchpauzes en de ondertussen welbekende cappuccino bij de Italiaan. Dit heeft heel wat middagen beter gemaakt.

EMU-verpleging: dank voor jullie interesse en aanmoedigen. Ik kan altijd bij jullie even m'n verhaal kwijt (terwijl jullie ondertussen 32 camerabeelden en 8 EEG-tracés in de gaten houden).

Ook wil ik de dames van het secretariaat bedanken. Iedereen weet dat er geen EEG geregistreerd kan worden zonder jullie ondersteuning.

Speciale dank aan de collega's (in Heemstede én in Zwolle) die mee hebben gedaan aan mijn interviews en voor de inzichten die jullie me hiermee hebben gegeven.

Tot slot wil ik mijn familie en vrienden bedanken voor jullie steun en interesse in dit onderzoek. In het bijzonder mijn ouders voor de aanmoedigingen tijdens dit promotietraject en bij de totstandkoming van dit eindwerk. Dankjulliewel dat jullie er voor me zijn.

En last but not least: Lieve Mark, de afgelopen jaren is er veel veranderd in ons leven met de komst van Lieuwe en Doris. Ondanks alles heb je me altijd de tijd en ruimte gegeven om dit proefschrift goed af te ronden. Dankjewel voor al je steun en liefde de afgelopen jaren.