GRADING BRAIN INJURY IN NEONATAL EEG USING SVM AND SUPERVECTOR KERNEL

Rehan Ahmed[1,2], *Andriy Temko*[1,2], *William Marnane*[1,2], *Geraldine Boylan*[1,3] *and Gordon Lightbody*[1,2]

[1]Neonatal Brain Research Group, Irish Center for Fetal and Neonatal Translational Research
[2]Department of Electrical and Electronics Engineering, University College Cork, Ireland
[3] Department of Pediatrics and Child Health, University College Cork, Ireland

ABSTRACT

Brain injury at the time of birth could lead to severe neurological dysfunction at an older age. Grading the brain injury in the early hours after birth could help doctors determine a prompt and reliable treatment. This work presents an automated neonatal EEG grading system based on a cross-disciplinary method of using Support Vector Machine and supervectors, initially developed for speaker identification. The EEG is classified into one of the four grades of neonatal brain injury. The preliminary results show promising performance and are an improvement on the previously published results.

1. INTRODUCTION

Hypoxic-ischaemic encephalopathy (HIE) is the most common cause of neonatal deaths and long-term neonatal neurological dysfunction [1] with reported incidences of 3-5 per 1000 births [2]. HIE injury is a result of the lack of oxygen to the neonatal brain around the time of birth. The outcomes of the HIE injury at the later ages depends on the severity of the HIE insult. Mild encephalopathy could have a normal outcome, moderate encephalopathy having a 20-40% risk of neurological disability and severe encephalopathy leading to severe neurological disability or death [3]. Thus an early detection of the grade of HIE injury is of utmost importance for doctors to prescribe an early treatment.

HIE is graded using Electroencephalography (EEG) in four main types [4] as briefly described in Table 1 and shown in Fig. 1. As can be seen the two main features that differentiate HIE grades are Inter-Burst-Interval (IBI) and the discontinuity of the background EEG. In clinical practice, grading HIE requires the presence of highly qualified neurophysiologists and considerable time. This expertise is not widely available in Neonatal Intensive Care Units (NICU). An automated system for grading HIE could be of great help for the medical staff.

Automatic grading of HIE affected EEG (HIE-EEG) is a relatively new area. In a recent study, Stevenson et. al. proposed an automated system for grading HIE using a multi-class linear discriminant classifier trained on the amplitude modulation and instantaneous frequency features [5]. An accuracy of 77.8% was reported.

The HIE-EEG exhibit various patterns as can be seen in Fig. 1. Some of these patterns could be similar across HIE grades and only their inter-pattern variability along the whole EEG recording characterizes its grade. In this sense, the classification of EEG into different grades resembles a problem of closed-set text-independent speaker identification where statistical information of the underlying phonetic variability within the whole utterance is used to classify a speaker [6].

The most important part of speaker identification system is the creation of the speaker models. Both generative and discriminative modeling approaches have been reported in literature [6]. A combination of these approaches resulted in development of the supervector Support Vector Machine (SVM) SVM, where stacked Gaussian Mixture Model (GMM) means are fed into the linear SVM [7]. This method has also shown promising results in other pattern recognition areas [8, 9, 10].

In this study, we have developed an automated HIE-EEG grading system using the supervector SVM approach.

The paper is organized as follows: Section 2 describes the dataset used in this work. GMM supervectors and their use with the SVM is outlined in Section 3. The developed automatic HIE-EEG grading system is explained in Section 4. Results and discussion are presented in Section 5 with conclusions drawn in Section 6.

2. DATASET

The dataset used in this study comprises approximately one hour long EEG selected from the recordings of 54 full term neonates with HIE. The data was recorded in the NICU of Cork University Maternity Hospital, Cork, Ireland. 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. 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. Each EEG recording had continuous presence of a specific HIE grade. The EEG data were free of seizures and major movement artifacts. The files were graded

This work was supported by Science Foundation Ireland (SFI) Principal Investigator Award (10/IN.1/B3036) and an SFI Research Centers Award (12/RC/2272).

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

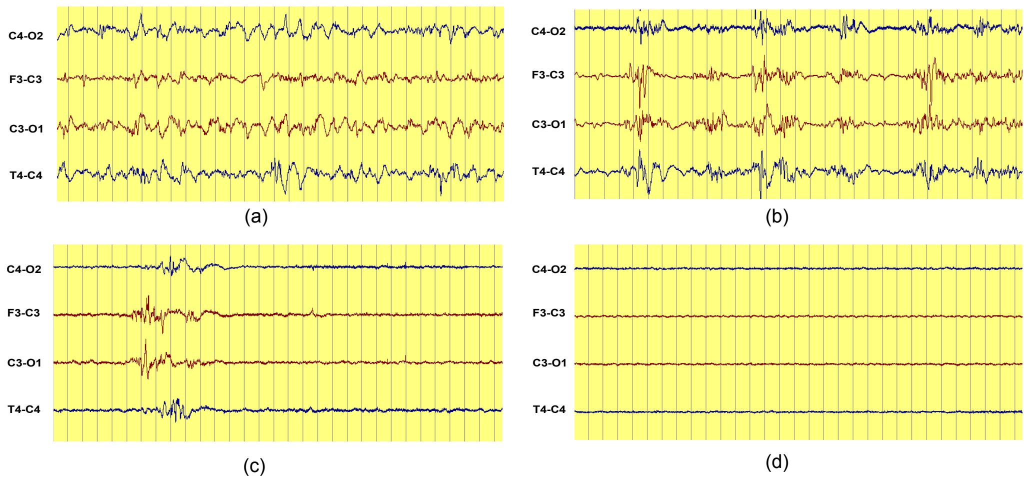

Fig. 1. Ideal epochs of EEG showing grades of HIE (a) Grade 1: Normal/Mild abnormalities. (b) Grade 2: Moderate abnormalities. (c) Grade 3: Major abnormalities. (d) Grade 4: Inactive.

by two independent EEGers using the system defined in [4] and summarized in Table 1. The same dataset has previously been used in [5] and thus a direct comparison of results is possible.

3. GMM SUPERVECTOR AND SVM

The GMM models a probability density function as a weighted sum of M Gaussian components:

$$p(\mathbf{x}) = \sum_{i=1}^{M} w_i \, g(\mathbf{x}|\mathbf{m}_i, \mathbf{\Sigma}_i), \tag{1}$$

where $\mathbf{x}$ is a feature vector, w_i is the mixture weight with $\mathbf{m}_i$ the mean and $\mathbf{\Sigma}_i$ the covariance matrix for the i^{th} mixture component $g(\mathbf{x}|\mathbf{m}_i, \mathbf{\Sigma}_i)$.

Given a set of N feature vectors $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ and a Universal Background Model (UBM), a new GMM model is created by adapting the means $\mathbf{m}_i$ of the UBM using Maximum-a-Posteriori (MAP) adaptation. A GMM supervector, $\mathbf{v}$, is created by concatenating all the means of the new model.

Kernels in the SVM are used to map the data from the input space into a high-dimensional space where the data become linearly separable. It is shown in [7] that an inner product between two supervectors, $\mathbf{v_a}$ and $\mathbf{v_b}$, is an upper bound of the Kullback-Leibler divergence between the two GMMs. Thus, the SVM kernel function is defined as,

$$\begin{aligned} K(\mathbf{v_a}, \mathbf{v_b}) &= \sum_{i=1}^{M} w_i (\mathbf{m}_i^a)^t \mathbf{\Sigma}_i^{-1} \mathbf{m}_i^b \\ &= \sum_{i=1}^{M} (\sqrt{w_i} \mathbf{\Sigma}_i^{-\frac{1}{2}} \mathbf{m}_i^a)^t (\sqrt{w_i} \mathbf{\Sigma}_i^{-\frac{1}{2}} \mathbf{m}_i^b). \end{aligned} \tag{2}$$

Table 1. HIE Grades

Grade	Description
1	Normal/Mild abnormalities: Continuous background pattern with mild asymmetric patterns, mild voltage depression
2	Moderate abnormalities: Discontinuous activity with IBI $\leq$ 10s
3	Major abnormalities: Discontinuous activity with IBI 10-60s, severe fading background patterns
4	Inactive: Background activity of $\leq 10\mu V$ or severe discontinuity of IBI $\geq$ 60s

It can be seen that Eq. 2 represents the dot product between the two supervectors scaled by a factor of $\sqrt{w_i}\mathbf{\Sigma}_i^{-\frac{1}{2}}$. Note that the scaling terms, weight w_i and covariance matrix $\mathbf{\Sigma}_i$, are the same for all sequences and can be computed beforehand. This will allow the use of a simple linear kernel inside the SVM.

4. AUTOMATED EEG GRADING SYSTEM

4.1. Preprocessing and Feature Extraction

An overview of the complete system is shown in the Fig. 2. The EEG from each channel is first down-sampled from 256Hz to 32Hz with an anti-aliasing filter set at 12.8Hz. The down-sampled and filtered EEG is then segmented into 8s epochs with 50% overlap. A total of 55 features are extracted from each EEG epoch. This feature set provides a generic EEG description which is computed from the frequency, time and information theory domains as described below.

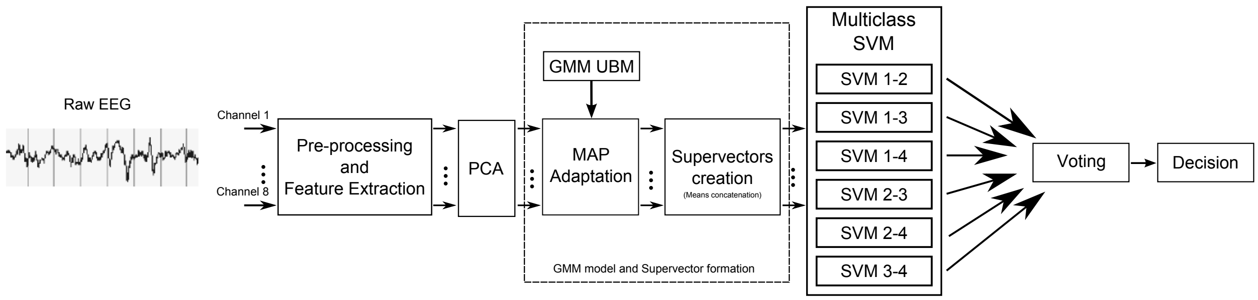

Fig. 2. Automated neonatal HIE-EEG grading system.

Frequency Domain: The power spectrum density (PSD) of each epoch is obtained using a 256 point Fast Fourier Transform (FFT). A number of features are extracted from the PSD of the epoch. Additionally, the EEG is decomposed into 8 coefficients using the Daubechies 4 wavelet, the energy in the 5th coefficient corresponding to 1-2Hz is used as a feature.

- Total power (0-12Hz)
- Peak frequency of spectrum
- Spectral edge frequency (80%, 90%, 95%)
- Power in 2Hz width subbands (0-2Hz, 1-3Hz, ...10-12Hz)
- Normalised power in subbands
- Wavelet energy

Time Domain: A number of features are extracted from the epoch of EEG and the first and second derivative of the EEG denoted by Δ and $\Delta\Delta$ respectively.

- Curve length
- Number of maxima and minima
- Root mean squared amplitude
- Hjorth parameters
- Zero crossings (raw epoch, Δ and $\Delta\Delta$)
- Autoregressive modelling error (model order 1-9)
- Skewness
- Kurtosis
- Nonlinear energy
- Variance (Δ and $\Delta\Delta$)

Information Theory: Features based on information theory were chosen based on an analysis of features by Faul et al. [11].

- Shannon entropy
- Singular value decomposition entropy
- Fisher information
- Spectral entropy

The usability of these features for EEG has been discussed in previous work on neonatal [12, 13, 14], adult seizure detection [15], and neurological outcome prediction [16]. Thus, the extracted features can capture the information required for grading HIE-EEG.

4.2. UBM and Supervector

After feature extraction, Principal Component Analysis (PCA) is then used to reduce the number of correlated features. This also allows the use of a diagonal covariance matrix within the GMM training. The UBM is created by training a GMM with all the training data. This model represents the diversity of EEG across all HIE grades, artifacts etc. This UBM compensates for the lack of training data available for direct training of individual GMMs for each HIE grade.

Then, the sequences of 20 feature vectors are created which span over 80 seconds of the EEG signal. A similar duration of EEG was chosen for processing in [5]. A GMM model is created for each sequence by the MAP adaptation of means of the UBM. The adapted means are then concatenated to form a supervector.

4.3. Classification

The supervectors are fed to the SVM. We have used a one-against-one approach for multi-class SVM classification [17]. Each SVM model is trained from the supervectors of two grades with 2-fold cross validation used on the training data to find the parameter C for the linear SVM.

For each sequence, the output of the multi-class SVM is a 6-dimensional vector which contains the grades assigned by each classifier. Majority voting is performed on this vector to get a decision for a given sequence. Decisions from all the sequences in a recording from all 8 channels are concatenated in one vector. An overall majority voting is then performed on this vector to determine the HIE grade for a complete EEG recording.

5. RESULTS AND DISCUSSION

The Leave One Out (LOO) performance assessment is used in this work. In this routine, the system is trained using the data from 53 recording and remaining one unseen recording

Table 2. Confusion Matrix of the output of the system's output and actual assigned grade by the EEGer.

Actual	System's output					
Grade	1	2	3	4	Total	Incorrect
1	21	1	0	0	22	1
2	5	9	0	0	14	5
3	1	1	10	0	12	2
4	0	0	1	5	6	1
Precision(%)	80	81	90	100		

is used to test the system. The process is repeated for each recording. The LOO routine produces an almost unbiased performance assessment of the developed system [18].

Table 2 shows the results of the proposed system. The overall accuracy of the proposed automated HIE grading system was 83.3%. These results are superior to the 77.8% presented in [5]. As can be seen 9 out of 54 recordings were misclassified, with most confusion observed between grades 1 and 2. This is similar to the results presented in [5].

The last row of the confusion matrix in Table 2 shows the precision of the system which is defined as the ratio between the number of correctly assigned decisions and the number of total decisions assigned to a specific grade. It can be seen that the system is over 90% precise for classifying the data of grades 3 and 4, whereas its precision is significantly lower for grade 1 and 2.

The confidence level of assigning a grade to a testing recording was also investigated. This can be extracted from the number of winning votes in the majority voting procedure. A threshold on the percentage of winning votes is used to label decisions as certain or uncertain.

Fig. 3 shows the accuracy of the system at various confidence levels. As can be seen the accuracy increases when only certain decisions are made. The improvement comes at the expense of not taking into account the recordings with uncertain decisions. It can be seen that the accuracy increases almost linearly with the increase of uncertain decisions. The circle shows the performance of the system without uncertainty as described in the Table 2 whereas the square indicates the best achievable accuracy of 96% at the expense of not making a decision for almost half of the recordings. This point corresponds to making a decision with at least 2/3 majority.

The distribution of certain and uncertain correct and incorrect decisions made by the system for the square marker is shown in Fig. 4. Of the total number of the correctly classified recordings (45), 27 were correctly classified with certainty, 18 were correctly classified with uncertainty. Among 9 mis-classifications, only 2 were misclassified with certainty and 7 were uncertain mis-classifications. As expected, most

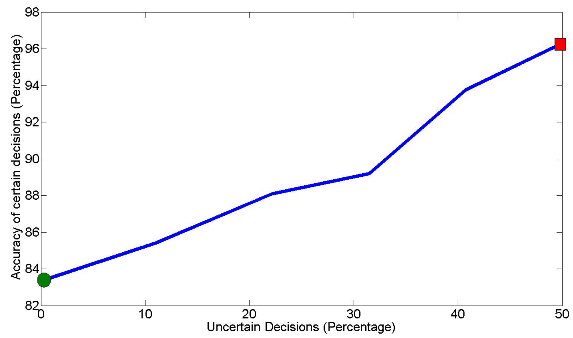

Fig. 3. Accuracy of the certain decisions at the expense of not accommodating the uncertain decisions

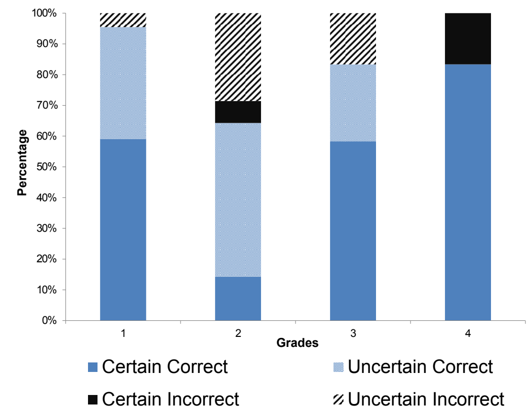

Fig. 4. The distribution of certainty of decisions made by the system.

uncertain decisions occurred between grade 1 and 2.

The major cause of the decreased accuracy of the system for classifying grade 1 and 2 is the similar morphology of the EEG signals in both grades, as could be seen in the Fig. 1. It was also observed that except for one recording, the second best assigned grade to the misclassified files was their actual correct grade.

6. CONCLUSIONS AND FUTURE WORK

A novel system of grading HIE injury in neonatal EEG has been developed. This system is based on a cross-disciplinary approach of using the supervector SVM technique that was originally developed for the speaker identification problem. Promising performance was reported when compared to the current state-of-the-art results [5]. Some of the key areas to be focused in the future is the modeling of sleep cycling states, incorporation of other physiological signals such as heart rate variability to the proposed framework, and refined post-processing steps.

7. REFERENCES

[1] R. C. Vannucci, "Hypoxic-ischemic encephalopathy," *American Journal of Perinatology*, vol. 17, no. 3, pp. 113–120, 2000.

[2] E. M. Graham, K. A. Ruis, A. L. Hartman, F. J. Northington, and H. E. Fox, "A systematic review of the role of intrapartum hypoxia-ischemia in the causation of neonatal encephalopathy," *American Journal of Obstetrics and Gynecology*, vol. 199, no. 6, pp. 587 – 595, 2008.

[3] C. Robertson and N. Finer, "Term infants with hypoxic-ischemic encephalopathy: Outcome at 3.5 years," *Developmental Medicine & Child Neurology*, vol. 27, no. 4, pp. 473–484, 1985.

[4] D. Murray, G. Boylan, C. A. Ryan, and C. Sean, "Early EEG findings in hypoxic-ischemic encephalopathy predict outcomes at 2 years," *Pediatrics*, vol. 124, no. 3, pp. 459–467, 2009.

[5] N. J. Stevenson, I. Korotchikova, A. Temko, G. Lightbody, W. P. Marnane, and B. Boylan, "An automated system for grading EEG abnormality in term neonates with hypoxic-ischaemic encephalopathy," *Annals of Biomedical Engineering*, vol. 41, no. 4, pp. 775–785, 2013.

[6] T. Kinnunen and H. Li, "An overview of text-independent speaker recognition: From features to supervectors," *Speech Communication*, vol. 52, no. 1, pp. 12 – 40, 2010.

[7] W. M. Campbell, D. E. Sturim, D. A. Reynolds, and A. Solomonoff, "SVM based speaker verification using a GMM supervector kernel and NAP variability compensation," in *Proc. ICASSP*, 2006, vol. 1, pp. 97–100.

[8] C. H. You, K. Lee, and H. Li, "An SVM kernel with GMM-supervector based on the bhattacharyya distance for speaker recognition," *Signal Processing Letters, IEEE*, vol. 16, no. 1, pp. 49–52, 2009.

[9] H. Hu, M. Xu, and W. Wu, "GMM supervector based svm with spectral features for speech emotion recognition," in *Proc. ICASSP*, 2007, vol. 4, pp. 413–416.

[10] X. Zhuang, X. Zhou, M. A. Hasegawa-Johnson, and T. S. Huang, "Real-world acoustic event detection," *Pattern Recognition Letters*, vol. 31, no. 12, pp. 1543 – 1551, 2010.

[11] S. Faul, G. Boylan, S. Connolly, W. Marnane, and G. Lightbody, "Chaos theory analysis of the newborn EEG - is it worth the wait?," in *Intelligent Signal Processing, IEEE International Workshop on*, 2005, pp. 381–386.

[12] E. M. Thomas, A. Temko, G. Lightbody, W. P. Marnane, and G. B. Boylan, "Gaussian mixture models for classification of neonatal seizures using EEG," *Physiological Measurement*, vol. 31, no. 7, 2010.

[13] J. Gotman, D. Flanagan, J. Zhang, and B. Rosenblatt, "Automatic seizure detection in the newborn: methods and initial evaluation," *Electroencephalography and Clinical Neurophysiology*, vol. 103, no. 3, pp. 356 – 362, 1997.

[14] B. R. Greene, S. Faul, W. P. Marnane, G. Lightbody, I. Korotchikova, and G. Boylan, "A comparison of quantitative EEG features for neonatal seizure detection," *Clinical Neurophysiology*, vol. 119, no. 6, pp. 1248 – 1261, 2008.

[15] S. Faul, A. Temko, and W. Marnane, "Age-independent seizure detection," in *Proc. EMBC*, 2009, pp. 6612–6615.

[16] O. M. Doyle, A. Temko, D. M. Murray, G. Lightbody, W. Marnane, and G. B. Boylan, "Predicting the neurodevelopmental outcome in newborns with hypoxic-ischaemic injury," in *Proc. EMBC*, 2010, pp. 1370–1373.

[17] H. Chih-Wei and L. Chih-Jen, "A comparison of methods for multiclass support vector machines," *Neural Networks, IEEE Transactions on*, vol. 13, no. 2, pp. 415–425, 2002.

[18] V. Vapnik, *Estimation of Dependences Based on Empirical Data: Empirical Inference Science*, Information Science and Statistics. Springer, Dordrecht, 2006.