Bio-Medical Materials and Engineering 26 (2015) S1569–S1578
DOI 10.3233/BME-151456
IOS Press

Patient prognosis based on feature extraction, selection and classification of EEG periodic activity

Alain Sánchez-González[a,*], Begoña García-Zapirain[a], Iratxe Maestro Saiz[b] and Izaskun Yurrebaso Santamaría[b]
[a]*DeustoTech-Life, University of Deusto, Av de las Universidades, 24, 48007, Biscay, Spain*
[b]*Cruces University Hospital and Bio-Cruces Health Research Institute, Biscay, Basque Country, Spain*

Abstract. Periodic activity in electroencephalography (PA-EEG) is shown as comprising a series of repetitive wave patterns that may appear in different cerebral regions and are due to many different pathologies. The diagnosis based on PA-EEG is an arduous task for experts in Clinical Neurophysiology, being mainly based on other clinical features of patients. Considering this difficulty in the diagnosis it is also very complicated to establish the prognosis of patients who present PA-EEG. The goal of this paper is to propose a method capable of determining patient prognosis based on characteristics of the PA-EEG activity. The approach, based on a parallel classification architecture and a majority vote system has proven successful by obtaining a success rate of 81.94% in the classification of patient prognosis of our database.

Keywords: Medical classification, bioinformatics, feature selection, EEG periodic activity

1. Introduction

The electroencephalogram (EEG) can provide valuable information about diagnosis and prognosis to aid physicians. EEG is recorded using a collection of scalp-electrodes that are fitted to the scalp and is used routinely to assess cerebral disorders [1].

The proposed research in this paper is based on previous work carried out by our research group [2], in which mathematical methods of analysis were developed for a particular type of EEG Periodic Activity (PA-EEG). This activity was observed in patients who were or might have been suffering from epilepsy, a disease whose early and accurate diagnosis, impossible in many cases, determines patient treatment and facilitates prognosis. The signals present in epilepsy consist of repetitive waveforms that exhibit relatively uniform morphology and duration and that occur at approximately regular, consecutive time intervals. They are commonly classified as lateralized periodic discharges (LPD), bilateral independent discharges (BIPDs) and generalized periodic discharges (GPDs), the latter of which often include diphasic waves. All forms of periodic discharges are commonly found in critical patients, especially in intensive care unit settings (ICU).

*Address for correspondence: Alain Sánchez-González, DeustoTech-LIFE, University of Deusto, Av de las Universidades, 24, 48007, Biscay, Spain. Tel.: +34 94413 90 00; Fax: 944 456 817; Email: sanchez.alain@deusto.es.

An attempt is made here to extend an analytical methodology that has been previously developed [2] to a larger population group, for instance patients with neurological pathologies of different aetiology and benignity, in which patterns of periodic activity in routine EEG recordings are commonly found.

The periodic appearance of the EEG activity is a feature common to several clinical entities. For example, it occurs frequently in patients with past experience of epilepsy or patients with a clinical compatible case of repeated seizures with or without clear motor symptoms, which normally facilitate the diagnosis.

There is a significant group of patients in which the occurrence of these periodic EEG patterns prompts the possible diagnosis of epileptic states. However the presented clinical signs are not always clearly suggestive of epilepsy. This lack of clarity raises doubts about the treatment and prognosis of the patients.

Establishing a mathematical clinical correlation among PA-EEG findings in these patients by making a classification according to their benign or malign prognosis will make it possible to adopt the most suitable treatment for the patient, thus improving their quality of life. The prognosis of the patients is determined by patient survival, which means that any assistance to the physicians is very valuable. It is therefore of great importance to develop algorithms to classify patients according their prognosis.

2. Materials and methods

2.1. Acquiring EEG data

The data was collected at a large tertiary hospital (Cruces University Hospital, Bilbao, Biscay, Spain) by specialists from the Clinical Neurophysiology Department. The recordings were obtained from the routine EEG exams of patients suspected of brain injury or disease. A Neurofax (Nihon-Khoden, Rosbach, Germany) EEG machine, which has 32 EEG channels in the standard 10–20 electrode placement system, was used to record the data. The sampling frequency was set to 500Hz. The data was recorded using a lateral-bipolar montage.

The processing stage at recording uses a finite impulse response (FIR) filter with a band-pass region of 0.3–40 Hz. To preserve the time-domain features of the signal, a zero-phase filter using the forward-reverse method [3] was employed.

2.2. Database

The population group is large, as it includes patients with neurological pathologies of different aetiology and prognosis. This makes it difficult to characterise and create an accurate and robust architecture for classification purposes. The pathologies of patients in the population group include: toxic and metabolic encephalopathy, postanoxic encephalopathy, convulsive and non-convulsive status epilepticus, cerebral infarction, tumours or dementia.

For the purpose of correctly analysing and interpreting EEG signals, the recordings have been marked by experts from the Clinical Neurophysiology Department at Cruces University Hospital, stipulating the time of appearance of the segments of detected activity and their respective channels. The database is composed of ten patients, seven of them with good prognosis and three of them with bad prognosis. These prognoses were obtained from the real evolution of the patients. While the

database is still incomplete, the number of patients with good and bad prognoses being less than necessary, the length of the recordings is large enough to allow for repeated random sub-sampling validation, as explained in Section 2.8, which is enough at this stage of the research.

2.3. Pre-processing of the periodic activity signal

The following processes were carried out on the periodic activity signal prior to extracting the features of the signal: notch filtering at 50 Hz to remove the line noise, normalization with respect to the mean by subtracting the mean to the data, downsampling the signal from the original sampling rate (500 Hz) to 50 Hz and finally, removal of large-amplitude artefacts using the following threshold approach. If the EEG signal s(t) at time instant t0 is greater than a predefined threshold, then set this signal to zero within a small window frame around t0; that is, if $s(t0) > \zeta$, then set $s(t0 + \tau) = 0$ for $-L/2 \leq \tau \leq L/2$; and do nothing otherwise. We set ζ to 400 μv and L to 5 seconds.

2.4. PA-EEG periods selection

PA-EEG periods are selected from EEG signals marked by specialists who define the channels and time of appearance. A necessary condition is that periods should be of a minimum duration of 20 seconds, a requisite demanded by experts in neurophysiology to ensure that segments contain periodic activity. Segments of longer duration are split into 20-second intervals which correspond to 1000 samples, for classification purposes. In this research 105 20s-long segments of good prognosis and 53 20s-long segments of bad prognosis have been used.

2.5. Classification methods

Automatic learning methods have proven their effectiveness in recognising EEG wave patterns [4-8]. Owing to the variety of aetiologies of patients from the current database, the instantaneous frequency, which was the chosen feature to separate epileptic from non-epileptic patients in [2], has been shown to be insufficient for the purpose of proper classification in the present study. For this reason a larger number of features and the use of classifiers have been proposed.

After having carried out many experiments using different classification algorithms, like Neural Networks, Support Vector Machines with Linear, Quadratic and RBF kernels, Adaboost M1 with decision trees and discriminant weak classifiers, a parallel architecture (committee) was implemented [9-11]. This was due to the robustness and reliability they provide to the classification.

The selection of algorithms as candidates to form part of the committee was based on the following diversity measures [12]: Q-statistics, ρ coefficient, disagreement measure and double fault measure. Using these measures allowed for a combination of classifiers to be obtained that best fit the data to be classified. The selected classifiers were Adaboost M1 with Decision Trees weak classifiers, Support Vector Machine Linear kernel and Support Vector Machine RBF kernel. For both kinds of Support Vector Machines, the parameters C and γ were set to 1 and 3 respectively. For the RBF kernel the δ scaling factor was set to 1. The decision regarding classification of the parallel architecture was taken by majority vote.

Realizing a parallel classification architecture adds complexity and computational cost to the system, but these are mainly due to the need for training the different classifiers that compose the committee and, in addition but to a lesser extent, due to implementing a majority-based voting system. Since training is done offline and in a phase prior to setting up the system for a possible online application,

and also since the additional complexity of the committee is negligible, online performance should remain unaltered. Figure 1 shows an outline of the architecture implemented.

2.6. Feature extraction

Feature extraction is a determining factor when classifying patterns. Features need to be insensitive to noise, discriminant and separate from each other. Their main purpose is to objectively describe certain aspects, in this case regarding the PA-EEG activity that may be characterized subjectively by specialists.

A collection of 84 features were selected based on previous studies [13-17] in the time, frequency, wavelet, time-frequency and instantaneous frequency domains.

In order to characterise PA-EEG signals several processing techniques have been employed:

- The Welch method with a 2 second long hamming window to transform the signal to the frequency domain.
- Wavelets to extract various features of the signals from time-frequency sub-bands. By using a Daubechies mother Wavelet with 5 levels of decomposition, the original spectral range is divided into 32 sub-ranges with corresponding detail and approximation coefficients.
- The Wigner-Ville distribution to transform the signal to the time-frequency domain through a lag-independent kernel.

To prevent features with higher absolute values from having greater weight in the learning process, the features were normalized based on the following Eq. (1).

The maximum and minimum values are calculated for each feature such that each is then normalized between 0 and 1.

$$\underline{Feature_i} = \frac{Feature_i - \min_j(Feature_j)}{\max_j(Feature_j) - \min_j(Feature_j)} \tag{1}$$

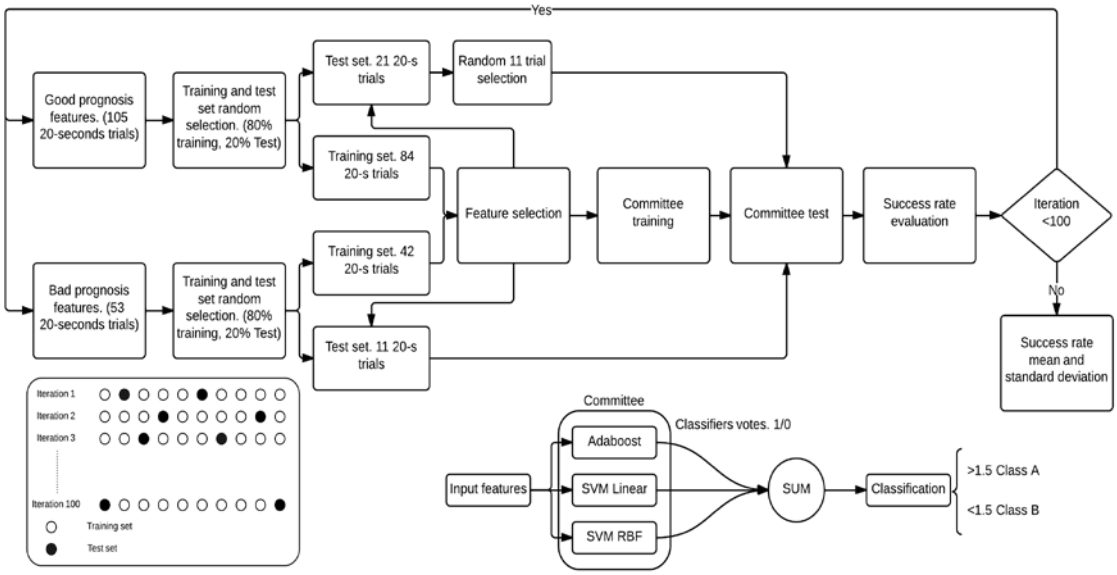

Fig. 1. Classification architecture.

2.7. Feature selection

Feature selection creates a subset of initial attributes improving their predictive performance and constructing patterns more efficiently. One of the major problems with classification algorithms is multidimensionality, which has an adverse effect on their decision-making power. By applying an attribute selection process, we attempt to select the smallest subset to obtain the greatest accuracy.

There are many methods used to select attributes. For further information, we refer the reader to [18]. The main difficulty lies in selecting the method that is best suited to the features to be classified. There are currently many studies on the subject, such as [19, 20] which apply feature selection techniques in order to obtain greater success rates in their classifications.

For simplicity, the features were reduced using the Wilcoxon method, which is built-in in Matlab. More specifically, the method was chosen because the statistical Shapiro-Wilk test showed that the data distribution was non-Gaussian. A subset of 11 features were selected from the original 84 calculated in the different domains, which make our classification method more accurate. Features are chosen based on their relevance, i.e. their discriminant capabilities. In order to obtain the optimal subset of features we have implemented an iterative algorithm that classifies data and computes the success rate, then removes the least relevant feature at each iteration. This is done from the original number of features until only 1 remains, and the subset of features that provided the highest success rate is chosen, in our case when only 11 where remaining. Table 1 shows the selected features.

2.8. Experimental set-up

The experimental set-up consists of a training phase followed by a testing phase, each of which is fed by signal segments (trials) that are obtained at random from the PA-EEG epochs marked by the specialists within the full recordings of the database of patients with good and bad prognosis. Since the database is composed of 7 and 3 patients of each kind, respectively, and the trials are 1000 samples long (20 seconds), there exist 105 trials of patients with good prognosis and 53 trials of patients with bad prognosis. Then, repeated random sub-sampling validation is performed 100 times, as follows:

- Generate independent training (80%) and test (20%) groups in a random fashion from the segments obtained before. Proceeding this way, training and testing are done akin to using a cross-validation method, but with completely random rather than permuted groups at each iteration, thereby increasing the objectiveness of the validation. To test the individual classifiers and the committee, 20% of the available bad prognosis 20-s segments were used (11 20-s segments) as well as the same number of good prognosis segments.
- Select the most relevant features for the training set of the current iteration, simulating a real case in which characteristics cannot be inferred from the recordings (they have to be obtained prior to setting up the system). These features are used for the upcoming experimental tests.

Table 1
Selected features (Mathematical expressions for calculating the features can be seen in the appendix)

1.Time-Frequency coefficients kurtosis. Eq. (2)	2.Time-Frequency coefficients skewness Eq. (3)	3. Wavelets D4 coefficients variance. Eq. (4)	4. Time-frequency sub-band energy. Eq. (5)
5.Time-frequency coefficients roll off. Eq. (6)	6.Wavelet D6 coefficients minimum	7. Wavelets D8 coefficients variance. Eq. (4)	8. Wavelets A8 coefficients variance. Eq. (4)
9.Wavelet D5 coefficients minimum	10. Area under the curve. Eq. (7)	11.Time-frequency sub-band energy. Eq. (5)	

- Train the individual classifiers of the committee. This is still part of the training phase.
- Classify data from the testing set using the features and individual classifiers as parameterized in the previous step.

At this point, it is already possible to calculate the success rate of the current iteration and reiterate the process from the first step.

The above procedure is executed 100 times, finally computing the average success rate and standard deviation of the 100 iterations. Figure 1 shows a detailed diagram of the algorithm outlined before.

3. Results

Table 2 shows the results obtained in the classification of PA-EEG signals, including the standard deviation and the statistical significance of the calculated feature selection. The statistical significance has been calculated using a Wilcoxon test by making a comparison between success rates with and without feature selection.

As mentioned in the abstract, the combined success rate using the committee is of 81.94% with a standard deviation of 9.81 and the p value shows that the results are statistically significant. This success rate is greater than that of any individual method and also better than the rate obtained with the complete set of features. The increment of the success rate of the parallel architecture compared to the single classifiers is not as high as originally expected possibly since the database is still incomplete.

However, the choice of the committee has prevailed in order to have better scalability and a higher improvement of the success rate in later stages of the project. Moreover, in a medical context any enhancement of the diagnoses is desirable, even if small.

Table 3 shows the confusion matrix of the committee obtained from the 100 iterations of the algorithm. It indicates the performance of the committee. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. From here the following statistical measures of the performance of the committee can be calculated: True Positive Rate: 0.83, False Positive Rate: 0.19, True Negative Rate: 0, 8 and False Negative Rate: 0.16. These results indicate that in fact the architecture accurately classifies patients of good and bad prognosis. According to our experiments, the classification problem seems to be nonlinear. We conclude this due to the lowest classification rate corresponding to the only linear classifier of the committee (Linear-SVM). However, Linear-SVM seems essential in the choice of classifiers that make up the committee

Table 2
Classification results, standard deviation and significance p value for the comparison between feature subset and raw features

Classifier	Success rate (Feature subset 11 features)	Std. Dev.	Success rate (raw 84 features)	Std. Dev.	p
Adaboost M1	78.61%	9.42	76.22%	9.47	0.108
SVM-Linear Kernel	76.55%	9.91	76.16%	8.44	0.743
SVM RBF kernel	79.77%	10.55	67.55%	8.45	0
Committee	81.94%	9.81	78.22%	8.63	0.003

Table 3

Committee classification confusion matrix (11 features)

	Good prognosis	Bad prognosis
Good prognosis	881	219
Bad prognosis	178	922

in order to increase the success rate compared to the individual results.

The results obtained in this first phase of the research are positive in several aspects. Firstly, it was ascertained that the characterization of PA-EEG signals and their subsequent classification according to the patient benign or malign prognosis are indeed feasible. This is a complex task for several reasons: there is a wide range of pathologies that generate PA-EEG signals, characterized by very varied wave patterns. What is more, the medication prescribed to patients also affects the form that EEG signals take in many cases. However, it can be ensured that the set of selected features is appropriate given the success rates obtained in this preliminary stage of the research.

Furthermore, feature selection has been shown to be beneficial. A subset of 11 features was selected from the original set increasing the success rate. The features have been evaluated according to their individual relevance, due to the computational cost of evaluating all possible subsets of features. Additionally, no redundancy has been taken into account, despite being one of the most determining factors in feature selection. Independence between features is desirable, because in many cases some of them provide the same information, or tend to mix the boundaries of the different classes decreasing the success rate.

Lastly, diversity measures [20] among the different classifiers comprising the committee have been calculated in order to assess the complementary nature of the parallel classification architecture and to select the most accurate classifiers for the purpose of our casuistry. The means of the diversity measures for the classification committee are: Q-statistics: 0.54, ρ coefficient: 0.36, disagreement measure: 0.21, double fault measure: 0.10. This task entails a lot of complexity, as it is necessary to perform each of the calculations for each of the possible sets of classifiers of the committee. In order to assess the statistical significance between the committee and the individual classifiers, a non-parametric Mann-Whitney test was applied, since the data does not follow a normal distribution. The results are shown in Table 4.

Taking all this into account, there are substantial grounds to believe that the study is in a promising line of work, one that we intend to continue.

4. Conclusions

After carrying out experimental tests based on available AP-EEG data, it can be concluded that accurate automatic classification of different patients according to their prognosis is possible, by no means a straightforward task for physicians because, as has already been stated, many times it is based on other clinical features and is mostly subjective.

Likewise, it has been demonstrated that the reduction in the number of features and the implementation of efficient classification architecture are indeed viable options. This in turn makes us feel optimistic about increasing the number of patients in the database, which will imply an expansion of patterns to be learnt by classifiers of the committee but should also increase the accuracy of the classification.

Table 4

Mann-Whitney test between committee and individual classifiers

	Feature subset 11 features			Original 84 features		
	Committee vs. Adaboost M1	Committee vs. SVM Linear	Committee vs. SVM RBF	Committee vs. Adaboost M1	Committee vs. SVM Linear	Committee vs. SVM RBF
p	0.011	0	0.131	0.183	0.120	0

Classifiers have been chosen to solve the problem at hand due to the variability of pathologies involved in the study. It would otherwise have been too complex to implement an ad hoc signal processing scheme that could differentiate among so many pathologies and succeed with patient diagnosis. On the contrary, by using classifiers and a multi-parameter characterization of PA-EEG recordings, patient groups are easier to handle. In addition, a parallel architecture has been preferred to enhance the scalability of the system in future stages with more extensive databases.

Moreover, the method may be used in a clinical environment in its current form. Once the committee is adequately trained, the classification system can be used to improve medical diagnosis of patients presenting PA-EEG bursts of similar characteristics to those included in the database.

The aim of this research has been to obtain objective data about PA-EEG signals in patients thereby providing significant assistance to the diagnosis issued by specialists from Clinical Neurophysiology units, with there being the chance to propose more effective forms of treatment. By attaining the objectives that have been set out, it may be possible to improve the quality of life of patients.

Lastly, after completing the first phase of the experiment, it is possible to state that characterization of PA-EEG bears little relation to that of the peak trains associated with epileptic patients researched in [2], as the instantaneous frequency does not emerge as one of the most relevant features.

5. Future work

Regarding future lines of research, it is hoped that the database will be expanded to at least 30 patients to improve the reliability of conclusions. Moreover, feature significance will be revised by expanding the database.

In the next phase of the study improved pre-processing of PA-EEG signals will be implemented using techniques such as Independent Component Analysis (ICA) in order to remove artefacts [21].

There is also a desire to implement ad hoc inference measures between the different features, which will facilitate the selection of the most important features in this casuistry according to their independence, discriminant power and complementary nature.

Acknowledgments

This publication has been funded by the eVIDA research group grant from the Education and Research Department of the Basque Country, by Deiker from the University of Deusto and by the Basque Government SAIOTEK program. The authors also wish to thank the anonymous reviewers for their helpful comments.

Appendix

Mathematical expressions for calculating the selected features:

$$Kurtosis = \frac{1}{(NK-1)\sigma_{(t,f)}^4}\sum_{k=1}^{K}\sum_{n=1}^{N}\left(p[n,k]-\mu_{(t,f)}\right)^4 \quad (1)$$

$$Skewness = \frac{1}{(NK-1)\sigma_{(t,f)}^3}\sum_{k=1}^{K}\sum_{n=1}^{N}\left(p[n,k]-\mu_{(t,f)}\right)^3 \quad (2)$$

$$Variance = \sigma^2 = \frac{1}{N}\sum_{n=1}^{N}(\mu-|z[n]|)^2 \quad (3)$$

$$Sub-bands\ energies = \sum_{n=1}^{N}\sum_{k=1}^{K_\delta}p[n,k] \quad (4)$$

$$Spectral\ Roll-off = \lambda\sum_{n=1}^{N}\sum_{k=1}^{K}p[n,k] \quad (5)$$

$$Area\ under\ the\ curve = \frac{1}{N}\sum_{n=1}^{N}|z[n]| \quad (6)$$

Notations: σ: Standard deviation. **μ**: Mean. p[n,k]: time-frequency domain (n: time, k: frequencies). z[n]: time domain signal (n: time). t, f: time and frequencies. λ: frequency under which 85% of the signal power besides.

References

[1] S. Seidel, S. Aull-Watschinger and E. Pataraia, The yield of routine electroencephalography in the detection of incidental nonconvulsive status epilepticus–A prospective study, Clinical Neurophysiology **123** (2012), 459–462.

[2] J.M. O'Toole, B.G. Zapirain, I.M. Saiz, A.B.A. Chen and I.Y. Santamaria, Estimating the time-varying periodicity of epileptiform discharges in the electroencephalogram, 2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA), Montreal, QC, Canada, 2012, pp. 1229–1234.

[3] A.V. Oppenheim, R.W. Schafer and J.R. Buck, Discrete-Time Signal Processing, Prentice-hall Englewood Cliffs, IEEE, 1989.

[4] A. Subasi and M. Ismail Gursoy, EEG signal classification using PCA, ICA, LDA and support vector machines, Expert Systems with Applications **37** (2010), 8659–8666.

[5] S. Sun, C. Zhang and D. Zhang, An experimental evaluation of ensemble methods for EEG signal classification, Pattern Recognition Letters **28** (2007), 2157–2163.

[6] P. Xiaomei, Z. Chongxun, X. Jin and B. Guangyu, Adaboost for improving classification of left and right hand motor imagery tasks, Proceedings of First International Conference on Neural Interface and Control, Wuhan, China, 2005, pp. 10–13.

[7] N. Acır and C. Güzeliş, Automatic spike detection in EEG by a two-stage procedure based on support vector machines, Computers in Biology and Medicine **34** (2004), 561–575.

[8] E. Haselsteiner and G. Pfurtscheller, Using time-dependent neural networks for EEG classification, IEEE Transactions on Rehabilitation Engineering **8** (2000), 457–463.

[9] I. Bonet, A. Rodríguez, M.M. García and R. Grau, Combinación de clasificadores para bioinformática, Computación y Sistemas **16** (2012), 191–201.

[10] L.I. Kuncheva, Combining pattern classifiers: Methods and algorithms, IEEE Transactions on Neural Networks **18** (2007), 964–964.
[11] S. Segrera Francia, M. García and M. Navelonga, Multiclasificadores: mÉtodos y arquitecturas, informe técnico DPTOIA-IT-2006-001, Universidad de Salamanca, Salamanca, 2006.
[12] L.I. Kuncheva and C.J. Whitaker, Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy, Machine Learning **51** (2003), 181–207.
[13] A. Aarabi, F. Wallois and R. Grebe, Automated neonatal seizure detection: A multistage classification system through feature selection based on relevance and redundancy analysis, Clinical Neurophysiology **117** (2006), 328–340.
[14] B.R. Greene, S. Faul, W.P. Marnane, G. Lightbody, I. Korotchikova and G.B. Boylan, A comparison of quantitative EEG features for neonatal seizure detection, Clinical Neurophysiology **119** (2008), 1248–1261.
[15] C. Yamaguchi, Fourier and Wavelet analyses of normal and epileptic electroencephalogram (EEG), Conference Proceedings First International IEEE EMBS Conference on Neural Engineering, Capri Island, Italy, 2003, pp. 406–409.
[16] S.S. Lodder and M.J. van Putten, Quantification of the adult EEG background pattern, Clinical Neurophysiology **124** (2013), 228–237.
[17] S. Hu, C. Xu, W. Guan, Y. Tang and Y. Liu, Texture feature extraction based on wavelet transform and gray-level co-occurrence matrices applied to osteosarcoma diagnosis, Biomedical Materials and Engineering **24** (2014), 129–143.
[18] Y. Saeys, I. Inza and P. Larrañaga, A review of feature selection techniques in bioinformatics, Bioinformatics **23** (2007), 2507–2517.
[19] L. Sun and J. Xu, Feature selection using mutual information based uncertainty measures for tumor classification, Biomedical Materials and Engineering **24** (2014), 763–770.
[20] J. Xu, L. Sun, Y. Gao and T. Xu, An ensemble feature selection technique for cancer recognition, Biomedical Materials and Engineering **24** (2014), 1001–1008.
[21] J.A. Urigüen and Begoña García Zapirain, EEG artifact removal–state-of-the-art and guidelines, Journal of Neural Engineering **12** (2015), 031001.