Improved Neonatal Seizure Detection Using Adaptive Learning

A. H. Ansari [1], P. J. Cherian [2], A. Caicedo [1], M. De Vos [3], G. Naulaers [4], and S. Van Huffel [1]

Abstract— **In neonatal intensive care units performing continuous EEG monitoring, there is an unmet need for around-the-clock interpretation of EEG, especially for recognizing seizures. In recent years, a few automated seizure detection algorithms have been proposed. However, these are suboptimal in detecting brief-duration seizures (< 30s), which frequently occur in neonates with severe neurological problems. Recently, a multi-stage neonatal seizure detector, composed of a heuristic and a data-driven classifier was proposed by our group and showed improved detection of brief seizures. In the present work, we propose to add a third stage to the detector in order to use feedback of the Clinical Neurophysiologist and adaptively retune a threshold of the second stage to improve the performance of detection of brief seizures. As a result, the false alarm rate (FAR) of the brief seizure detections decreased by 50% and the positive predictive value (PPV) increased by 18%. At the same time, for all detections, the FAR decreased by 35% and PPV increased by 5% while the good detection rate remained unchanged.**

I. Introduction

Occurrence of neonatal seizures usually denotes an underlying serious neurological dysfunction. As the majority of these seizures have a subtle or no clinical correlate, their timely detection requires continuous electroencephalographic monitoring (cEEG). However, due to lack of expert support, there is a significant unmet gap in implementing cEEG in neonatal intensive care units (NICUs) [1]–[3]. Furthermore, occurrence of seizures may potentially aggravate brain injury, as has been shown in laboratory animal studies, although it is still debatable in human newborns [4]–[8]. Brain monitoring is usually performed in NICUs by using a compressed single channel EEG trend, known as amplitude integrated EEG (aEEG), because of its ease of use and interpretation. However, this has many limitations. Visual interpretation of multichannel cEEG by an experienced clinical neurophysiologist is the ideal method for detecting seizures. However, such expertise is often not available around the clock. Moreover, this method is expensive and labor-intensive [3], [9]. Therefore, an accurate automated neonatal seizure detector can be a valuable supportive tool for NICUs and it can potentially improve the quality and decrease the expenses of continuous monitoring of brain function.

In recent years, many publications have described different methods of automated neonatal seizure detection, varying from heuristic methods [10]–[14], data-driven classifiers [15]–[18], and their combinations [19]–[22]. One of the greatest challenges for these methods is the detection of brief-lasting seizures [12], [15], [23], [24] because typical characteristics of seizures, such as evolutionary patterns, are less clear-cut and at the same time they may be misinterpreted as artifacts [22]. Therefore, improved detection of these can boost the performance of an automated detection algorithm. The main goal of this paper is improving the overall performance of a previously developed multi-stage neonatal seizure detector particularly by improving the performance of the short seizure detections. To this end, after alarming for each detected seizure, the feedback from the expert is collected and used immediately to enhance the detector for subsequent detections. Thus the detector is constantly tuned for different channels and patients.

II. Methods

A. Dataset

All used data (training and test) in this work were recorded at the Sophia Children's Hospital (part of the Erasmus University Medical Center Rotterdam, The Netherlands) and all neonates had hypoxic ischemic encephalopathy (HIE). To develop and train the methods, 17 neonates are included, each of which has 27 hours on average of recorded EEG-polygraphy data, collected in a database, DB1, including electrocardiogram (ECG), electrooculogram (EOG), chin or limb surface electromyogram (EMG), and abdominal respiratory movement (Resp.). For testing the proposed method, we used 31 different neonates with the same measurement protocol from the same center. Among these, for 18 neonates, the scored seizures of the whole recordings were available (DB2) while for 13 neonates, only 2 hours in which at least one seizure appeared were scored by an expert and were available (DB3). In total, the database includes 461 hours of training and 516 hours of test data. The mean and median seizure duration of the test sets, including 1975 seizures, are

* This research is supported by Research Council KUL (BOF): CoE PFV/10/002 (OPTEC), SPARKLE IDO-13-0358, C24/15/036; VLAIO: projects: SWT 150466 - OSA+; iMinds Medical Information Technologies SBO 2016; Belgian Federal Science Policy Office: IUAP P7/19/ (DYSCO, 'Dynamical systems, control and optimization', 2012-2017); Belgian Foreign Affairs- Development Cooperation: VLIR UOS programs (2013-2019); EU: EU MC ITN TRANSACT 2012, #316679, European Research Council: ERC Advanced Grant, #339804 BIOTENSORS. AC is a postdoctoral research of the FWO and supported by HIP Trial (FP7/2007-2013) #260777

[1] A. H. Ansari, A. Caicedo, and S. Van Huffel are with the KU Leuven, Department of Electrical Engineering-ESAT, STADIUS, and imec, Leuven, Belgium (e-mail: amirhossein.ansari@kuleuven.be)

[2] P. J. Cherian, Section of Clinical Neurophysiology, Department of Neurology, Erasmus MC, University Medical Center, Rotterdam, The Netherlands and, Division of Neurology, Department of Medicine, McMaster University, Hamilton, Canada

[3] M. De Vos, Institute of Biomedical Engineering, Department of Engineering, University of Oxford, Oxford, UK

[4] G. Naulaers, Department of Development and Regeneration, University Hospitals Leuven, Neonatal Intensive Care Unit, KU Leuven, Leuven, Belgium

respectively 67 and 40 seconds. Before using the EEG signals, all the data were filtered between 1 and 20 Hz and then 20 bipolar EEG channels based on the full montage of the 10-20 International System explained in [25] are computed.

B. The multi-stage detector

The multi-stage algorithm is described in [22], in detail. Briefly, in the first stage, a heuristic algorithm mimicking a human expert is applied on the EEG signals in order to define some potential seizures. Afterwards, in the second stage, the EEG-polygraphic signal of each potential seizure is split into shorter epochs (8s, 50% overlap) and 50 features from each epoch are extracted. Then, a pre-trained support vector machine (SVM) is applied on the extracted features and a probability is assigned to each epoch. Finally, all probabilities of each potential seizure are compared with a predefined threshold (TH) and aggregated by a majority voting technique. In this way, some potential seizures primarily detected by the heuristic algorithm are removed as falsely detected seizures by the post-processor in the 2nd stage. As a result, the false alarm rate (FAR) significantly diminishes by 64% while the good detection rate (GDR) decreases only by 7%, when the TH is fixed to 0.3 based on used training data.

C. Proposed Method

The main idea of the proposed method is that in order to improve the performance of detecting short seizures, the sensitivity of the detector should vary according to the three following facts: 1) different patients may have different grades of background activity abnormality. As it is shown in [2], there is a relation between the background activity and characteristics of seizures, such as duration, repetition, rhythmicity, and evolution. The neonates with severe brain dysfunction have usually shorter and more repeated seizures. Therefore it is expected that the sensitivity of the detector should change from patient to patient. 2) Many short seizures are either focal, i.e., when a seizure activity appears only on one EEG electrode, or regional, i.e., when two or more adjacent electrodes display the seizure activity. Hence, only a limited number of channels need to be investigated for detecting seizures. In that case, using other irrelevant channels may inadvertently increase the false alarm rate. Thus, it is expected that the detector operates differently on different channels. 3) In some cases, especially for neonates with more severe brain injury and very short and repeatedly occurring seizures, the repetitiveness may disappear after treatment with anti-epileptic drugs (AEDs). Accordingly, it is expected that when a few seizures with robust characteristics are detected, the method becomes tuned to detect new seizures in that channel, even if these may not express all the characteristics of the previous seizures. Conversely, when there are no well-defined seizures or if some artifacts are wrongly detected in a channel, the detector needs to become less sensitive to the detection of dubious seizure-like patterns for that specific channel. To take these facts into account, the threshold (TH) of the second stage of the mentioned multi-stage seizure detector will be adaptively retuned in a third stage using the output of the second stage and feedback of an expert user. Every false/true detection will increase/decrease the TH of that channel and other channels in the neighborhood. This change applied on the TH will gradually be forgotten unless other similar feedback is regularly being received. The penalizing/ rewarding rate as well as the forgetting factor should be optimized based on the training dataset. Figure 1 shows a simple block diagram of the proposed multi-stage classifier. As it is clear in this diagram, after receiving feedback for each detected seizure by the post-processor, the new value of TH is recalculated based on the received feedback, the elapsed time from the last feedback, the channel in which the current seizure is detected, and the old value of the TH. The following steps thoroughly describe the proposed scenario when a potential seizure is detected on channel ch^* in the first stage:

1. The potential seizure is split into 8s epochs with 50% overlap.
2. The features proposed in [22] are extracted for each epoch.
3. A pre-trained SVM (with probabilistic output) is applied on the features and a probability is assigned to each epoch.
4. The thresholds of all available channels are primarily updated by eq. (1), derived from the transient and steady-state responses of resistor–capacitor circuits with an initial condition.

$$TH_{ch} \leftarrow TH^0 + (TH_{ch} - TH^0)e^{-\frac{\Delta t}{\tau}} \quad (1)$$

TH_{ch} is the sensitivity threshold of the updating channel. TH^0 is a predefined steady-state sensitivity threshold, Δt denotes the elapsed time (per second) from the previous update, and τ is a keeping factor (inverse of the forgetting factor) which shows how long a feedback will affect the threshold (the feedback will be used in step 6).

5. Then, all assigned probabilities of the current potential seizure are compared with the new TH_{ch^*}. If a majority of epochs has probabilities smaller than TH_{ch^*}, the potential seizure is labeled as false detection and removed. Otherwise, it is labeled as seizure and the detector alarms.
6. If a new seizure is detected, the real label (y: seizure, dubious, non-seizure) of that seizure scored by an expert is used to update the thresholds of all available channels by

$$TH_{ch} \leftarrow TH_{ch} - \frac{\gamma}{d(ch,ch^*)+1} R(y) \quad (2)$$

where γ is a pre-defined learning rate, R is a rewarding function defined by (3), and $d(ch, ch^*)$ is the graph-based distance between the updating channel (ch) and the detected channel (ch^*) on the montage graph.

$$R(y) = \begin{cases} +1 & if\ y = Seizure \\ 0 & if\ y = Dubious \\ -1 & if\ y = Nonseizure \end{cases} \quad (3)$$

As a consequence, if a seizure is detected on a channel and the expert user believes that the seizure has truly been detected, the threshold of that channel will decrease by γ, and subsequently, the thresholds of its neighbor channels with one common electrode decrease by $\gamma/2$, and likewise for other

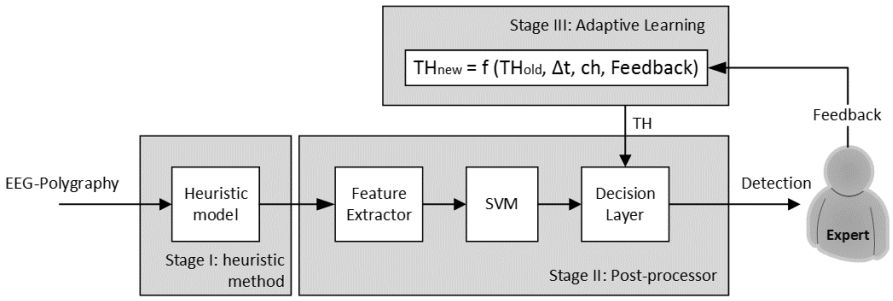

Figure 1. Diagram of the proposed technique. The TH used in the post-processor is adaptively returned by the proposed 3rd stage based on the current feedback, elapsed time from that last feedback, detected channel, and the old value of TH.

channels. Conversely, if the user scores the detected seizure as false alarm, the threshold increases by γ. This change ($\pm\gamma$) added to the threshold exponentially decays in time based on eq. (1) so that if no alarm is found or no feedback received for 3τ seconds, the threshold will become almost equal to TH^0. If the seizure is scored as dubious, it has no effect on the threshold.

III. Results and Discussion

In order to compare the performance of the proposed method with the original multi-stage seizure detector (II. A), the good detection rate (GDR %), the false alarm rate (FAR), and the positive predictive value (PPV %) are measured by (4-6).

$$GDR = \frac{TP}{TP+FN} \times 100 \quad (4)$$

$$FAR = \frac{FP}{total\ time\ (in\ hour)} \quad (5)$$

$$PPV = \frac{TP}{TP+FP} \times 100 \quad (6)$$

In these equations, TP, FP, and FN are respectively the number of truly detected, falsely detected, and missed seizures.

In this work, the used learning rate (γ) and keeping factor (τ) were optimized by a leave-'one patient'-out (LOPO) cross -validation method applied on the training dataset and equal to 0.1 and 10min respectively.

Figure 2 shows the performance of the proposed method as well as the original method with fixed threshold for the test DBs. As it is clear in this figure, the overall area under the ROC increased by the proposed method. In the reference paper of the original method [22], it was shown that TH=0.3 is a good trade-off between the GDR and the FAR based on the training data. The same threshold has been used here as TH^0 (squares and circles in the figure). As a result, the overall FAR (for DB2&3) decreased by 35% while the GDR did not decrease. At this point, the PPV also increased by 5%.

Figure 3 displays the histogram of the detected seizures, the number of false alarms, and the PPVs as a function of seizure duration for both methods and both DBs together. Almost all seizures longer than 1min were detected by both methods (upper chart), while the short seizures are still not entirely detected. Nevertheless, the number of false detections of very short seizures ($< 30s$) decreased by 50% which is the main contribution of this work. The PPV of very short seizures also increased from 41% to 59% by the proposed method which shows higher reliability of the alarms.

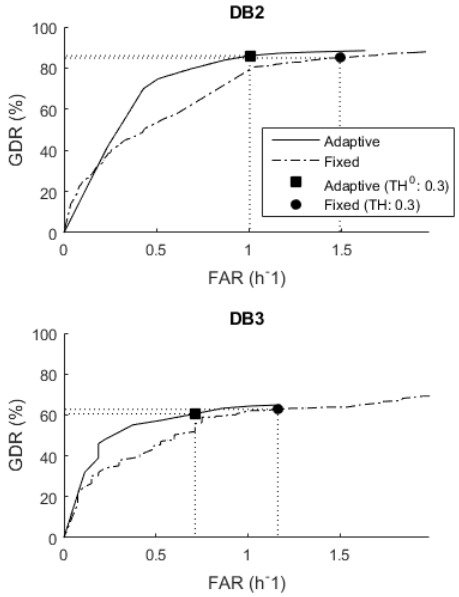

Figure 2. The ROC curve of the seizure detector with fixed threshold (dashed line) as well as the proposed adaptive threshold (continuous line) for DB2 (top) and DB3 (bottom).

The other advantage of this method is that since the feedback of the expert user is continuously used to retune the threshold, the method can tailor to the needs of the user. In other words, if an expert user prefers to detect more seizures including the dubious and definite ones or to detect only definite seizures, the method can adaptively learn these preferences. Consequently, it is possible that this method would have higher overall performance in a multi-rater scored database compared to other kinds of fixed detectors. But, this needs to be tested in a future work.

The main drawback of this technique is the need for collecting feedback from the expert users in a real NICU environment. A major problem is the limited EEG expertise among front-line carers in the NICUs, the nurses and neonatologists. These primary caregivers generally call a neurologist or clinical neurophysiologist for consultation when they see repetitive patterns that frequently occur or are

considering initiating treatment with anti-seizure medications. In this research it is presupposed that after each alarm, the feedback of an expert EEG reader is available. In future research, we will address the practical limitations, such as the limited knowledge of primary caregivers, a delay between an alarm and its feedback, and limited number of available feedback for each patient.

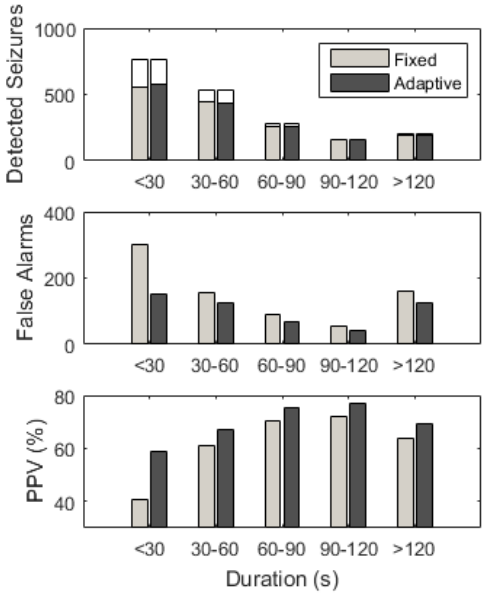

Figure 3. Histogram of the detected seizures (top), the number of false alarms (middle), and the PPVs (bottom), for the original method with fixed threshold (light gray) and the proposed adaptive learning method (dark gray). In the upper chart, the white stacked bars show the number of missed seizures.

IV. Conclusion

In this work, an adaptive learning method was proposed to decrease the false alarm rate and subsequently increase the positive predictive value of a previously developed multi-stage neonatal seizure detector particularly for very short seizures and false alarms. This method effectively improves the performance such that the FAR decreases by 35% and PPV increases by 5% on average, while the good detection rate remains unchanged. Moreover, for very short events, which are very challenging problems in automated neonatal seizure detection, the FAR decreases by 50% and PPV increases by 18%.

References

[1] J. J. Volpe, *Neurology of the Newborn*, 5th ed. Philadelphia: Saunder WB, 2008.

[2] J. C. Perumpillichira, *Improvements in Neonatal Brain Monitoring after Perinatal Asphyxia*. 2010.

[3] J. M. Rennie, G. Chorley, G. B. Boylan, R. Pressler, Y. Nguyen, and R. Hooper, "Non-expert use of the cerebral function monitor for neonatal seizure detection," *Arch. Dis. Child.-Fetal Neonatal Ed.*, vol. 89, no. 1, pp. F37–F40, 2004.

[4] H. C. Glass, D. Glidden, R. J. Jeremy, A. J. Barkovich, D. M. Ferriero, and S. P. Miller, "Clinical Neonatal Seizures are Independently Associated with Outcome in Infants at Risk for Hypoxic-Ischemic Brain Injury," *J. Pediatr.*, vol. 155, no. 3, pp. 318–323, Sep. 2009.

[5] E. C. Wirrell, E. A. Armstrong, L. D. Osman, and J. Y. Yager, "Prolonged seizures exacerbate perinatal hypoxic-ischemic brain damage," *Pediatr. Res.*, vol. 50, no. 4, pp. 445–454, Oct. 2001.

[6] J. Y. Yager, E. A. Armstrong, H. Miyashita, and E. C. Wirrell, "Prolonged neonatal seizures exacerbate hypoxic-ischemic brain damage: correlation with cerebral energy metabolism and excitatory amino acid release," *Dev. Neurosci.*, vol. 24, no. 5, pp. 367–381, 2002.

[7] S. Mitra *et al.*, "Changes in Cerebral Oxidative Metabolism during Neonatal Seizures Following Hypoxic–Ischemic Brain Injury," *Front. Pediatr.*, vol. 4, Aug. 2016.

[8] G. L. Holmes, "Seizure-induced neuronal injury Animal data," *Neurology*, vol. 59, no. 9 suppl 5, pp. S3–S6, Nov. 2002.

[9] P. J. Cherian *et al.*, "Validation of a new automated neonatal seizure detection system: A clinician's perspective," *Clin. Neurophysiol.*, vol. 122, no. 8, pp. 1490–1499, Aug. 2011.

[10] W. Deburchgraeve *et al.*, "Automated neonatal seizure detection mimicking a human observer reading EEG," *Clin. Neurophysiol.*, vol. 119, no. 11, pp. 2447–2454, Nov. 2008.

[11] M. De Vos *et al.*, "Automated artifact removal as preprocessing refines neonatal seizure detection," *Clin. Neurophysiol.*, vol. 122, no. 12, pp. 2345–2354, Dec. 2011.

[12] S. B. Nagaraj, N. J. Stevenson, W. P. Marnane, G. B. Boylan, and G. Lightbody, "Neonatal Seizure Detection Using Atomic Decomposition With a Novel Dictionary," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 11, pp. 2724–2732, Nov. 2014.

[13] N. J. Stevenson, J. M. O'Toole, L. J. Rankine, G. B. Boylan, and B. Boashash, "A nonparametric feature for neonatal EEG seizure detection based on a representation of pseudo-periodicity," *Med. Eng. Phys.*, vol. 34, no. 4, pp. 437–446, May 2012.

[14] P. Celka and P. Colditz, "A computer-aided detection of EEG seizures in infants: a singular-spectrum approach and performance comparison," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 5, pp. 455–462, May 2002.

[15] A. Temko, E. Thomas, W. Marnane, G. Lightbody, and G. Boylan, "EEG-based neonatal seizure detection with Support Vector Machines," *Clin. Neurophysiol.*, vol. 122, no. 3, pp. 464–473, Mar. 2011.

[16] H. Hassanpour, M. Mesbah, and B. Boashash, "Time–frequency based newborn EEG seizure detection using low and high frequency signatures," *Physiol. Meas.*, vol. 25, no. 4, p. 935, Aug. 2004.

[17] J. G. Bogaarts, E. D. Gommer, D. M. W. Hilkman, V. H. J. M. van Kranen-Mastenbroek, and J. P. H. Reulen, "EEG Feature Pre-processing for Neonatal Epileptic Seizure Detection," *Ann. Biomed. Eng.*, vol. 42, no. 11, pp. 2360–2368, Aug. 2014.

[18] A. Zwanenburg *et al.*, "Using trend templates in a neonatal seizure algorithm improves detection of short seizures in a foetal ovine model," *Physiol. Meas.*, vol. 36, no. 3, p. 369, 2015.

[19] J. Mitra *et al.*, "A Multi-stage System for the Automated Detection of Epileptic Seizures in Neonatal EEG," *J. Clin. Neurophysiol. Off. Publ. Am. Electroencephalogr. Soc.*, vol. 26, no. 4, pp. 218–226, Aug. 2009.

[20] A. Aarabi, R. Grebe, and F. Wallois, "A multistage knowledge-based system for EEG seizure detection in newborn infants," *Clin. Neurophysiol.*, vol. 118, no. 12, pp. 2781–2797, Dec. 2007.

[21] A. H. Ansari, V. Matic, M. De Vos, G. Naulaers, P. J. Cherian, and S. Van Huffel, "Improvement of an automated neonatal seizure detector using a post-processing technique," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 5859–5862.

[22] A. H. Ansari *et al.*, "Improved multi-stage neonatal seizure detection using a heuristic classifier and a data-driven post-processor," *Clin. Neurophysiol.*, vol. 127, no. 9, pp. 3014–3024, Sep. 2016.

[23] S. R. Mathieson *et al.*, "Validation of an automated seizure detection algorithm for term neonates," *Clin. Neurophysiol.*, vol. 127, no. 1, pp. 156–168, Jan. 2016.

[24] L. S. Smit, R. J. Vermeulen, W. P. F. Fetter, R. L. M. Strijers, and C. J. Stam, "Neonatal Seizure Monitoring Using Non-Linear EEG Analysis," *Neuropediatrics*, vol. 35, no. 6, pp. 329–335, Nov. 2004.

[25] P. J. Cherian, R. M. Swarte, and G. H. Visser, "Technical standards for recording and interpretation of neonatal electroencephalogram in clinical practice," *Ann. Indian Acad. Neurol.*, vol. 12, no. 1, pp. 58–70, 2009.