34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Temporal Evolution of Seizure Burden for Automated Neonatal EEG Classification

Andriy Temko, Nathan Stevenson, William Marnane, Geraldine Boylan, Gordon Lightbody

Abstract— **The aim of this paper is to use recent advances in the clinical understanding of the temporal evolution of seizure burden in neonates with hypoxic ischemic encephalopathy to improve the performance of automated detection algorithms. Probabilistic weights are designed from temporal locations of neonatal seizure events relative to time of birth. These weights are obtained by fitting a skew-normal distribution to the temporal seizure density and introduced into the probabilistic framework of the previously developed neonatal seizure detector. The results are validated on the largest available clinical dataset, comprising 816.7 hours. By exploiting these priors, the ROC area is increased by 23% (relative) reaching 96.75%. The number of false detections per hour is decreased from 0.72 to 0.36, while maintaining the correct detection of seizure burden at 75%.**

I. Introduction

HYPOXIC ischemic encephalopathy (HIE) is the most common cause of seizures in the sick full term neonate. The incidence of neonatal seizures is generally reported as around as between 1-3 per 1000 but may be much higher in very preterm babies [1]. In reality, these values are probably inaccurate estimates as only about one third of all neonatal seizures are clinically visible and many remain undetected in the busy Neonatal Intensive Care Unit (NICU) [2]. Failure to detect seizures and the resulting lack of treatment may result in brain damage and in severe cases, death. A system that could automatically detect and annotate seizures on the neonatal EEG would be extremely useful for clinicians in the NICU. Although a number of methods and algorithms have been proposed previously in an attempt to automatically detect neonatal seizures [3] – [7], to date their transition to clinical use has been limited due to poor performance. Navakatikyan et al. [5] reported that their system correctly detected 82.8% of the seizure burden (the total amount of time the newborn spends in seizure) at a cost of 2 false detections per hour. A recent study by Cherian et al. [6] reported the correct detection of on average 59% of seizure burden at a cost of 0.58 false detections per hour (FD/h). With the exclusion of the four most difficult and worst performing patients, the number of FD/h was shown to be reduced from 0.58 to 0.28.

There are two key directions in automated neonatal seizure detection. The first follows analytical learning principles [8] and focuses on the creation of a set of heuristic rules and thresholds from clinical prior knowledge [3] – [6]. The resultant detectors analyze EEG using a small number of the descriptors from which a decision is made using empirically derived thresholds. The second approach relies on inductive learning [8] and utilizes statistical classifier based methods [7], [9], which employ elements of machine learning to classify a set of features using a data-driven decision rule.

It is known that good solutions to most practical learning problems result from a combination of these two approaches. Unlike analytical rules and thresholds, where binary decisions are obtained, the classifier-based approach often outputs continuous (probabilistic) values and thus provides confidence or credibility to the decisions made. While working in the classifier domain, prior information can easily be introduced using the existing well-defined probabilistic framework.

Domain prior knowledge on neonatal seizures can come in different ways. It can come in terms of statistics of neonatal seizure spatial locations or estimated patient-specific history of previous seizure spatial locations [10]. In the current study, knowledge about the temporal evolution of seizures in neonates with HIE is used to temporally weight the output of a seizure detection algorithm. In particular, the previous work of our group [11] has shown that the distribution of neonatal seizures in time is not uniform. In neurologically compromised neonates, seizures are more or less likely to happen in a certain period after birth. This information is converted in our work into probabilistic weights and integrated into a previously designed system of neonatal seizure detection.

II. Neonatal Seizure Detectors

A. Dataset

The dataset in our work is composed of EEG recordings from 18 newborns recruited from the NICU, Cork University Maternity Hospital, Cork, Ireland. The patients were full term babies ranging in gestational age from 39 to 42 weeks. All newborns had seizures secondary to hypoxic ischemic encephalopathy (HIE). A Carefusion NicOne video EEG monitor was used to record multi-channel EEG at

Manuscript received 15/05/2012. This work was supported in part by the SFI (10/IN.1/B3036) and Welcome Trust (085249/Z/08/Z).

Andriy Temko, William Marnane, and Gordon Lightbody are with the Department of Electrical and Electronic Engineering and the Neonatal Brain Research Group, University College Cork, Ireland. {andreyt}@eleceng.ucc.ie, {l.marnane, g.lightbody}@ucc.ie

Nathan Stevenson and Geraldine Boylan are with the Department of Pediatrics and Child Health and the Neonatal Brain Research Group, University College Cork, Ireland. {n.stevenson, g.boylan}@ucc.ie

256Hz using the 10-20 system of electrode placement, modified for neonates. The standard protocol for EEG recording in the NICU required the following 9 active electrodes: T4, T3, O1, O2, F4, F3, C4, C3, and Cz. Then, the following 8 EEG bipolar pairs were used to annotate the data: F4-C4, C4-O2, F3-C3, C3-O1, T4-C4, C4-Cz, Cz-C3 and C3 - T3. All electrographic seizures were annotated independently by two experienced neonatal electro-encephalographers using simultaneous video EEG. The combined length of the EEG recordings totaled 816.7 hours (median per patient, 48.5h) and contained 1389 electrographic seizures (median per patient, 53 seizures). The dataset contains a wide variety of seizure types including both electrographic-only and electro-clinical seizures of focal, multi-focal and generalized types. The continuous EEG recordings were not edited to remove the large variety of artifacts and poorly conditioned signals that are commonly encountered in the real-world NICU environment.

B. *Automated seizure detection system architecture*

The neonatal seizure detection system is shown in Fig. 1. The EEG from the 8 above-mentioned channels was down-sampled from 256Hz to 32Hz with an anti-aliasing filter set at 12.8Hz. The EEG was then split into 8s epochs with 50% overlap between epochs. Fifty-five features were extracted from each channel which represent both time and frequency domain characteristics as well as information theory based parameters. Various system architecture choices (such as epoch length, epoch shift, features, etc) are detailed in [9].

Neonatal seizures can be localized to a single EEG channel; thus per-channel annotations are needed to train a classifier. For the seizure class, the training dataset consists of approximately 20 minutes of EEG per patient for which individual channel annotations are available, which sum up to M*20 minutes per patient for seizures involved in M channels. For example, if a training dataset consists of 17 patients for which 20 minutes of seizure are transcribed on the per channel basis and on average 4 channels are involved in every seizure, then for an epoch length of 8 seconds with an overlap of 4 seconds, the seizure class of training data will consist of 17patients*(1200s/4s) *4channels = 19200 epochs. It may be more or less depending on the number of channels involved in every seizure for every patient. These were used to represent a seizure class, while 40000 epochs were randomly selected from the non-seizure data for representation of the non-seizure class. The training data for the classifier were normalized anisotropically by subtracting the mean and dividing by standard deviation to assure commensurability of the various features. This normalizing template was then applied to the testing data.

The normalized features extracted from each epoch were then fed to a support vector machine (SVM) classifier with a Gaussian kernel. Nested cross-validation model selection on the training data was performed to choose suitable model parameters. The outputs of the SVM were converted to probability-like values [12] and smoothed with a moving average filter. The maximum of the averaged probabilities across all channels was computed to represent the final support of a seizure. It was then compared to a threshold from the interval [0 1]. After comparison, a binary decision was taken: 1 for seizure and 0 for non-seizure. The 'collar' technique was applied last – every seizure decision was extended from either side to account for the delay introduced by the moving average filter. The system emits a continuous pseudo-probabilistic output which allows for selection of a desired operating point depending on clinical needs. In [13], the system was shown to outperform the existing alternatives using various metrics and the standardized performance assessment.

C. *Performance assessment and metrics*

In clinical practice, samples of testing patient data are never available beforehand in the NICU. It is therefore necessary to develop patient-independent neonatal seizure detector. For this reason, the leave-one-out (LOO) cross-validation method was used to assess the performance of the system for patient-independent seizure detection [9]. This way, all but one patients' data from the dataset were used for training and development and the remaining seizure patient's data was used for testing. This procedure was repeated until each seizure patient had been a test subject and the mean result was reported.

The metrics used in this work are epoch-based sensitivity and specificity values which are defined as the epoch-wise accuracy of each class (seizure and non-seizure) separately. Sensitivity corresponds to detected seizure burden, i.e. the total amount of time the baby spends in seizure. Seizure burden is the most important metric to indicate whether a patient should be treated or not [2]. We also report the number of FD/h. By thresholding the probability of seizure (in the range from 0 to 1), it is possible to report the curves of performance in contrast to reporting a performance for a single operating point. The area under Receiver Operating Characteristic (ROC) curve, which plots sensitivity over

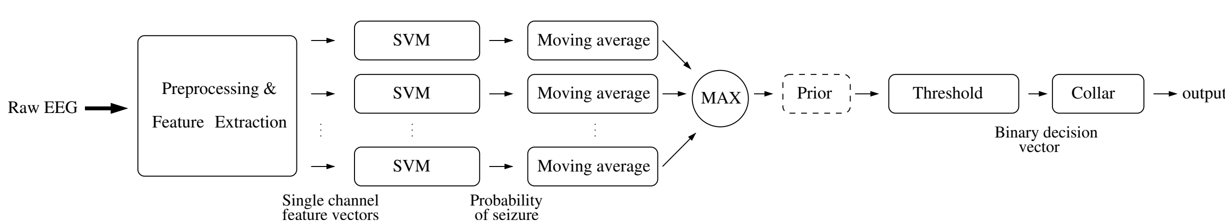

Fig. 1. Neonatal seizure detection system diagram

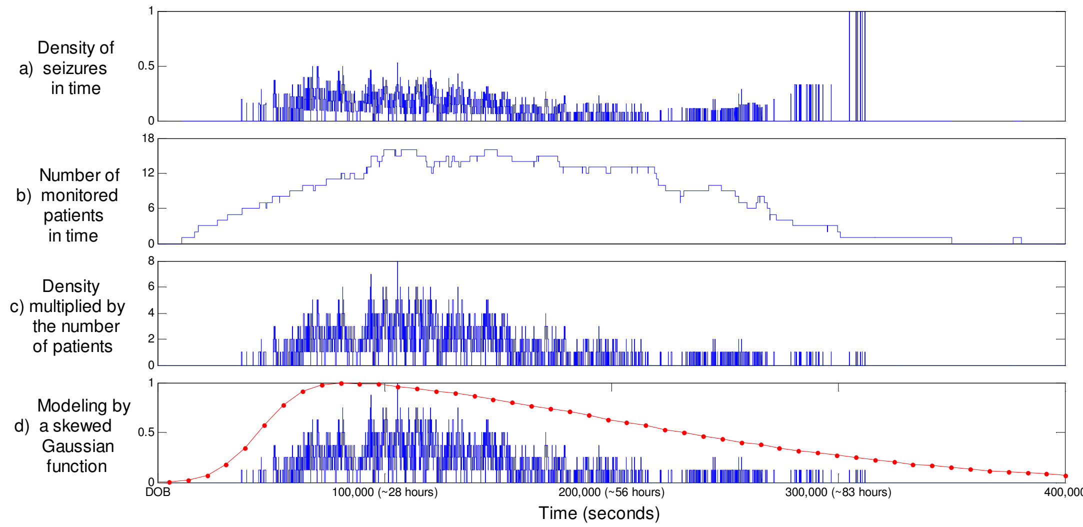

Fig. 2. Modeling the clinical prior. The x-axis indicates time elapsed after date of birth (DOB). a) Density of seizures. b) The number of monitored patients. c) Density multiplied by the number of monitored patients. d) A skew-normal distribution fitting the normalized data-driven clinical priors.

specificity values, is used in this work.

III. Clinical Prior

The process of modeling the clinical priors is shown in Fig. 2. First, the temporal density of seizures was calculated from the dataset as shown in Fig. 2 (a). The x-axis indicates time elapsed after date of birth (DOB). For example, it can be seen from Fig. 2 (a) that at approximately 30 hours after birth around 40% of monitored patients had seizures. Naturally, this data-driven temporal seizure density measure needs to be normalized by some credibility function. Fig. 2 (b) plots the number of patients monitored versus time. From Fig. 2 (b), it can now be explained that the unit density in Fig. 2 (a) resulted from the fact that only a single patient was monitored at that time. In effect, Fig. 2 (b) is used in our work to weight the data-driven seizure density. It can be also seen from Fig. 2 (b) that there is not a single point in time, relative to the time of birth, where all 18 patients from the dataset were monitored – at most, 16 patients were simultaneously monitored (e.g. ~30h after birth). Fig. 2 (c) shows the density from plot (a) multiplied by the credibility function from plot (b). It can be seen from Fig. 2 (c) that the resultant measure follows a skewed normal distribution with a long tail. It was decided here to approximate the normalized final measure by a skew-normal distribution. The implementation proposed in [14] is used in our work. Defining $\upsilon = \dfrac{x-\xi}{\omega}$, the probability density function of the skew-normal distribution parameterized by location ξ, scale ω, and shape α is given as:

$$f(x) = \frac{2}{\omega}\phi(\upsilon)\Phi(\alpha\upsilon) \quad (4)$$

$$\phi(\upsilon) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\upsilon^2}{2}\right), \quad (5)$$

$$\Phi(\alpha\upsilon) = \int_{-\infty}^{\alpha\upsilon}\phi(s)ds \quad (6)$$

When shape parameter α=0, the skewness vanishes, and the standard normal density is obtained. As α increases (in absolute value), the skewness of the distribution increases. The sign of α defines whether the distribution is left or right-skewed. It has been shown in [14] that there is no closed-form expression available for maximum likelihood estimates of the distribution parameters (ξ, ω, and α). For this reason, the parameters are manually greedy-searched to get a reasonable fit over the data. In our work, ξ=0.4, ω=1.5, and α=19 were used (scaled by 10^5 to fit the actual time axis in seconds). The resultant function, normalized to be between 0 and 1, is shown in Fig. 2 (d) in red.

The modeled priors are used to weight the probability of a seizure event. In particular, as shown in Fig. 1, the maximum probability of seizure computed across channels is multiplied by the corresponding temporal weight.

IV. Results And Discussion

A. ROC area

The absolute differences in per-patient performance with and without temporal priors are shown in Fig. 3. It can be seen from Fig. 3 that the separability of seizure and non-seizure probabilistic activity for most patients increases when weighted with the designed prior function. The *p*-values of the two-tailed statistical significance test of the difference between the two ROC areas [15] indicate that only for patients 9 and 16 the proposed priors have no statistically significant effects. For the remaining patients, the corresponding *p*-values are close to 0.

The overall positive effect of the proposed temporal weighting of probabilities can be seen through the average ROC area across all patients which has been increased from 95.77 to 96.75; this corresponds to a ~23% relative improvement (96.75-95.77)/(100-95.77).

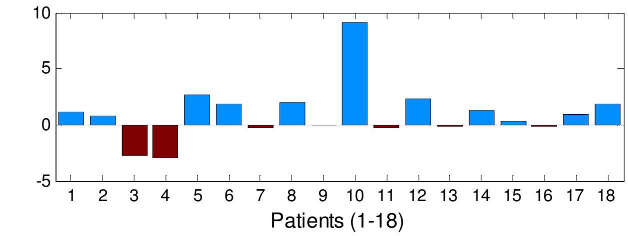

Fig. 3. Absolute per-patient differences in ROC areas for the neonatal seizure detector obtained with and without using designed clinical priors.

It is worth reemphasizing that unlike other studies which report performance increases obtained on datasets of several carefully selected minutes of EEG [4], the results in our study are obtained on the largest available dataset, which comprises 816 hours of continuous unedited neonatal EEG, and thus these results are stable and significant.

B. *Seizure burden and FD/h*

The contribution of the introduced weights in terms of the clinically important metric is shown in Fig. 4. It can be seen that the number of false detection per hour is consistently lower when exploiting the designed weights for all operating points. In particular, the number of FD/h can be reduced from 0.72 to 0.36 while maintaining the correct detection of seizure burden as high as 75%.

The provided performance curve not only enables the comparison of our system with alternatives but also facilitates the comparison among the alternatives which report the same metrics. It can be seen from Fig. 4 that results reported in [5] and [6] could effectively be thought of as belonging to the same curve which is almost equally distanced from the curve of our results. It can be concluded that the claimed reduction in the number of FD/h in [6] comes therefore at a commensurable cost of reduced detected seizure burden; although the works in [5 and [6] are separated by 5 years of active research.

C. *Future work*

A neurological assessment can be performed soon after birth in the neonate and the severity of HIE can be graded. Different prior functions can be introduced for each HIE grade which will allow more accurate modeling of the temporal evolution of the seizure burden. It is also known that patient cooling procedures affect seizures in neonates with HIE. It is therefore necessary to re-estimate the reported priors on a cohort of newborns who are cooled to properly reflect the differences in seizure burden distributions.

V. Conclusions

A significant improvement in the performance of a patient independent neonatal seizure detector was achieved by the inclusion of a temporal prior weighting. This temporal weighting function was designed from the statistics of seizure location distributions relative to time of birth. The statistics are calculated from the largest available datasets of

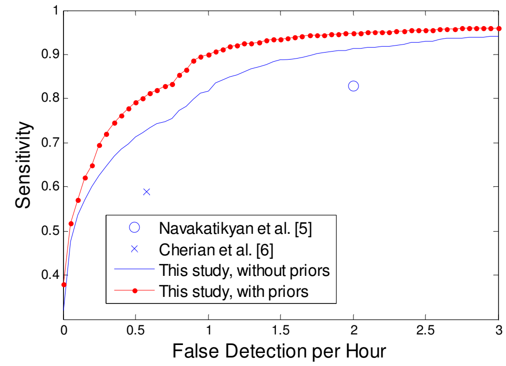

Fig. 4. Sensitivity vs. number of false detection per hour.

neonatal seizures. It was shown that the inclusion of the reported weighting function in the existing probabilistic framework results in the significant increase of the seizure detection performance measured by both epoch-based and event-based metrics. The designed priors can be exploited in existing neonatal seizure detectors.

References

[1] J. Rennie, G. Boylan, "Treatment of neonatal seizures", *Arch Dis Child Fetal Neonatal Ed*, v. 92, pp. 148-50, 2007.

[2] D. Murray, G. Boylan, I. Ali, C. Ryan, B. Murphy, S. Connolly, "Defining the gap between electrographic seizure burden, clinical expression and staff recognition of neonatal seizures", *Arch Dis Child Fetal Neonatal Ed*, v. 93, pp.187-91, 2008.

[3] J. Mitra, J. Glover, P. Ktonas, A. Kumar, A. Mukherjee, N. Karayiannis, J. Frost, R. Hrachovy, E. Mizrahi, "A Multistage System for the Automated Detection of Epileptic Seizures in Neonatal Electroencephalography", *J Clin Neurophysiol*, v. 26, pp. 1-9, 2009.

[4] P. Celka, P. Colditz, "A computer-aided detection of EEG seizures in infants, a singular-spectrum approach and performance comparison", *IEEE Tran Biomed Eng*, v. 49, pp. 455–462, 2002.

[5] M. Navakatikyan, P. Colditz, C. Burke, T. Inder, J. Richmond, C. Williams, "Seizure detection algorithm for neonates based on wave sequence analysis", *Clin Neurophysiol*, v. 117, pp. 1190-1203, 2006.

[6] P. Cherian, W. Deburchgraeve, R. Swarte, M. De Vos, P. Govaert, S. Van Huffel, G. Visser, "Validation of a new automated neonatal seizure detection system: A clinician's perspective", *Clin Neurophysiol*, v. 122(8), pp. 1490-1499, 2011.

[7] A. Aarabi, R. Grebe, and F. Wallois, "A multistage knowledge-based system for EEG seizure detection in newborn infants", *Clin Neurophysiol*, v. 118, pp. 2781-97, 2007.

[8] T. Mitchell, *Machine Learning*, McGraw Hill, 1997.

[9] A. Temko, E. Thomas, W. Marnane, G. Lightbody, G. Boylan, "EEG-based neonatal seizure detection with support vector machines", *Clin Neurophysiol*, v. 122, i. 3, pp. 464-473, 2011.

[10] A. Temko, G. Lightbody, E. Thomas, G. Boylan, W. Marnane, "Instantaneous Measure of EEG Channel Importance for Improved Patient-Adaptive Neonatal Seizure Detection", *IEEE Trans Biomed Eng*, to appear.

[11] N .Lynch, N. Stevenson, V. Livingstone, J. Rennie, G. Boylan, "The temporal evolution of electrographic seizure burden in neonatal hypoxic ischemic encephalopathy", *Epilepsia*, Wiley, to appear.

[12] J. Platt, "Probabilistic outputs for SVM and comparison to Regularized likelihood methods", *Advances in Large Margin Classifiers*, 1999.

[13] A. Temko, E. Thomas, W. Marnane, G. Lightbody, G. Boylan, "Performance assessment for EEG-based neonatal seizure detectors", *Clin Neurophysiol*, v. 122, i. 3, pp. 474-482, 2011.

[14] A. Azzalini, "A class of distributions which includes the normal ones", *Scand J Statist*, v. 12, pp. 171–178, 1985.

[15] J. Hanley, B. McNeil, "A method of comparing the areas under receiver operating characteristic curves derived from the same cases", Radiology, v. 148, i. 3, pp. 839-843, 1983.