Detecting Neonatal Seizures using Short Time Fourier Transform and Frechet Distance

Aleksandar Jeremic[1] and Dejan Nikolic[2]
[1]*Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada*
[2]*Physical Medicine and Rehabilitation, University Childrens Hospital, Faculty of Medicine,*

Keywords: Seizure Detection, Information Fusion, Machine Learning.

Abstract: Recently there has been an increase in the number of long-term cot-bed EEG systems being implemented in clinical practice in order to monitor neurological development of neonatal patients. Consequently a significant research effort has been made in the development of automatic EEG data analysis tools including but not limited to seizure detection as seizure frequency and/or intensity are one of the most important indicators of brain development. In this paper we propose to evaluate time dependent power spectral density using short time Fourier transform and using Frechet distance measure to detect presence and/or absence of seizures. We propose to use three different distance measures as they capture different properties of the corresponding PSD matrices. We evaluate the performance of the proposed algorithms using real data set obtained in the NICU of the McMaster University Hospital. In order to benchmark performance of our proposed techniques we trained and tested a support vector machine (SVM) classifier.

1 INTRODUCTION

Continuous EEG monitoring and analysis remain important clinical tools in neonatal intensive care units (NICU) for early stage evaluation / detection / diagnosis of various types of encephalopathies. In the recent decade there has been a significant advancement in utilizing advanced EEG techniques for improving outcomes for neonatal patients experiencing various degrees of neurological developmental issues.(Faul, 2005). To this purpose long-term EEG monitoring is being applied to cot-beds for a wider spectrum of patients and not only to those with severe neurological problems (Temko et al., 2015). Consequently the amount of data being generated by such systems cannot be reviewed by experts due to the limited resources (number of personnel, time, etc.) One of the most important and critical emergencies phenomenons that is being monitored in NICUs is occurrence of seizures as it allows better understanding of brain function in a variety of patients, from the extremely premature newborn to the term baby with acute injury. In addition, neonatal EEG facilitates rapid diagnosis of seizures, identification of epileptic encephalopathy, and may provides useful prognostic information.

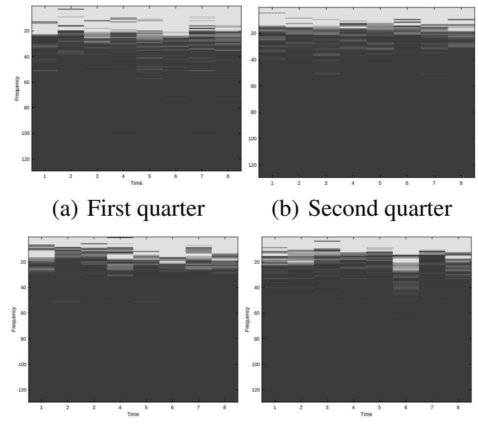
A seizure is defined clinically as a paroxysmal alteration in neurologic function, i.e., behavioural, motor, or autonomic function. It is a result of excessive electrical discharges of neurones, which usually develop synchronously and happen suddenly in the central nervous system (CNS). It is critical to recognize seizures in newborns, since they are usually related to other significant illnesses. Seizures are also an initial sign of neurological disease and a potential cause of brain injury (Volpe, 2001). In a clinical settings physicians are able to detect seizures based on EEG data however the process may be time consuming considering the number of cot-beds in regular size NICU department. To this purpose development of computer-aided diagnosis would be extremely beneficial as such system would be important from both academic and clinical standpoint of view. From the academic stand point automatic recording of seizures and consequently analysis of these data would provide insight into frequency of occurrence and correlate it with the dynamic of neurological development. From clinical standpoint it has been demonstrated that the neonatal patient outcomes can be improved due to early detection of certain encephalopathies.

In the last decade a significant number of short-time Fourier transform and wavelet transform tech-

Jeremic, A. and Nikolic, D.
Detecting Neonatal Seizures using Short Time Fourier Transform and Frechet Distance.
DOI: 10.5220/0009178703420347
In *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2020) - Volume 4: BIOSIGNALS*, pages 342-347
ISBN: 978-989-758-398-8; ISSN: 2184-4305

niques have been proposed. Wavelet transform techniques commonly attempt to decompose the signal into the frequency bands of interest and thus achieve discriminatory goals. On the other hand, STFT attempts to analyze non-stationary properties by analyzing various features contained in the time-dependent power spectral density. To this purpose several machine learning algorithms have been implemented but their performance relies heavily on significant amount of data which may not be available since the patient-to-patient variability of seizure features is significant. In our previous work, we proposed two distributed detection algorithms for neonatal seizure detection using some of the commonly used single channel seizure detection algorithms and extended this approach to the detection of seizures using Frechet distance measure of sample EEG covariances. In this paper we develop a distributed detection algorithm using a single-channel detection algorithm based on short-time Fourier transform (STFT) and three Frechet distance based detectors between the training ensemble of STFT matrices and actual data.

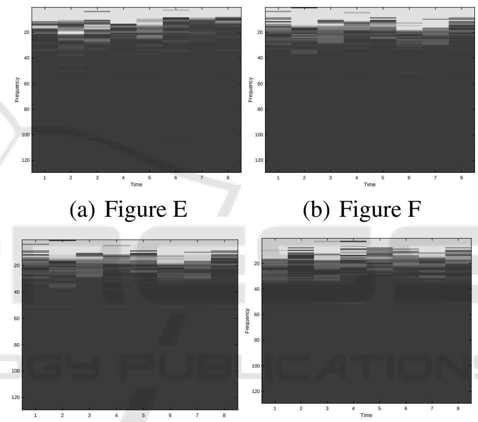
First, we present an estimator of the Frechet mean of the power spectral density (PSD) matrix on the manifold $\mathcal{M}$ using the different measures of Riemannian distances. Then we introduce the Fréchet mean based on two Riemannian distances and discuss computational algorithms for calculating the proposed distance means. In Section 3 we illustrate applicability of our results using data set of NICU patients. Finally, in Section 4 we discuss future directions.



(a) First quarter (b) Second quarter

(c) Third quarter (d) Fourth quarter

Figure 1: PSD matrices in the presence of seizures - 1s epoch with 25% overlap.



(a) Figure E (b) Figure F

(c) Figure G (d) Figure H

Figure 2: PSD matrices in the absence of seizures - 1s epoch.

2 SIGNAL MODEL

2.1 Short-time Fourier Transform

Let $x(t)$ denote the uniformly sampled EEG signal, then the discrete STFT can be written as

$$\mathcal{F}\{F(n,k)\} = \sum_{m=0}^{M-1} f(n-m)w(m)e^{-2\pi mk/N} \quad (1)$$

where $f(n)$ is the EEG signal, $w(m)$ is a supporting window, M is the window length, and N is the number of samples used for calculating STFT. Therefore this algorithm can be viewed as a continuous calculation of STFT using a sliding windows and hence contains in itself temporal variation of the EEG frequency spectrum. The visual representation of matrix F is commonly referred to as a spectrogram of the corresponding signal. In order to apply the Frechet distance measure to the ensemble of the corresponding STFT spectrograms F_i we calculate the corresponding power spectral density matrix given by

$$S_i = F_i F_i^H \quad (2)$$

where superscript H denotes Hermitian transpose due to the fact that the entries of F are complex numbers. In Figures 1 we illustrate sliding window PSD output of STFT for an arbitrary seizure epoch and in Figure 2 we illustrate similar results fin the absence of seizures.

2.2 Frechet Distance

To measure the distance between two $M \times M$ covariance matrices $\mathbf{A}$ and $\mathbf{B}$ on manifold of positive definite matrices $\mathcal{M}$, we consider the metrics which have been developed to measure distance between two points on the manifold itself.

The first metric is obtained by measuring distance between projections on the subspace spanned by unitary matrices (Li and Wong, 2013)

$$d_{R_1}(\mathbf{A},\mathbf{B}) = \sqrt{\text{Trace}(\mathbf{A}) + \text{Trace}(\mathbf{B}) - 2\text{Trace}(\mathbf{A}^{\frac{1}{2}}\mathbf{B}\mathbf{A}^{\frac{1}{2}})} \tag{3}$$

The second metric is obtained by measuring the distance between their projections on the subspace spanned by identity matrices. It has been shown (Li and Wong, 2013) that this distance is equivalent to:

$$d_{R_2}(\mathbf{A},\mathbf{B}) = \sqrt{\text{Trace}(\mathbf{A}) + \text{Trace}(\mathbf{B}) - 2\text{Trace}(\mathbf{A}^{\frac{1}{2}}\mathbf{B}^{\frac{1}{2}})} \tag{4}$$

Let the points $\mathbf{A},\mathbf{B} \in \mathcal{M}$ and let $\mathbf{X}$ be a the point on the manifold at which we construct a tangent plane (it is usually denoted as $T_{\mathcal{M}}\mathbf{X}$). According to the inner-product $\langle \mathbf{A},\mathbf{B}\rangle_{\mathbf{X}} = \text{Trace}(\mathbf{X}^{-1}\mathbf{A}\mathbf{X}^{-1}\mathbf{B})$ the log- Riemannian metric is given as (Moakher, 2005):

$$d_{R_3}(\mathbf{A},\mathbf{B}) = \left\|\log(\mathbf{A}^{-\frac{1}{2}}\mathbf{B}\mathbf{A}^{-\frac{1}{2}})\right\|_2 = \sqrt{\sum_{i=1}^{M}\log^2(\mathcal{L}_i)} \tag{5}$$

where the $\mathcal{L}_i$'s are the eigenvalues of the matrix $\mathbf{A}^{-1}\mathbf{B}$ (Absil et al., 2009). (Metric d_{R3} has been developed in various ways and has, for a long time, been used in theoretical physics).

In order to solve the corresponding minimization problems we presented detailed computational algorithms for calculating these distances in (Jahromi et al., 2015). In all the cases certain iterative procedures are necessary however we demonstrated existence of unique solutions (means) for all the proposed distances.

2.3 Local Detectors

We then construct two separate ensembles: in the absence of seizures we calculate sequence of PSD matrices $F_i^j|H_0$ for $i = 1,\ldots,p$ where p is the total number of windows for the jth EEG channel and similarly, when the seizure is present we calculate $F_i^j|H_1$ $i = 1,\ldots,q$ where q is the total number of windows for the jth channel and particular seizure epoch. Note that in the preliminary approach we will use a single channel detection system using $C3$ channel based on $10-20$ system labels.

For each ensemble we then calculate the centre of the ensemble by using a Frechet mean given as the point which minimizes the sum of the squared distances (Barbaresco, 2008):

$$\mathcal{S}|\hat{\mathbf{H}}_\mathbf{l} = \text{argmin}_{\mathcal{S}\in\mathcal{M}} \sum_{i=1}^{n} d^2(\mathbf{S}_i|H_l, \mathcal{F}|\mathbf{H}_\mathbf{l}) \tag{6}$$

where $d(.,.)$ denotes the metric being used respectively. Therefore the above expression can be interpreted as a way of calculating an averaged sample psd matrix using a sliding window where S_i represents an $i-th$ PSD window. Then we calculate empirical pdf by finding by modelling the pdf as a set of radial basis functions using $\mathcal{S}|\hat{\mathbf{H}}_\mathbf{l}$ as a centre i.e.

$$\hat{p}_j(d|H_i) = \sum_{l=1}^{n} \alpha_l e^{-\|d-\|\mathcal{S}\|_{d_j}^2/\beta_l^2} \tag{7}$$

where subscript i denotes hypothesis (seizure present or no seizure) and subscript j denotes with respect to which of the three aforementioned distances was used. Note that we obtain 6 different empirical pdfs using two hypotheses and three distances. The unknown coefficients α_l and β_l are obtained by applying least squares fit on the empirical counts based on the training set (expert annotations).

The local decisions $u_n, n = 1,2,3$ can be expressed as

$$u_n = \begin{cases} 0, & \text{the } n\text{th detector favours } H_0 \\ 1, & \text{the } n\text{th detector favours } H_1 \end{cases} \tag{8}$$

where "favours" should be interpreted in the following way. If the prior probabilities are know pick i so that $P(H_i)\hat{p}_n(d|H_i) \geq P(H_{1-i})\hat{p}_n(d|H_{1-i})$ (i.e. maximum a posteriori detector) and if they are being treated as equally likely then pick $\hat{p}_n(d|H_i) \geq \hat{p}_n(d|H_{1-i})$ (i.e. maximum likelihood detector). In the remainder of the paper we will be using MAP detector as the patients admitted to NICU have sufficient number of seizure epochs.

2.4 Distributed Detection System

Each of the metric detectors presented in the previous section can be considered as a single channel i.e. local detector. In order to improve the overall performance of a single detectors we propose to combine the existing single detectors and utilize their strengths by extending previous results on blind multichannel information fusion (Liu et al., 2007).

Figure 3 shows the structure of a typical parallel distributed detection system with N detectors. The local detectors transmit local decisions u_n based on a particular metric that they are using. Obviously in

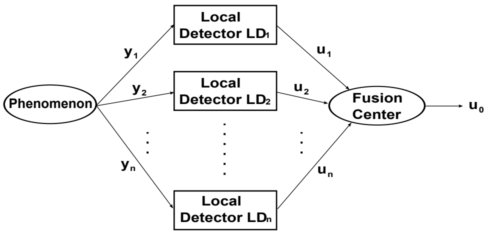


Figure 3: Parallel Distributed Detection System.

our case there are three local detectors as we are using three different metrics. All the local decisions are then sent to the fusion centre, where the global decision u_0 is made based on a fusion rule in order to minimize the overall probability of error. Additional detectors can be added into the system whenever more information is required to make final decision.

The local decisions u_n, $n = 1,2$ can be expressed as

$$u_n = \begin{cases} 0, & \text{the } n\text{th detector favours } H_0 \\ 1, & \text{the } n\text{th detector favours } H_1 \end{cases} \quad (9)$$

where "favours" should be interpreted in the following way: pick i so that $\hat{p}_n(d|H_i) \geq \hat{p}_n(d|H_{1-i})$.

After receiving the local decisions, the fusion centre makes the global decision by applying an optimal fusion rule in order to minimize the final error probability. The authors provided the optimality criterion for N local detectors in the sense of minimum error probability in (Varshney, 1986). We recall it here for the case of $N = 3$.

$$u_0 = \begin{cases} 1, & \text{if } w_0 + \sum_{n=1}^{3} w_n > 0 \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

where,

$$w_0 = \log\left(\frac{P_1}{P_0}\right) \quad (11)$$

and

$$w_n = \begin{cases} \log((1-P_n^m)/P_n^f), & \text{if} \quad u_n = 1 \\ \log(P_n^m/(1-P_n^f)), & \text{if} \quad u_n = 0 \end{cases} \quad (12)$$

The probabilities of false alarm and missed detection of the nth local detector are denoted as P_n^f and P_n^m, respectively. The optimal fusion rule tells us that the global decision u_0 is determined by the a priori probability and the detector performances, i.e., P_1, P_n^f and P_n^m. In our previous work we considered these probabilities to be unknown (Mirjalily, 2003),(Liu et al., 2007). In the current work we assume prior probabilities $P(H_0)$ and $P(H_1)$ are unknown as they can change

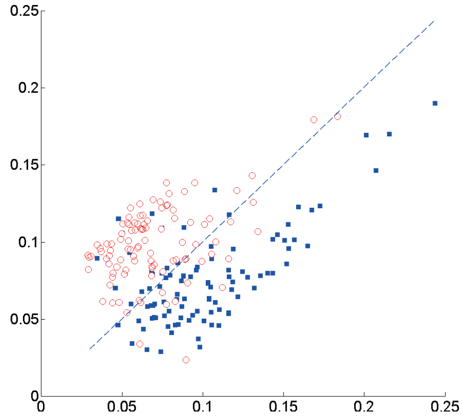


Figure 4: Scatter plot of detection performance using blind method.

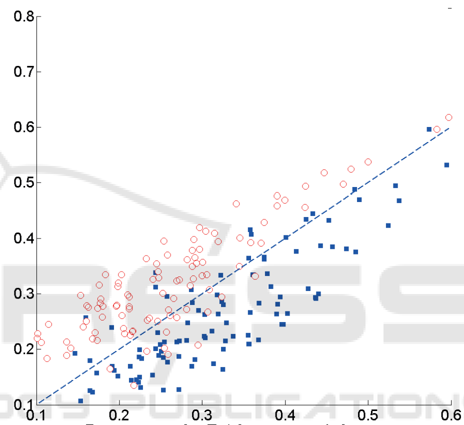


Figure 5: Scatter plot of detection performance using MAP method.

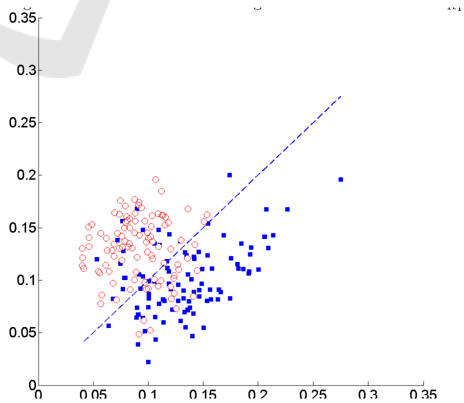


Figure 6: Scatter plot of detection performance using ML method.

significantly with time and depend on the neonate's state but we assume that the anomalies are estimated from the empirical pdf distributions given in 7.

In order to make the final decision, we need to utilize the information available to us: the local binary decisions u_n. Note that in the presence of the training set (annotations) the initial guesses for unknown parameters can be obtained from the training set but they are non-stationary and change with time. The details of the implementation are given in our previous work, (Liu et al., 2014), (Jeremic and Nikolic, 2019).

3 RESULTS

We evaluate the performance of the proposed algorithms on the data set consisting of preterm infants (GA less than 32 weeks) admitted to the Neonatal Intensive Care Unit at McMaster Hospital. Due to physical limitations we were able to obtain prior expect knowledge on a very limited time length and limited set of patients. We selected only patients with seizure epochs and obtained expert annotations on a limited length (2 hours per patient).

For illustrational purposes in Figures 4-6 , we plot the detection performance as a scatter diagram of windows selected from testing data. Note that in the presence of motion artifacts the actual performance will actually vary significantly. Furthermore because the original system design was based on no-seizures the system was calibrated so that the probability of false alarm is controlled. Due to motion artifacts and reaction to pain stimuli during medical procedures in NICU it is quite likely that local detectors will identify these manifestation in EEG as false seizure. In Table 1 we present the results of our previously proposed blind system (Jeremic and Nikolic, 2019) without any training in which the detector anomalies and priors are estimated and the local detectors are based on (Rankine et al., 2007), (Gotman, 1997) and (Celka and Colditz, 2002). In Tables 2 and 3 we illustrate our two proposed algorithms with average probability of error averaged of 1000 randomized training set runs. As expected the proposed system performs better due to the fact that expert annotations are available.

For comparison purposes we also implemented a support vector machine (SVM) classifier which attempts to find an optimal hyperplane in the feature (or reduced dimension) space which minimizes overall probability of classification error. To this purpose we use PSD images and reduce their dimensionality using principal component analysis (PCA) as feature reduction preprocessing technique. The overall average accuracy of SVM was 82% for seizure-free epochs and 78% for seizure epochs. The number of features selected was set to 20 in order to capture 85% of the variance (arbitrarily set). Note that the number of features can be selected optimally and it will be addressed in future work.

Table 1: Average seizure detection performance - blind.

	d_{R1}	d_{R2}	d_{R3}	Fused
false seizures	0.14	0.15	0.16	0.11
missed seizures	0.17	0.14	0.16	0.15

Table 2: Average seizure detection performance - training set maximum a posteriori.

	d_{R1}	d_{R2}	d_{R3}	Fused
false seizures	0.07	0.09	0.12	0.05
missed seizures	0.09	0.08	0.11	0.07

Table 3: Average seizure detection performance - training set based maximum likelihood.

	d_{R1}	d_{R2}	d_{R3}	Fused
false seizures	0.09	0.11	0.15	0.09
missed seizures	0.11	0.10	0.14	0.12

4 CONCLUSIONS

Automatic systems for seizure detection have been subject of considerable research interest in the past. One of main advantages lies in the fact that expert time is potentially required only during the training session. Furthermore, for newborn patients admitted to NICU such systems enable continuous monitoring of seizure events and hence can provide better insight into neurological development. In recent years significant effort has been placed on developing systems that predict seizures in order to potentially counter them with appropriately generated electrical stimuli. To this purpose in this paper we examined possibility of detecting seizures by measuring different distances using STFT. To achieve this goal we define local detectors using empirically determined parameters and fuse their local decisions using our previously developed information fusion algorithm for seizure detection. We demonstrated the applicability of the proposed algorithms using a real data set consisting of multiple NICU patients and expert annotations.

Our results indicate that training techniques offer better performance if adequate expert annotations are available. An effort should be placed on examining possibility of using machine learning techniques which would enable efficient management of resources. Due to patient-to-patient variability we expect that certain semi-supervised approach will have to be used in order to adjust parameters of the detection system to a particular patient as number of seizures may be insufficient in the beginning stage immediately after admission to NICU. Nonetheless, we expect that semi-supervised machine learning and/or deep learning techniques could provide adequate seizure detection with acceptable error levels once the feature reduction algorithm is optimally designed.

REFERENCES

Absil, P.-A., Mahony, R., and Sepulchre, R. (2009). *Optimization algorithms on matrix manifolds*. Princeton University Press.

Barbaresco, F. (2008). Innovative tools for radar signal processing based on Cartan's geometry of SPD matrices & information geometry. *Radar Conference, 2008. RADAR'08. IEEE*, pages 1–6.

Celka, P. and Colditz, P. (2002). A computer-aided detection of EEG seizures in infants: asingular-spectrum approach and performance comparison. *IEEE Trans. on biomedical engineering*, 49(5):455–462.

Faul, S. e. a. (2005). An evaluation of automated neonatal seizure detection methods. *Clinical Neurophysiology*, 116(7):1533–1541.

Gotman, J. e. a. (1997). Automatic seizure detection in the newborn: methods and initial evaluation. *Electroencephalography and clinical neurophysiology*, 103(3):356–362.

Jahromi et al., M. (2015). Estimating Positive Definite Matrices Using Frechet Mean. In *Biosignals 2015*.

Jeremic, A. and Nikolic, D. (2019). Detecting neonatal seizures using sample covariance estimation. *Proc. Biosignals 2019*, 4(1):225–230.

Li, Y. and Wong, K. M. (2013). Riemannian distances for EEG signal classification by power spectral density. *IEEE journal of selected selected topics in signal processing*.

Liu, B., Jeremic, A., and Wong, K. (2007). Blind adaptive algorithm for M-ary distributed detection. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007*, volume 2.

Liu, B., Jeremic, A., and Wong, K. (2014). Optimal distributed detection of multiple hypotheses using blind algorithm. *IEEE Trand. on Aerospace and Electronic Systems*, 50:1190–1203.

Mirjalily, G. e. (2003). Blind adaptive decision fusion for distributed detection. *IEEE Transactions on Aerospace and Electronic Systems*, 39(1):34–52.

Moakher, M. (2005). A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 26(3):735–747.

Rankine, L., Stevenson, N., Mesbah, M., and Boashash, B. (2007). A nonstationary model of newborn EEG. *IEEE Trans. on Biomed. Eng.*, 54(1):19–28.

Temko, A., Marnane, W., Boyland, G., and Lightbody, G. (2015). A computer-aided detection of EEG seizures in infants: asingular-spectrum approach and performance comparison. *Decision Support Systems*, 70:86–96.

Varshney, P. (1986). Optimal data fusion in multiple sensor detection systems. *IEEE Trans. on Aerospace and Electronic Systems*, pages 98–101.

Volpe, J. (2001). *Neurology of the newborn*. WB Saunders Co.