Automated spike detection: which software package?

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

Seizure: European Journal of Epilepsy 95 (2022) 33–37

Contents lists available at ScienceDirect

Seizure: European Journal of Epilepsy

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

Automated spike detection: Which software package?

E.E.M. Reus [a,*], F.M.E. Cox [a], J.G. van Dijk [b], G.H. Visser [a]

[a] *Department of Clinical Neurophysiology, Stichting Epilepsie Instellingen Nederland*
[b] *Department of Neurology, Leiden University Medical Centre, Leiden, the Netherlands*

A R T I C L E I N F O

Keywords:
Automatic detection
Spike detection
Epilepsy
Video-EEG monitoring
EMU

A B S T R A C T

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

1. 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.

2. Methods

2.1. 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

* Corresponding author at: Stichting Epilepsie Instellingen Nederland (SEIN). PO Box 540, 2130 AM Hoofddorp, the Netherlands.
E-mail address: ereus@sein.nl (E.E.M. Reus).

https://doi.org/10.1016/j.seizure.2021.12.012
Received 27 October 2021; Received in revised form 22 December 2021; Accepted 24 December 2021
Available online 26 December 2021

waived according to Dutch rules.

2.2. 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 (Fig. 1).

2.3. 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).

2.3.1. 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.

2.3.2. 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].

2.3.3. 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.

2.3.4. 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.

2.4. 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 s [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

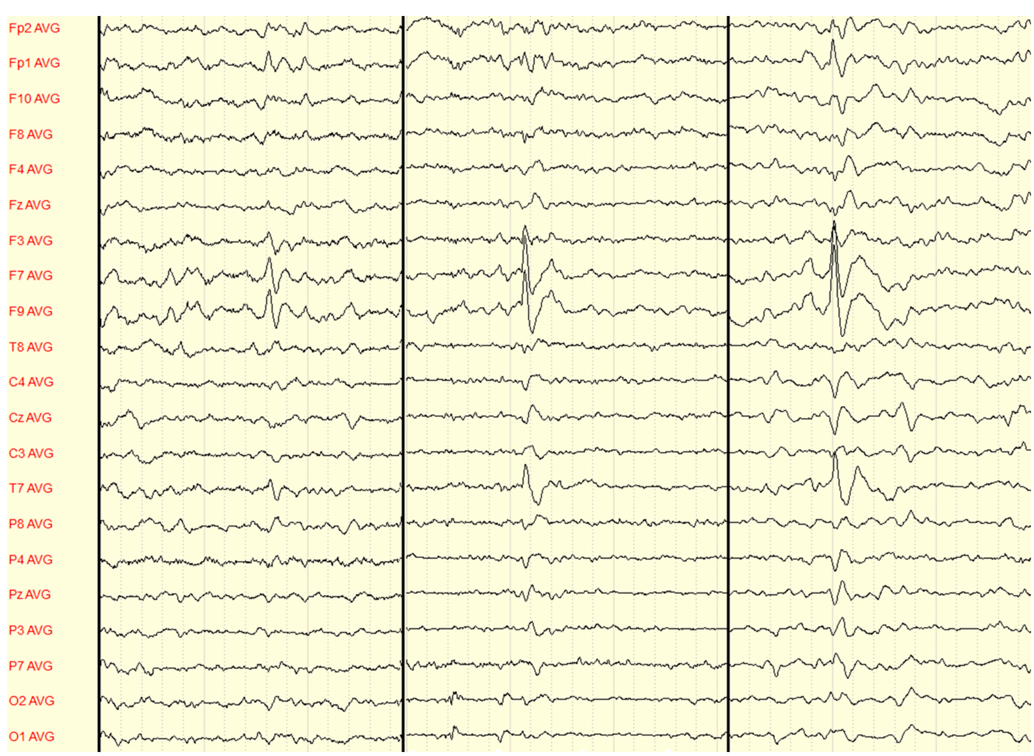

Fig. 1. On the left side is a possible IED. On the middle and right side are definite IEDs. IEDs are shown in an average reference montage, but reviewers could use any montage. IED = Interictal Epileptiform Discharge.

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.).

3. Results

3.1. 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.

3.2. Performance

3.2.1. 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.

All three software packages missed epileptiform discharges in one particular record containing multiple IEDs, which could be interpreted as eye closure sensitivity (Fig. 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.

3.2.2. 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 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.

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.

4. 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 pediatric 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

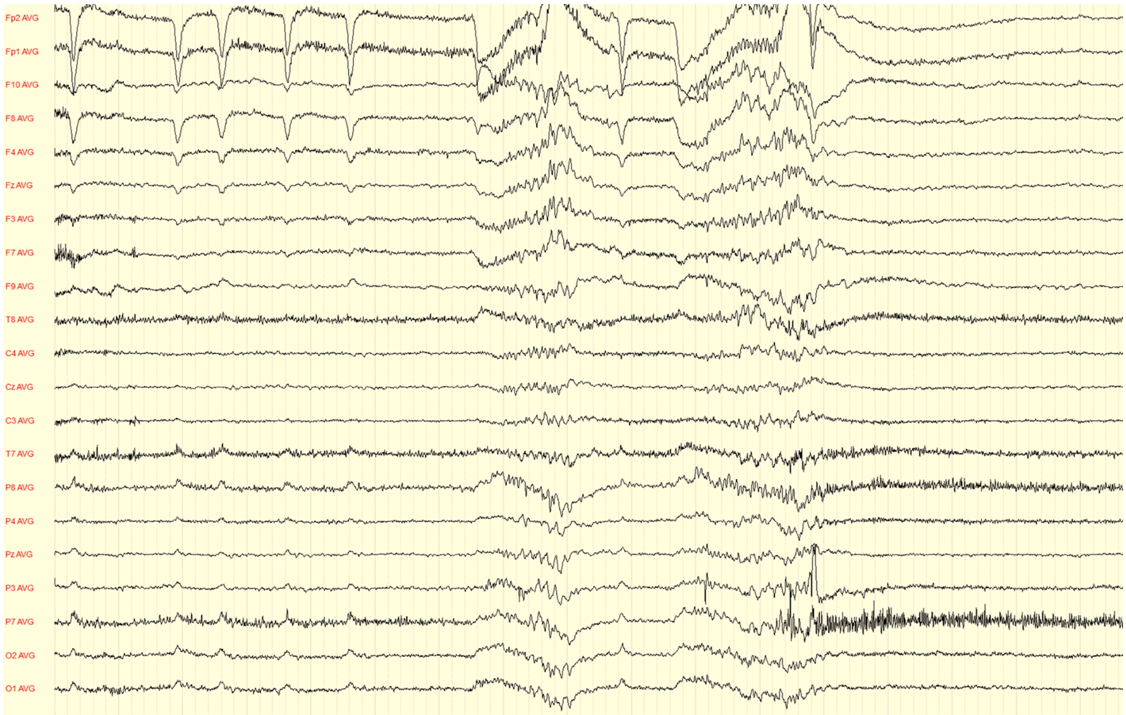

Fig. 2. Example of the eye closure sensitivity all three software packages missed. The example is shown in an average reference montage.

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	99% (CI	0.57 (CI
BESA -	150	174	4731	46–54%)	98–99%)	0.53–0.61)

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.

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.

5. Conclusions

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.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Neither this work nor the authors have ever been funded by Persyst Development Corporation, AIT Austrian Institute of Technology or BESA Epilepsy.

Declaration of competing interest

None of the authors has any conflict of interest to disclose.

Acknowledgment

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. None had any influence on the design, results or decision to publish this paper.

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 LaA, Siew-Na Lim LaA, Frank Lien LbB, Wua Tony T. * Duration of electroencephalographic recordings in patients with epilepsy. Seizure 2013;22: 438–42. https://doi.org/10.1016/j.seizure.2013.02.016. https://doi.org/.

[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] Koca * Guray, Morkavukb Gulin, Akkayab Efdal, Karadasa Omer, Leventoglub Alev, Unayc Bulent, Gokcild Zeki. 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.

[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) 2013. https://www.randomizer.org/ [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. https://doi.org/10.1016/j.cnp.2017.07.002. https://doi.org/.

[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. In: 2011 Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE.; 2011. p. 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):1196–204. 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;35:375–80. 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;131:1174–9. 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, SG Abdel-Baki, 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, Dijk van JG, Cox FME. Automated seizure detection in an EMU setting: are software packages ready for implementation? Manuscript submitted for publication 2021.