Improving Classification of Epileptic and Non-Epileptic EEG Events by Feature Selection

* Evangelia Pippa[1], Evangelia I. Zacharaki[1], Iosif Mporas[1]
Vasiliki Tsirka[2], Mark P. Richardson[2], Michael Koutroumanidis[2] and Vasileios Megalooikonomou[1]
[1]Multidimensional Data Analysis and Knowledge Management Laboratory
Dept. of Computer Engineering and Informatics, University of Patras
26500 Rion-Patras, Greece
[2]Dept. of Clinical Neurophysiology and Epilepsies
Guy's & St. Thomas' and Evelina Hospital for Children, NHS Foundation Trust/ King's College,
London, United Kingdom
*Corresponding Author's email: pippa@ceid.upatras.gr

Abstract: Correctly diagnosing generalized epileptic from non-epileptic episodes, such as psychogenic non epileptic seizures (PNES) and vasovagal or vasodepressor syncope (VVS), despite its importance for the administration of appropriate treatment, life improvement of the patient, and cost reduction for patient and healthcare system, is rarely tackled in the literature. Usually clinicians differentiate between generalized epileptic seizures and PNES based on clinical features and video-EEG. In this work, we investigate the use of machine learning techniques for automatic classification of generalized epileptic and non-epileptic events based only on multi-channel EEG data. For this purpose, we extract the signal patterns in the time domain and in the frequency domain and then combine all features across channels to characterize the spatio-temporal manifestation of seizures. Several classification algorithms are explored and evaluated on EEG epochs from 11 subjects in an inter-subject cross-validation setting. Due to large number of features feature ranking and selection is performed prior to classification using the ReliefF ranking algorithm within two different voting strategies. The classification models using feature subsets, achieved higher accuracy compared to the models using all features reaching 95% (Bayesian Network), 89% (Random Committee) and 87% (Random Forest) for binary classification (epileptic versus non-epileptic). The results demonstrate the competitiveness of this approach as opposed to previous methods.

Keywords: epileptic seizures, PNES, vasovagal syncope, classification, machine learning.

1. Introduction

One of the most common and challenging medical cases in everyday clinical practice is that of patients reporting one or more episodes of paroxysmal loss of consciousness or altered awareness. The management of these medical cases may be proven to be demanding, time consuming and expensive and finally, in spite of the extensive and exhaustive investigation, the underlying diagnosis may remain elusive [1,2,3]. The differential diagnosis that a clinician usually faces is mainly that of an epileptic seizure, a possible psychogenic non epileptic seizure (PNES) and a probable vasovagal syncope (VVS).

The diagnosis of epilepsy and its differentiation from other causes of TLoC is typically based on historical information and is assisted by specific tests [2]. However clinical information is commonly fragmented or even missing because patients may have limited or no recall of the event and a witness account might not be available to describe diagnostically decisive clinical phenomena [1,2]. Even when a witness is available, diagnosis may be difficult and often remains uncertain because convulsive syncope, a seizure-like reaction resulting from global cerebral hypoperfusion, can mimic epileptic seizures [3,4]. Agreement between physicians as to the nature of a single event may also be limited [5]. Such diagnostic uncertainty has cost both in terms of mortality and ongoing morbidity and in terms of the financial burden associated with hospitalization and repeated investigations.

Epileptic seizures are brief episodes of abnormal excessive or synchronous neuronal activity in the brain of patients suffering from epilepsy [6]. During an epileptic seizure there are several specific changes recorded in the electroencephalogram (EEG) which is a sensitive and important test used to evaluate patients with suspected epilepsy. There are certain characteristic ictal neurophysiological patterns that support the identification and detection of epileptic events and postictal and/or interictal abnormalities that can provide supplementary information. Fig. 1 shows generalized spike wave abnormalities from an epileptic patient. Specifically, there is a burst of generalized 3-5 Hz spike and slow wave complex lasting approximately 5 secs.

FIGURE 1

Pshychogenic non-epileptic seizures (PNES) are sudden paroxysmal changes in behavior or consciousness, that resemble epilepsy but are not accompanied by the electrophysiological changes that characterize an epileptic seizure [7]. Although the clinical history can help differentiate these episodes,

it is not unlikely to have inconclusive and insufficient event description by the patient and witnesses, not being able to confidently exclude an underlying epileptic disorder. In these cases the diagnosis of PNES can be supported by video-EEG monitoring, especially if a psychogenic event is captured, since in the case of PNES there are no specific EEG changes. Fig 2 shows an EEG fragment during a PNES. No EEG correlates can be seen and the recording is frequently marred by muscular artifacts.

FIGURE 2

Vasovagal or vasodepressor syncope is a common type of syncope and various mechanisms have been postulated for explaining the characteristic association of hypotension and bradycardia. The term "vasovagal" was introduced by Lewis [8] to indicate that both blood vessels and heart were implicated and since atropine reversed the bradycardia but not the hypotension he considered vasodilatation as the primary responsible factor. During a vasovagal syncopal attack there may be some characteristic EEG changes starting with progressive generalized theta slowing of background rhythms followed by sometimes hypersynchronous delta activity of high voltage, (beta / alpha → theta → delta) and appearance of progressively lower voltage rhythms until isoelectric suppression [9,10] (see Fig. 3). This pattern is progressively reversed after the patient's fall, during his/her recovery. These changes do not include any ictal activity.

FIGURE 3

Several methods have been proposed for the classification of EEG captured events into epileptic or normal [11,12,13,14,15]. The problem of the discrimination between ictal and interictal EEG signals has been studied [16], too. However, only a few studies deal with the differentiation between epileptic and other paroxysmal episodes of loss of consciousness such as PNES and vasovagal syncope. It is worth to note that the discrimination between different types of non-epileptic events is considerably more useful in diagnostic procedure given the semiological resemblance between the aforementioned paroxysmal attacks. Furthermore, according to [7] the one third of PNES patients may have clinical convincing GrandMal like seizures. This makes discrimination between PNES and epileptic seizures a challenging task, especially in an online monitoring system for automatic detection of epileptic events, such as [17], where false alarms caused by events similar to epilepsy are undesired.

To the best of our knowledge, only a few studies have been proposed in the literature for automated classification between epileptic and non-epileptic pathological events from EEG. Poulos et al. [18] proposed an algorithm which estimates a number of auto-correlated coefficients extracted from

an appropriately selected epileptic EEG segment and examines whether these coefficients are correlated with the coefficients of the unknown EEG segments in order to classify the latest into epileptic or non-epileptic. Their algorithm obtained a sensitivity of 83% for 90% specificity. Papavlasopoulos et al. [19] trained a LVQ1 neural network on an appropriately extracted set of auto-correlation coefficients (codebook) and used the resulting model to classify the corresponding feature vectors of the unknown EEG segments. The LVQ1 network achieved 86% accuracy. The feature extraction methods of the aforementioned classification frameworks, as well as the achieved results, can be found in [20]. Statistical analysis of the results based on chi-square test showed that the LVQ neural network method is superior than the cross-correlation one [20].

Regarding the features used for the classification of EEG segments the relevant works in the literature are considerably more. In the majority of them, the analysis is based on the estimation of the EEG channels' spectral magnitude [11, 15, 21, 22]. Other EEG features that have been reported are the autoregressive filter coefficients, the continuous and discrete wavelet transform, as well as energy per brain wave (delta, theta, alpha, beta, gamma) bands [15,21, 23]. Finally, time domain features have been proposed, such as zero-crossing rate [24] and statistics of the EEG samples per channel [15,21].

In this study, we evaluate a large set of time and frequency domain features which have been widely used for the analysis of EEG signals in the literature. In addition to the reported evaluations found in the literature, we extend the non-epileptic class to both PNES and VVS events. The diagnosis of epilepsy is more challenging compared to the detection of seizure onset due to the semiological resemblance between epileptic and non-epileptic events, especially when video-EEG monitoring is not incorporated [25]. Also, the classification of abnormal episodes into different types requires a broad knowledge of EEG patterns across patients, while seizure detection can rely on patient-specific models which are easier to learn, especially for generalized seizures [26]. For the evaluation, we examined a number of different classification algorithms. Our classification methodology can be used as part of our previous seizure detection architecture [26,27] in order to discriminate the detected events into epileptic or non-epileptic.

In a further step, feature ranking investigation using two different strategies (one based on frequency of feature appearing in a specific rank and the other based on sum of the weights assigned by the ReliefF ranking algorithm) was performed. The classification models using subsets of N best features were evaluated and revealed the most significant features for the classification task.

The rest of this paper is organized as follows. In Section 2 the classification methodology is presented and details about the evaluation data are provided. Section 3 describes the experimental protocol followed and presents the achieved results. Finally, in Section 4 we conclude this work.

2. Material and methods

2.1 Methodology for classification of generalized epileptic and non-epileptic events

The presented architecture for classification between generalized epileptic and non-epileptic EEG events is part of an end-to-end system for monitoring and analysis of brain disorders, the ARMOR framework [17]. Within the ARMOR framework patients suffering from seizures are monitored through sensors and the multi-parametric data are processed automatically (real-time by software tools) or semi-manually (offline with the support of software tools and visualizations) by neurology experts [26,27].

The proposed classification methodology can be used as additional module after the seizure detection and focal-vs-generalized events classification components [26,27] in order to discriminate the detected events into generalized epileptic, manifested by Generalized Spike Wave discharges (GSW) or non-epileptic, such as PNES and VVS. The block diagram of the overall architecture is illustrated in Figure 4.

FIGURE 4

Initially, the multidimensional EEG data are preprocessed by applying notch filtering (at 50Hz), baseline correction and re-sampling at 250 Hz (in order to obtain a common resolution level for all data coming from different patients and acquisition systems). Frame blocking of the incoming EEG streams to epochs of constant length w (= 2 seconds) is performed with constant time-shift and without time-overlap between successive epochs. Each epoch is a $N \times w$ matrix, where N is the number of selected EEG electrodes. A large number of features is extracted for each one of the N electrodes to characterize the temporal patterns and frequency content of each epoch. The extracted time domain and frequency domain features from all electrodes are concatenated to a single feature vector as a representative signature for each epoch. Details on the type of extracted features are provided in section 2.3.

All epochs are used as input to ARMOR's seizure detection module which detects paroxysmal events. The epochs classified as normal are ignored whereas epochs classified as seizure are further

entered to the seizure type classifier. In this final step, models for binary classification between generalized epileptic or non-epileptic events (PNES or VVS), which have been previously built in a training phase, are used in order to label the epochs. Each epoch is classified independently and no temporal constraints (across epochs) are applied, such as taking into consideration the class label of the precedent or subsequent epoch or the total event duration.

During the training phase of the classification architecture, epochs with known class labels (labeled manually by medical experts) are used to train binary classification models, i.e. generalized epileptic (GSW) vs non-epileptic (PNES and VVS).

During the test phase the unknown multidimensional EEG signal is preprocessed and parameterized with similar setup as in the training phase. Each extracted feature vector is provided as input to the seizure detector and according to the decision to the seizure classifier.

2.2 Data

The previously described classification methodology was evaluated on multi-parametric recordings performed within the ARMOR project [17]. The recordings were performed in the Department of Clinical Neurophysiology and Epilepsies in St Thomas' Hospital in London and acquired from 11 patients in total. All participants had at least one of their typical epileptic or non epileptic events captured during the recording procedure. The epileptic group, consisted of patients with known diagnosis of idiopathic generalized epilepsy, manifested clinically with absence seizures and they had at least one clinical episode captured during the recording associated with generalized spike wave discharges on the EEG. The non epileptic group included patients who had sustained a vasovagal syncope (2 participants) or a psychogenic non epileptic attack (5 participants) during their monitoring. The epilepsy group contains 105 generalized seizures while the non-epilepsy groups include 21 events (19 PNES and 2 VVS). Patients with focal seizures were excluded from this analysis.

The recordings were performed using conventional AgCl EEG electrodes positioned according to the extended international 10–20 system. A subset of the main EEG channels was selected for analysis which included the following channels: Fp2, F8, F4, T4, C4, A2, P4, T6, O2, Fp1, F7, F3, A1, C3, T3, P3, T5, O1, Fz, Cz, Pz. The recordings were manually annotated by expert Neurologists of the King College London. Only epochs during paroxysmal events were considered for training and for testing. All data were stored in EDF+ formatted files [28].

2.3 Feature Extraction and Classification Algorithms

After preprocessing, time domain and frequency domain features were extracted for each epoch. In particular, each of the EEG channels was parameterized using the following features: (i) time-domain features: minimum value, maximum value, mean, variance, standard deviation, percentiles (25%, 50%-median and 75%), interquartile range, mean absolute deviation, range, skewness, kyrtosis, energy, Shannon's entropy, logarithmic energy entropy, number of local maxima and local minima, zero-crossing rate, and (ii) frequency-domain features: 6-th order autoregressive-filter (AR) coefficients, power spectral density, frequency with maximum and minimum amplitude, the power of continuous wavelet transform using symlet 5 mother wavelet of scale 25 and 32, the power of discrete wavelet transform with mother wavelet function Daubechies 16 and decomposition level equal to 8. This resulted to 55 variables for each of the N=21 EEG channels producing a feature vector of dimensionality equal to 1155 in total.

The computed feature vectors, V, were used to train classification models. In order to evaluate the ability of the above features to discriminate between epileptic and non-epileptic epochs we examined several classification algorithms, including BayesNet [29,30], RandomCommittee, Random Forest [31], IBk [32] and SMO [33,34] with RBF kernel, which were implemented by the WEKA machine learning toolkit [35]. The classifiers in our study were selected in an attempt to evaluate representative algorithms for each one of the main categories of machine learning classification methods including probabilistic networks (BayesNet), decision trees (RandomForest), support vector machines (SMO), ensemble classifiers (RandomCommittee and RandomForest) but also simple methods such as k nearest neighbors (IBk).

During the test phase, the EEG recordings were pre-processed and parameterized as during training. Each classification model was used to label each of the detected seizure epochs. In the present evaluation no additional rules (e.g. knowledge based rules regarding events duration) were applied on the classification decision .

Evaluation was performed in a leave-one-out cross-validation setting. Specifically, each time one subject was left-out for testing, while the rest of the subjects were used for training. For the left-out subject, all epochs between seizure onset and offset were used as testing samples. Table 1 shows the number of epochs that were extracted from each subject during the seizure(s).

TABLE 1

The purpose of this study was to evaluate the seizure classification module, thus only paroxysmal events were used for training and testing of the classifiers. Evaluation of the total ARMOR framework may include the combined use of seizure detection and seizure classification in future work.

2.4 Feature Ranking and Feature Subsets Evaluation

In a further step we examined the discriminative power of the extracted features for the classification of epileptic and non-epileptic EEG events. The ReliefF algorithm [36] (which is an extension of an earlier algorithm called Relief [37]) was used for estimating the importance of each feature in binary classification (generalizing to polynomial classification by decomposition into a number of binary problems). In the ReliefF algorithm the weight of any given feature decreases if the squared Euclidean distance of that feature to nearby instances of the same class is more than the distance to nearby instances of the other class. ReliefF is considered one of the most successful feature ranking algorithms due to its simplicity and effectiveness [38, 39,40] (only linear time in the number of given features and training samples is required), noise tolerance and robustness in detecting relevant features effectively, even when these features are highly dependent on other features [38,41]. Furthermore, ReliefF avoids any exhaustive or heuristic combinatorial search compared with conventional wrapper methods and usually performs better compared to filter methods due to the performance feedback of a nonlinear classifier when searching for useful features [40].

In this study, ranking is performed by following a leave-one-out strategy on the available subjects. Specifically, for each leave-one-out experiment, feature ranking is performed using the ReliefF algorithm in each training subset. We combine the rankings of all leave-one-out experiments and calculate the total rank of features using two different strategies. The first strategy calculates the total rank of features according to the frequency of a feature appearing in a specific rank. For example the top-ranked feature is assumed to be the one that more frequently has the highest ranking score, regardless of the distribution of the scores it receives across experiments. The second strategy calculates the total rank of features according to the sum of the weights assigned by ReliefF in each training set. We examined the performance of the method, in terms of accuracy, sensitivity and specificity, for different number of N-best features (N =10, 20, 30, ...100, 200, 300, ... ,1100), with respect to the above strategies of feature ranking.

3. Results and Discussion

The classification methodology presented in Section 2.1 was evaluated using the classification algorithms and the cross-validation scheme described in Section 2.3. The accuracy, sensitivity and specificity are defined as:

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (2)$$

$$sensitivity = \frac{TP}{TP+FN} \quad (3)$$

$$specificity = \frac{TN}{FP+TN} \quad (4)$$

where TP denotes the true positives, TN the true negatives, FP the false positives and FN the false negatives. The results of the method using all features are shown on the left of Table 2. Here we consider the epileptic class as the positive and the non-epileptic class (PNES or VVS) as the negative.

TABLE 2

As can be seen in Table 2, the overall highest accuracy of the proposed methodology for classification between epileptic and non-epileptic EEG events is 86% for BayesNet classification. RandomCommittee and Random Forest classification models follow with 83% and 74% accuracy, respectively. For the classifier with the highest accuracy (BayesNet), the sensitivity (or recall), i.e. the fraction of actual epileptic events which are correctly identified as such, is 92% and the specificity, i.e. the proportion of non-epileptic events (either PNES or VVS) which are correctly classified as such, is 78%.

In a further step, we applied feature ranking using the ReliefF algorithm and the two strategies described in Section 2.4. The performance of the classification, in terms of accuracy, for different number of N-best features (N =10, 20, 30 ..., 100, 200, 300, ... ,1100) and for each algorithm separately are shown in Fig. 5 for the 1st ranking strategy and in Fig. 6 for the 2nd ranking strategy.

FIGURE 5

FIGURE 6

As can be seen in the above figures the highest classification accuracy is achieved when a small subset of discriminative features are used. Specifically, when the 1st ranking strategy is used the highest accuracy is achieved for a subset of 10 best features with a percentage of 95% for the Bayesian

Network, which is sufficiently high in comparison to the accuracy achieved when all features are used. Similarly, Random Committee and Random Forest achieve their highest accuracies for a subset of 300 and 200 best features respectively. The reported accuracies for these subsets of features are 92% and 87% for each algorithm respectively. IBk and SMO follow with an accuracy of 86% when a subset of 40 best features is used and 87% for a subset of 200 best features, respectively.

The 2nd ranking strategy shows similar behavior. For the Bayesian Network the highest accuracy (94%) is achieved for a subset of 50 best features. Random Committee, Random Forest, IBk and SMO follow with 85% for a subset of 70 best features, 89% for a subset of 50 best features, 84% for a subset of 70 best features and 90% for a subset of 60 best features, respectively.

In general, Random Forest and Random Committee seem to be the less stable algorithms, SMO on the other hand, although not the most accurate classifier, shows a more stable behavior as function of number of retained features, with the accuracy decreasing significantly for more than 200 features.

Tables 3 and 4 show the 50 best features according to the ranking strategy 1 and 2 respectively.

TABLE 3

TABLE 4

As can be seen, in general the two ranking strategies overall agree. Both of them rank features *nmin* (number of local minima), *nmax* (number of local maxima), *aryule3* (the 3rd coefficient of 6th order autoregressive filter), *minfreq* (frequency with minimum power), *cwt25* and *cwt32* (the coefficients of continuous wavelet transform using symlet 5 mother wavelet of scale 25 and 32) in the top 50 features.

The number of local minima (*nmin*) and the number of local maxima (*nmax*) seem to be the features with the highest discriminative ability. Since these features measures the smoothness of the signal it seems that the smoothness of the epileptic epochs is different from the one of non-epileptic epochs and aids the discrimination among them. Such a claim can be verified from the distributions of the values of the *nmin* (see Fig. 7) and *nmax* (see Fig. 8) features for the epileptic and non-epileptic class.

FIGURE 7

FIGURE 8

In both figures 7 and 8, the blue boxes indicate the distribution of the feature values on the epileptic class and the black boxes the distribution of the feature values on the non-epileptic class. As can be seen, there is a perfect discrimination between the epileptic and non-epileptic main boxes with the non-

epileptic epochs having a significantly larger number of local minima and maxima indicating less smooth signal compared to the generalized spike waves. The only overlap is observed between the extreme values of two classes (whiskers of the boxplots). The evaluation of our framework using only these two features (*nmin* and *nmax*) extracted from all the available channels resulted in 91% accuracy for 92% sensitivity and 89% specificity. The performance in terms of accuracy increases slightly when *nmin* and *nmax* are extracted from the 5 best channels (Fp1, T4 , T5, F7, Fp2), reaching 92%. This increase indicates that the frontotemporal regions in the brain covered by the aforementioned channels might be more important in discriminating generalized spike waves from PNES or VVS. The next most important features for discriminating epileptic from non-epileptic events are *aryule3*, *minfreq* and *cwt25* and *cwt32*. The autoregressive model specifies whether the EEG epoch depends linearly on its own previous values by expressing the signal with lagged terms of itself. In particular, the AR model residual (i.e. the prediction error) shows how possible is to model each sample as a linear combination of its previous ones. The lower absolute values of the AR coefficients of the non epileptic class (see Fig. 9 for the *aryule3* feature values) indicates that the signal of the non-epileptic class is much more noisy and stochastic-like compared to the epileptic signals which seem to be more structured and deterministic-like. Such an experimental result is consistent with our intuition about the two types of signals and the clinicians description of the events.

FIGURE 9

Differentiation is also observed on the frequency with the minimum power (*minfreq*) in the spectrogram of epileptic and non-epileptic epochs (see Fig. 10), with the *minfreq* of the epileptic class having a much greater range of values compared to the non-epileptic class in which the *minfreq* values are clustered around 50Hz. Note that this finding is not due to notch filtering since the same preprocessing was applied to all data (both epileptic and non-epileptic).

FIGURE 10

Finally, the expression of each epoch as a linear combination of the chosen wavelet basis functions captures the frequency content of the epoch in a localized area of the signal which seem to highlight the differences between the two classes (see Fig. 11 and 12).

FIGURE 11

FIGURE 12

Finally, in order to examine the ability of the BayesNet classifier to discriminate each type of pathological events (GSW, PNES or VVS) from the others, we performed binary classification of all possible pairs of pathological events (GSW-PNES, GSW-VVS and PNES-VVS). The results in terms of classification accuracy for different number of N-best features (10, 20, ..50) are shown in Fig. 13.

FIGURE 13

As can be seen, the PNES-VVS classification problem is the most difficult case for the classifier. The best classification accuracy (76%) for PNES-VVS pair is achieved when all features (1155) are used. On the other hand, GSW-PNES and GSW-VVS pairs are much easier cases for the classifier obtaining their maximum accuracy for the 10 best features. Specifically, GSW-PNES classification achieves 96% accuracy for 96% sensitivity and 100% specificity while GSW-VVS classification results in slightly lower percentages, i.e. 93% accuracy for 96% sensitivity and 87% specificity. Since generalized spike waves are very specific ictal neurophysiological patterns, they present much more consistent features (compared to the other types) which makes their detection an easier task. On the other hand, PNES has no specific EEG patterns but is frequently accompanied by muscular artifacts which present a variability across subjects. Similar variability appears even between consecutive epochs of VVS examples since there are several changes that happen successively in time during such an episode (beta / alpha $\rightarrow$ theta $\rightarrow$ delta $\rightarrow$ lower voltage rhythms $\rightarrow$ isoelectric suppression). It seems that the variability in the feature values of the PNES and VVS epochs is high (in respect to the available training data) impeding the learning of a discrimination model.

While 19 PNES appear to make a rather limited dataset, we believe that are sufficient given the lack of ictal EEG changes and the fact that their variability reflects only muscle and movement activities. The main problem is the really small sample of the 2 VVS-patients. However, VVS typically occur very rarely, in most patients annually, and only in very few patients more frequently, say monthly. It is therefore extremely unlikely to record them on standard EEG that is a 20min to one hour "snapshot" of brain activity. Still, because of the rather predictable sequence of EEG changes (alpha-theta delta etc) we believe that reasonable learning of a discrimination model is achievable / possible.

The proposed methodology takes into account features extracted from all the available channels by concatenating them in a single feature vector. The spatial localization of the features is encoded in their location within the feature vector presented to the classifier. Since the seizure onset patterns in focal seizures appear over a small subset of channels close to or at the epileptic focus, a strategy to

overcome the problem that different focal seizures appear on different channels is required. Such a strategy that successfully tackles the aforementioned problem has already been proposed in the literature [42]. In order to remove the information about the spatial location of the seizure from the training set, the authors in [42] proposed a sorting operation on the extracted features that reorders the features from the different channels in the feature vector before feeding it to the classifier.

However, since a seizure with focal onset (as manifested electroencephalographically) is always epileptic, here we have implemented a simplified version of a focal-vs-generalized seizure classification rule (as part of the ARMOR project) that automatically detects and labels the focal seizures. The focal-vs-generalized seizure classification rule is part of the online seizure detector [26][27] which performs a per channel analysis followed by the imposition of spatiotemporal constraints before taking the final decision (clear, focal, generalized) for each tested epoch. The classification rule is based on a mimetic approach requiring the seizure to be detected in at least 65% of the channels in order to be characterized as generalized; otherwise it is characterized as focal. Due to the different type of analysis (fusion of channel-based decisions versus fusion of features per channel to reach a decision), we are not presenting results of focal seizure classification in this paper, but rather focus on the classification of generalized events, which is the last component in our seizure analysis framework.

For a clinician the differentiation between focal and generalized events is important because it will play a crucial role in the medication/ treatment and general management choices. Such a rule (appearance in at least 65% of channels) is mainly useful when the events are focal, since focal EEG onset always indicates focal epileptic seizure activity. This rule has no clinical utility to the other event types, since the EEG expression of both psychogenic non-epileptic seizures and vasovagal syncope which leads to impairment of consciousness are “generalized”. However, this step was introduced to facilitate the solution methodologically. Upon the characterization of focal events, the method presented here can be used to discriminate the remaining events into epileptic or non-epileptic.

On the clinical usefulness front, it is true that a competent seizure detection algorithm or set of algorithms should be able to detect both focal and generalized seizures, and either of these from non-epileptic events. The reason is that impairment of consciousness can be seen in temporal lobe seizures or seizures with secondary generalization. However, initial prodromal clinical symptoms and some typical EEG characteristics can be used for the differential diagnosis. Due to the big variability of

seizure presentation there needs to be a detailed analysis of adequate number of representative cases of different evolving patterns and this will be the part of our next work.

Until however we will be able to evaluate the method more extensively on a large dataset with an adequate number of representative cases for focal seizures we applied the proposed methodology on a dataset of 9 patients (2 subjects with focal seizures, 5 subjects with PNES and 2 subjects with VVS). We developed also a different algorithm to remove the spatial content from the features. The algorithm sorts the features for each channel according to feature type and then extracts the standard deviation of each feature type across channels, and the difference between maximum and minimum values of each feature (max-min). We introduced the standard deviation and the max-min values to the BayesNet classifier and achieved 74% accuracy, 70% sensitivity and 76% specificity when the 20 best features are used. The above results were obtained with a leave-one-patient-out strategy for validation. Since the dataset size is small we assessed the method also in a leave-one-epoch-out strategy, as performed in some other studies and achieved 90.2% accuracy, 86.7% sensitivity and 91.5% specificity. The achieved accuracy in this case is much higher, as expected. However, we do not emphasize the importance of these results since they might not generalize to other data.

Although direct comparison with other studies is not possible due to the different characteristics of each dataset (e.g. different seizure types, lack of PNES or VVS examples in most studies or use of single channel data), the achieved classification accuracy is higher than the one reported in the literature. In particular, the achieved accuracy in [22] is 86%, lower than the accuracy of BayesNet in our methodology (95%). Furthermore, in [21] the reported sensitivity (83%) and specificity (98%) are lower than the sensitivity of the majority of the classification methods evaluated in our work and the specificity achieved by our framework (98%) when a subset of 10 discriminative features are used with respect to BayesNet classification.

Finally, although an initial work was held to reliably solve the problem of discrimination between different types of paroxysmal event and reveal the most discriminative features from a large set of time and frequency domain features given a dataset of 11 patients, there are some limitations that should be taken into account. The number of non-epileptic examples especially those of VVS) is limited and might not capture well the variability of the corresponding EEG events while the available generalized spike waves seem to be enough to describe such a consistent group of patterns. Under this scope we plan to start EEG recordings during tilt table test, which provokes VVS and therefore we

shall have a substantial number for further analysis. Furthermore, we aim to perform a more in depth analysis of focal seizures.

4. Conclusions

In this paper, we investigated the problem of classification between epileptic and non-epileptic events from multi-channel EEG data using a large number of time-domain and frequency domain features. The proposed methodology was evaluated in EEG data from 11 subjects. Examination of several classification algorithms showed that the best accuracy is achieved by BayesNet. Feature ranking investigation and evaluation of the classification models using subsets of features were performed and revealed the most significant features for the classification task. The use of the most discriminative features ($N = 10$) increased significantly the performance of BayesNet classification at 95% accuracy (94% sensitivity for 98% specificity). The method has been evaluated using cross-validation across subjects and showed that it can generalize satisfactorily providing the means for diagnosis support.

5. Acknowledgement

This study is partially funded by the EC under the FP7/2007-2013 with grant ARMOR, Agreement Number 287720. This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the NSRF - Research Funding Program: Thales. Investing in knowledge society through the European Social Fund.

6. References

[1] Al. McKeon, C. Vaughan and N. Delanty, “Seizure versus syncope,” vol. 5, Lancet Neurol., 2006, pp. 171-80.

[2] F. Brigo, R. Nardone and L.G. Bongiovanni, “Value of tongue biting in the differential diagnosis between epileptic seizures and syncope,” vol. 21, Seizure, 2012, pp. 568-572.

[3] J.T. Lin, D.K. Ziegler, C.W. Lai, W. Bayer, “Convulsive syncope in blood donors,” vol. 11, Annals of Neurology, 1982, pp. 525–28.

[4] W. Wieling, R.D. Thijs, N. van Dijk, A.A. Wilde, D.G. Benditt and J.G. van Dijk, " Symptoms and signs of syncope: a review of the link between physiology and clinical clues," vol. 132, Brain, 2009, pp. 2630–2642.

[5] W.A.J. Hoefnagels, G.W. Padberg, J. Overweg and R.A.C. Roos, "Syncope or seizure? A matter of opinion," vol. 94, Clinical Neurology and Neurosurgery, 1992, pp.153–156.

[6] R. S. Fisher, W. V. E. Boas, W. Blume, C. Elger, P. Genton, P. Lee and J. Engel, "Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE) ," vol. 46, Epilepsia, 2005, pp. 470-472.

[7] L. Szabo, "Semiologic analysis and classification of childhood psychogenic non-epileptic seizures based on Video-EEG monitoring data," PhD thesis, 2013.

[8] T. Lewis, "Vasovagal syncope and the carotid sinus mechanism," vol. 1, The British Medical Journal, 1932, pp. 873-876.

[9] F. Ammirati, F. Colivicchi, G. Di Battista G, F.F. Garelli, C. Pandozi and M. Santini, "Variable cerebral dysfunction during tilt induced vasovagal syncope," vol. 21, Pacing and Clinical Electrophysiology, 1998, pp. 2420-5.

[10] O. Mecarelli, P. Pulitano, E. Vicenzini, N. Vanacore, N. Accornero and M. De Marinis, "Observations on EEG patterns in neurally-mediated syncope: an inspective and quantitative study," vol. 34, Clinical Neurophysiology, 2004, pp. 203-207.

[11] B. Greene, G. Boylan, R. Reilly, P. Chazal and S. Connolly, "Combination of EEG and ECG for improved automatic neonatal seizure detection," vol. 118, Clinical Neurophysiology, 2007, pp. 1348-1359.

[12] Meier, H. Dittrich, A. Schulze-Bonhage and A. Aertsen, "Detecting epileptic seizures in long-term human EEG: a new approach to automatic online and real time detection and classification of polymorphic seizure patterns," vol. 25, Clinical Neurophysiology, 2008, pp. 119-131.

[13] A. Shoeb, H. Edwards, J. Connolly, B. Bourgeois, S.T. Treves and J. Guttag, "Patient-specific seizure onset detection," vol. 5. Epilepsy Behavior, 2004, pp. 483-98.

[14] A. Shoeb and J. Guttag, "Application of machine learning to epileptic seizure onset detection," 27th International Conference on Machine Learning (ICML), 2010.

[15] M. Valderrama, S. Nikolopoulos, C. Adam, V. Navarro and M. Van Quyen, “Patient-specific seizure prediction using a multi-feature and multi-modal EEG-ECG classification,” In Proc. of IFMBE, 2010.

[16] M.Z. Parvez and M. Paul, “Epileptic seizure detection by analyzing EEG signals using different transformation techniques,” vol. 145, Neurocomputing, 2014, pp.190-200.

[17] ARMOR project. Web-site: http:// www.armor-project.eu/.

[18] M. Poulos, F. Geogiacodis, V. Chrissicopoulos and A. Evangelou, “Diagnostic test for the discrimination between interictal epileptic and non-epileptic pathological EEG events using auto-cross-correlation methods,” vol. 43, American Journal of Electroneurodiagnostic Technology, 2003, pp. 228-264.

[19] S. Papavlasopoulos, M. Poulos and A. Evangelou, “Feature extraction from interictal epileptic and non-epileptic pathological EEG events for diagnostic purposes using LVQ1 neural network,” Proceedings of 7th International Conference on Mathematics Methods in Scattering Theory and Biomedical Technology, BIOTECH'7, 2005.

[20] S. Papavlasopoulos, M. Poulos, G. Bokos and A. Evangelou, “Classification control for discrimination between interictal epileptic and non-epileptic pathological EEG events,” vol. 1, International Journal of Medical World Academy of Science, Engineering and Technology, Health, Pharmaceutical and Biomedical Engineering, 2007.

[21] H.R. Mohseni, A. Maghsoudi and M.B. Shamsollahi, “Seizure detection in EEG signals: a comparison of different approaches,” In Conf. Proc IEEE Eng Med Biol Soc. ,2006, 6724-7.

[22] S. Nasehi and H. Pourghassem , “ Seizure Detection Algorithms Based on Analysis of EEG and ECG Signals: a Survey,” vol. 44, Neurophysiology, 201, pp. 174-186.

[23] S. Tong and N.V. Thankor , “ Quantitative EEG Analysis Methods and Applications,” Engineering in Medicine & Biology, 2009.

[24] B. Hunya, M. Signoretto, W. Van Paesschen, J.A. Suykens, S. Van Huffel and M. De Vos, “Incorporating structural information from the multichannel EEG improves patient-specific seizure detection,” vol. 123, Clinical Neurophysiology , pp. 2352-2361.

[25] E. Asanoa, C. Pawlak, A. Shah, J. Shah, A.F. Luat, J. Ahn-Ewing, H.T. Chugani, “The diagnostic value of initial video-EEG monitoring in children—Review of 1000 cases,” Epilepsy Research 66, 2005, pp. 129–135.

[26] I. Mporas, V. Tsirka, E. Zacharaki, M. Koutroumanidis and V. Megalooikonomou, “Online Seizure Detection from EEG and ECG signals for Monitoring of Epileptic Patients,” 8th Hellenic Conference on Artificial Intelligence (SETN), Lecture Notes in Computer Science, vol. 8445, 2014, pp 442-447.

[27] I. Mporas, V. Tsirka, E.I. Zacharaki, M. Koutroumanidis and V. Megalooikonomou, “Evaluation of Time and Frequency Domain Features for Seizure Detection from Combined EEG and ECG signals,” 7th Int. Conf. on Pervasive Technologies Related to Assistive Environments (PETRA), 2014.

[28] B. Kemp and J. Olivan, “European data format 'plus' (EDF+), an EDF alike standard format for the exchange of physiological data,” vol. 114, Clinical Neurophysiology, 2003 pp. 1755–1761.

[29] G.F. Cooper and E. Herskovits, “A Bayesian method for constructing Bayesian belief networks from databases,” 1990.

[30] G. Cooper and E. Herskovits, “A Bayesian method for the induction of probabilistic networks from data, ” vol. 9, Machine Learning, 1992, pp. 309-347.

[31] L. Breiman, “ Random Forests, ” vol. 45, Machine Learning, 2001, pp.5-32.

[32] D. Aha and D. Kibler, “Instance-based learning algorithms,” vol. 6, Machine Learning, 1991, pp. 37-66.

[33] S.S. Keerthi, S.K. Shevade, C. Bhattacharyya and K.R.K. Murthy, “Improvements to Platt's SMO algorithm for SVM classifier design,” vol. 13, Neural Computation, 2001, pp. 637-649.

[34] J. Platt, “Fast Training of Support Vector Machines using Sequential Minimal Optimization, ” Advances in Kernel Methods - Support Vector Learning, 1998.

[35] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten, “The WEKA Data Mining Software: An Update,”, vol. 11, SIGKDD Explorations, 2009.

[36] I. Kononenko, “Estimating attributes: Analysis and extension of RELIEF,” vol. 784, Proc. Euro. Conf. Mach. Learn., 1994, pp. 171– 182.

[37] K. Kira and L. A. Rendell, “A practical approach to feature selection,” Proc. 9th Int. Conf. Mach. Learn., 1992, pp. 249 – 256.

[38] T. G. Dietterich, “Machine learning research: Four current directions,” vol. 18, Artificial Intelligence Magazines, 1997, pp. 97–136.

[39] Y. Sun and J. Li, “Iterative RELIEF for feature weighting,” Proc. 21st Int. Conf. Mach. Learn., 2006, pp. 913–920.

[40] Y. Sun and D.Wu, “A RELIEF based feature extraction algorithm,” In Proceedings of SIAM International Conference on Data Mining, 2008.

[41] I. Kononenko, E.Simec and M. Robnik-Sikonja, “Overcoming the Myopic of Inductive Learning Algorithms with RELIEF-F,” Applied Intelligence, 1997

[42] B. R. Greene, W. P. Marnane, G. Lightbody, R. B. Reilly and G. B. Boylan, “Classifier models and architectures for EEG-based neonatal seizure detection,” vol. 29, Physiological measurement, 2008

Subject	Class	Number of Epochs	Number of Seizures
1	**GSW**	**59**	**52**
2	**GSW**	**29**	**19**
3	**GSW**	**16**	**14**
4	**GSW**	**19**	**20**
5	**PNES**	**1**	**1**
6	**PNES**	**1**	**1**
7	**PNES**	**1**	**1**
8	**PNES**	**13**	**13**
9	**PNES**	**3**	**3**
10	**VVS**	**45**	**1**
11	**VVS**	**18**	**1**

Table 1 Number of seizures and number of seizure epochs (2 seconds) per subject

Classification Model	Statistical Measures before Feature Selection			Statistical Measures after Feature Selection		
	Accuracy	***Sensitivity***	***Specificity***	***Accuracy***	***Sensitivity***	***Specificity***
BayesNet	**86%**	**92%**	78%	**95%**	**94%**	**98%**
RandomCommittee	83%	88%	77%	92%	89%	77%
RandomForest	74%	77%	70%	87%	92%	79%
IBk	69%	86%	43%	86%	94%	76%
SMO (RBF kernel)	68%	55%	**87%**	87%	85%	91%

Table 2 Classification performance before and after Feature Selection.

Feature	Channels
nmin	Fp1, T4, T5, Fz, F7, Fp2, A1, T6, Cz, O1, F4, F3, F8, P4
nmax	Fp1, T4, T5, F7, Fp2, P3, Fz, T3, T6, A1, Cz, O1, F4, F3, F8, P4
max	Fp1
std	Fp1
aryule3	T6, T4, F4, C4, F7, T5
maxfreq	T3, O2
aryule2	T4, T5, P4, Fp1, O1
logee	C4
cwt25	A1
cwt32	A1, Fp2
minfreq	F7

Table 3 Best features according to the top 50 ranking of STRATEGY 1 and the channels they appear on.

Feature	Channels
nmax	T4, Fp1, T5, F7, Fp2, Fz, T3, O1, P3, T6, F3, F8, F4, A1, Cz, P4
nmin	T4, Fp1, T5, F7, Fp2, Fz, T3, O1, P3, T6, F3, F8, F4, A1, Cz, P4
aryule3	T4, F4, T6, C4, T5, F7, O1
aryule2	T4, T3, T5, Fp2, F4, P4
Minfreq	T4
cwt32	A1, Fp2
cwt25	A1, Fp2

Table 4 Best features according to the top 50 ranking of STRATEGY 2 and the channels they appear on.

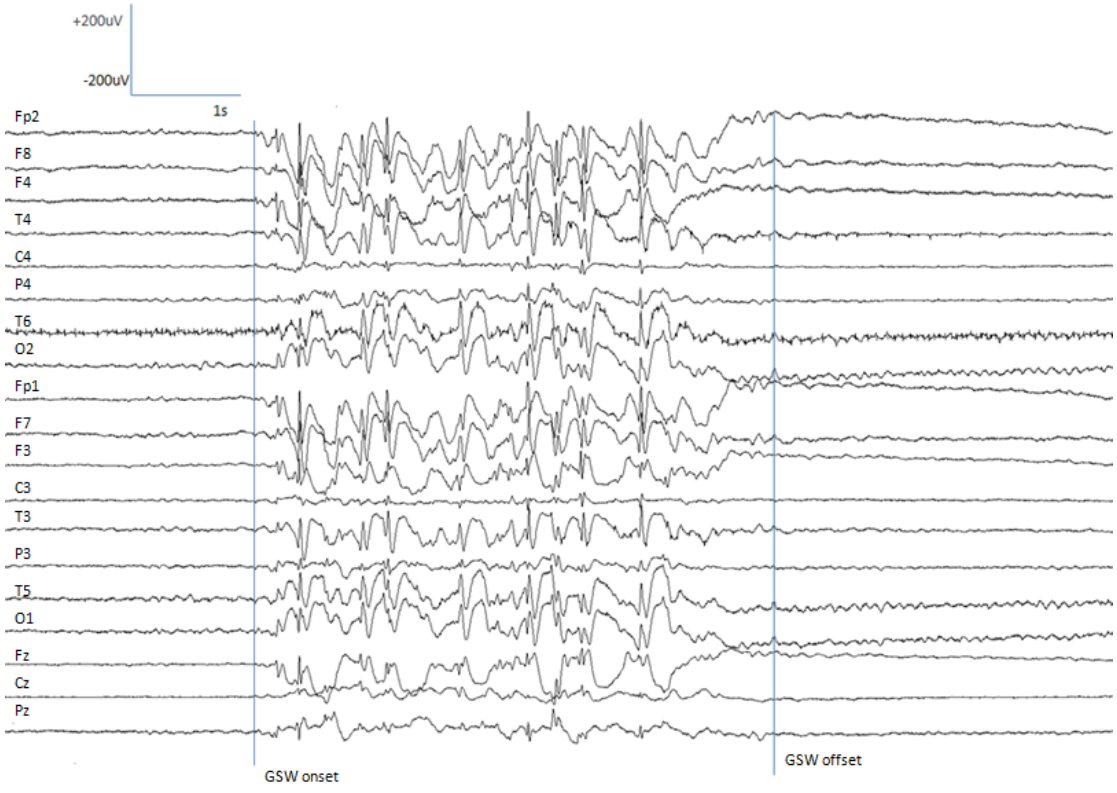


Figure 1 Generalized Spike Wave (GSW) example. The first marker indicates the beginning of the GSW event and the second marker its end.

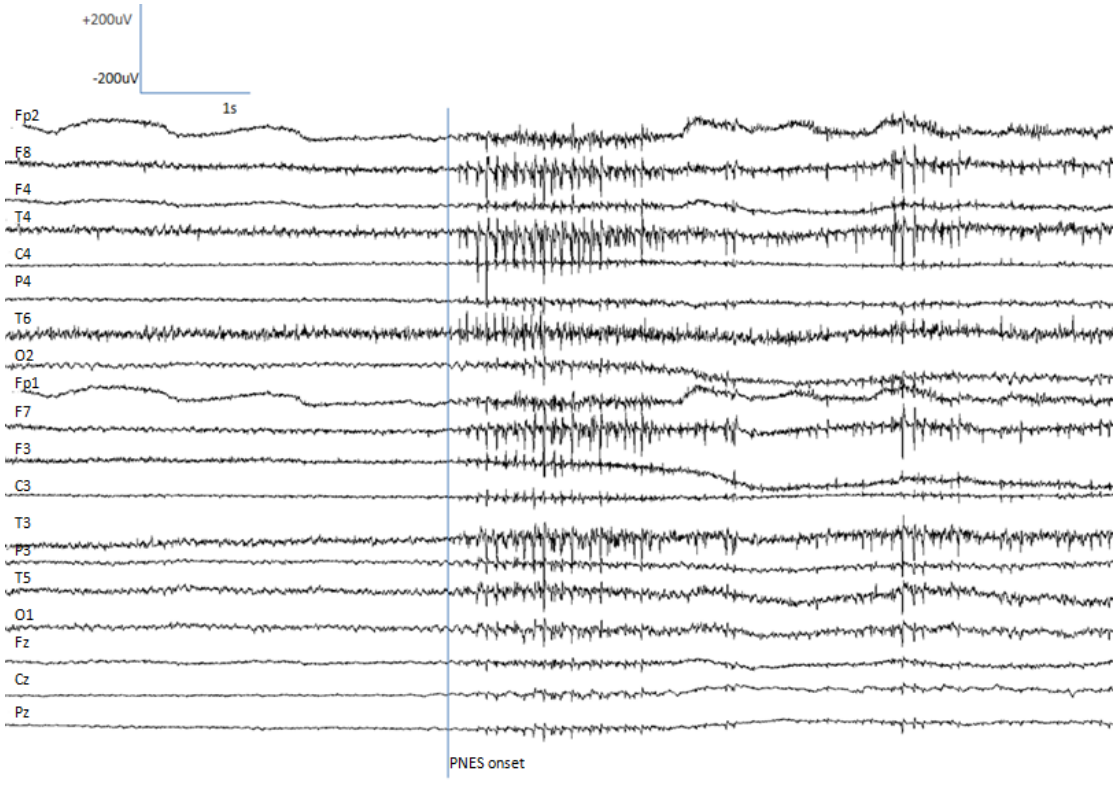


Figure 2 Psychogeninc Non Epileptic Seizure (PNES) example. The marker indicates the beginning of the PNES event.

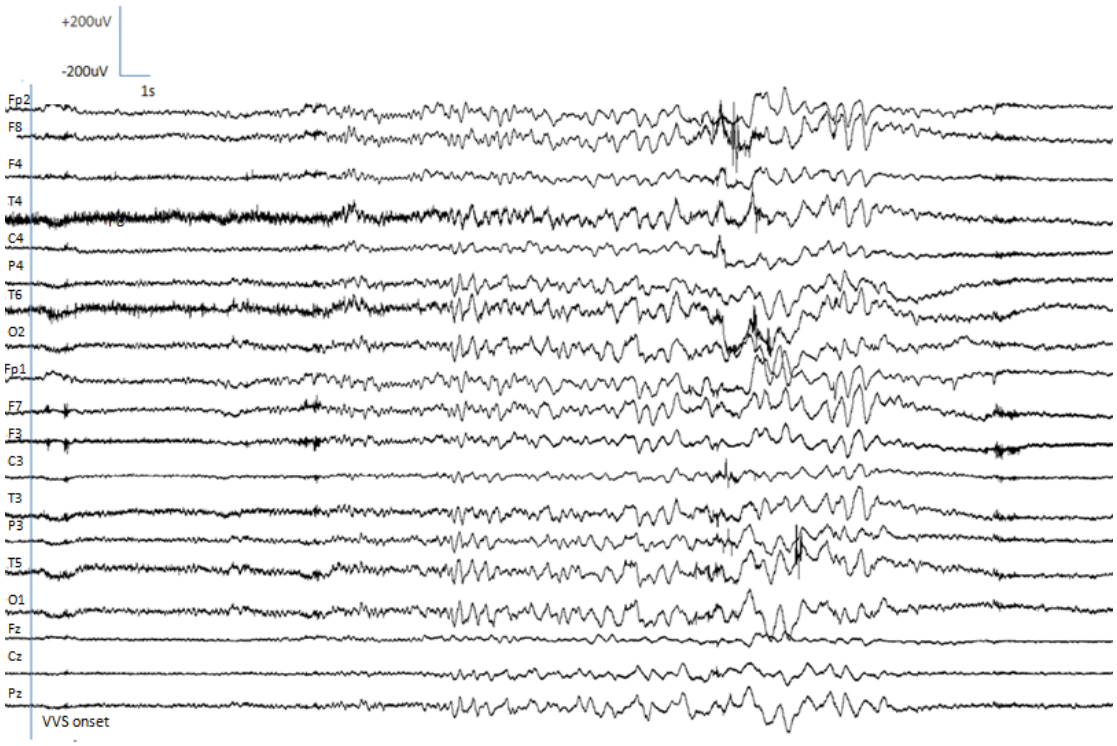


Figure 3 Vasovagal Syncopal Event (VVS) example. The marker indicates the beginning of the VVS event.

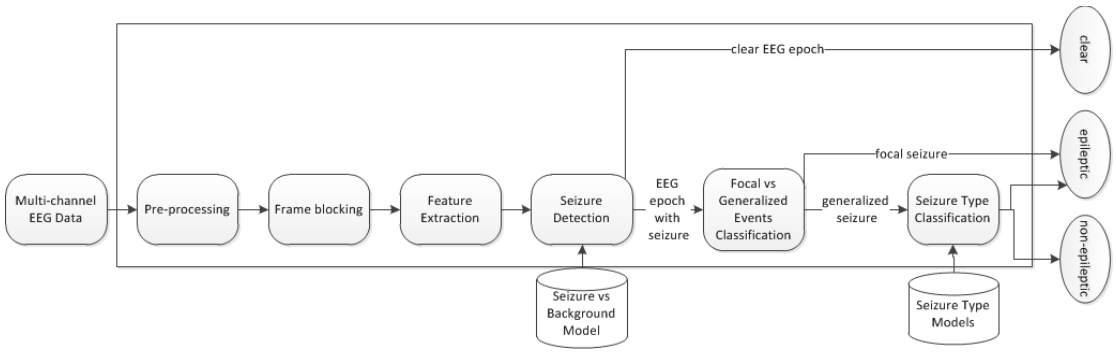


Figure 4. The concept of seizure detection and classification within the ARMOR framework.

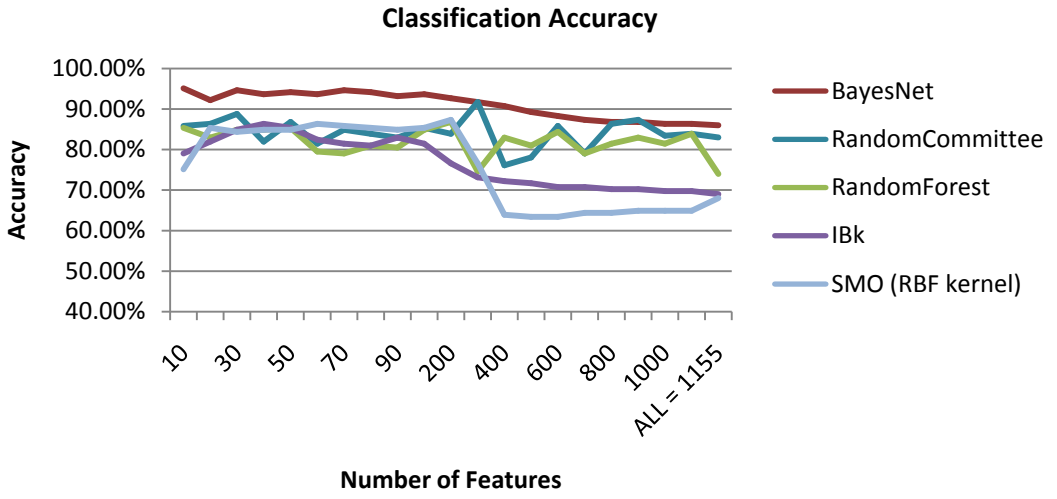


Figure 5 Classification Accuracy when 1st Ranking Strategy (based on Frequency) is used.

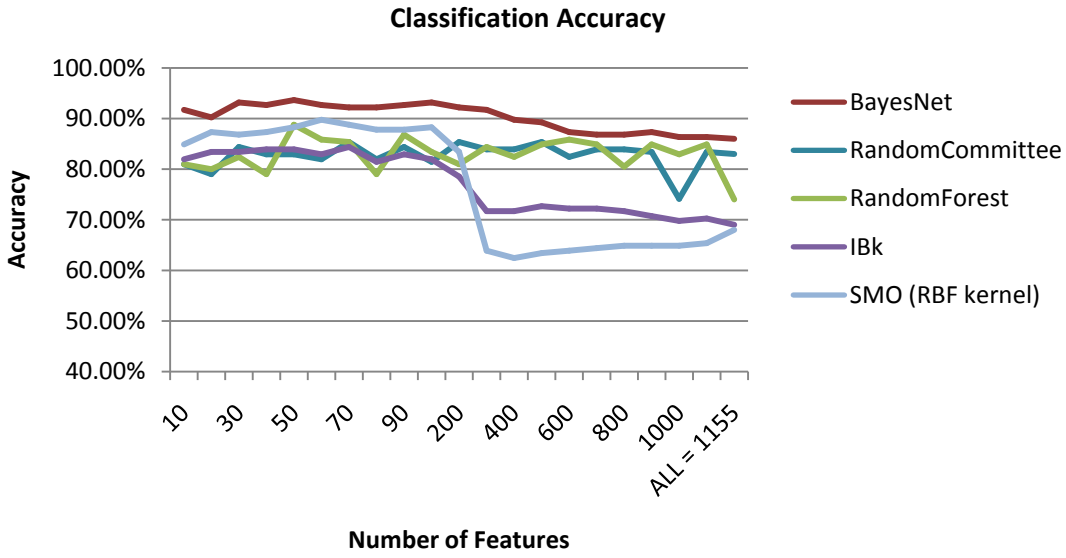
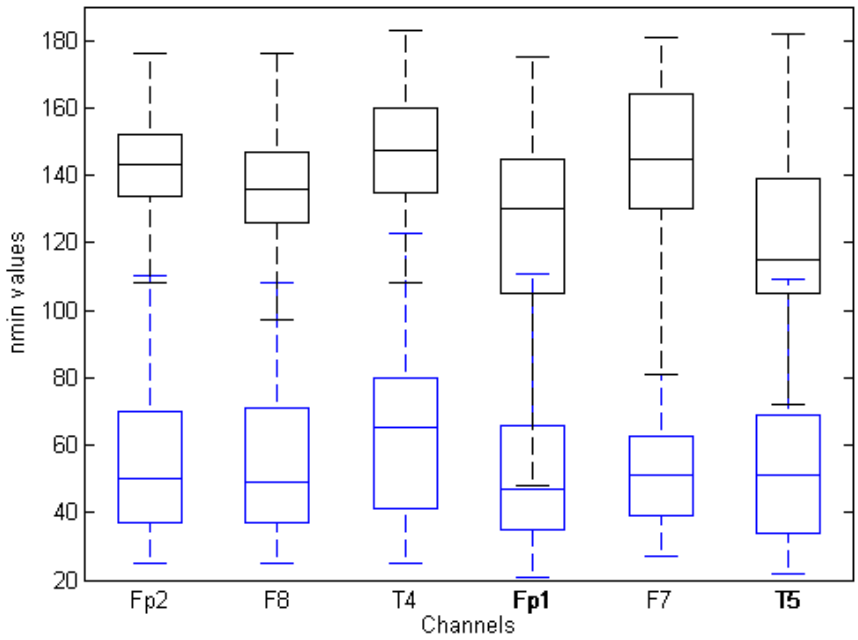
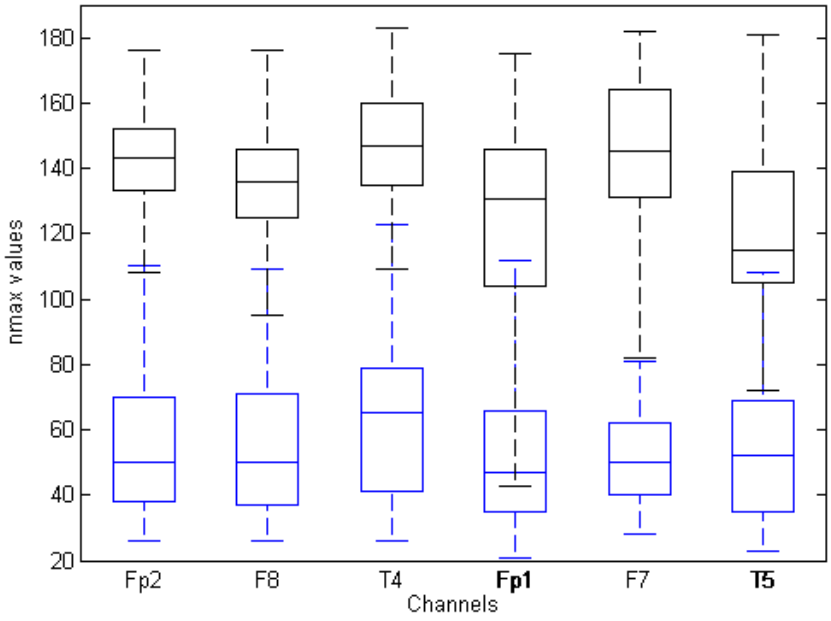


Figure 6 Classification Accuracy when 2nd Ranking Strategy (based on Sum of ReliefF Weights) is used.



1

2 **Figure 7 Values of the feature *nmin* of epileptic (blue boxes) and non-epileptic (black boxes) EEG epochs.**



3

4 **Figure 8 Values of the feature *nmax* of epileptic (blue boxes) and non-epileptic (black boxes) EEG epochs.**

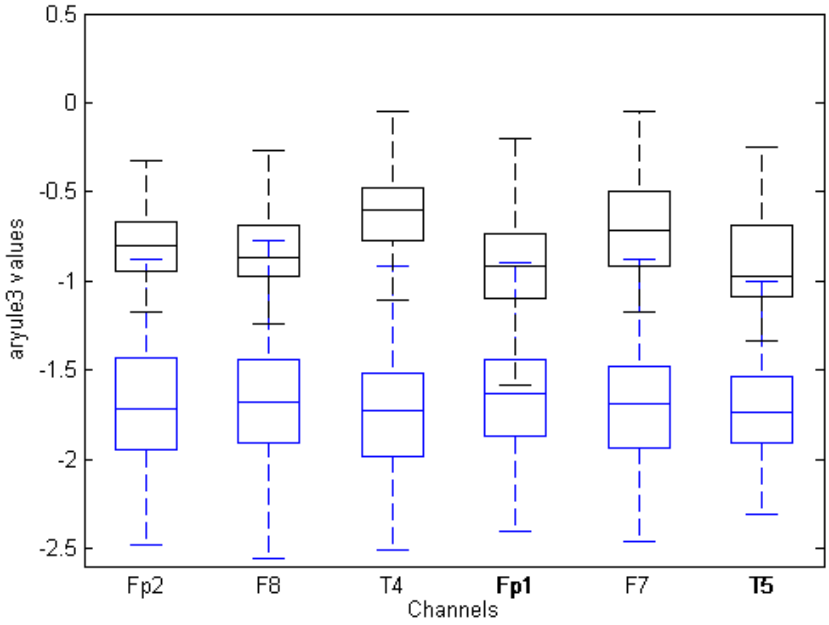


Figure 9 Values of the feature *aryule3* of epileptic (blue boxes) and non-epileptic (black boxes) EEG epochs.

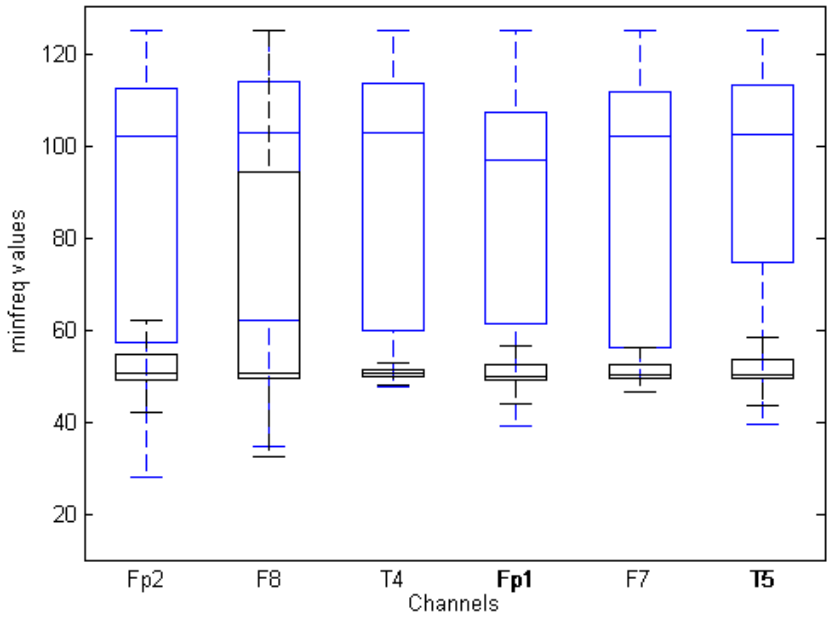


Figure 10 Values of the feature *minfreq* of epileptic (blue boxes) and non-epileptic (black boxes) EEG epochs.

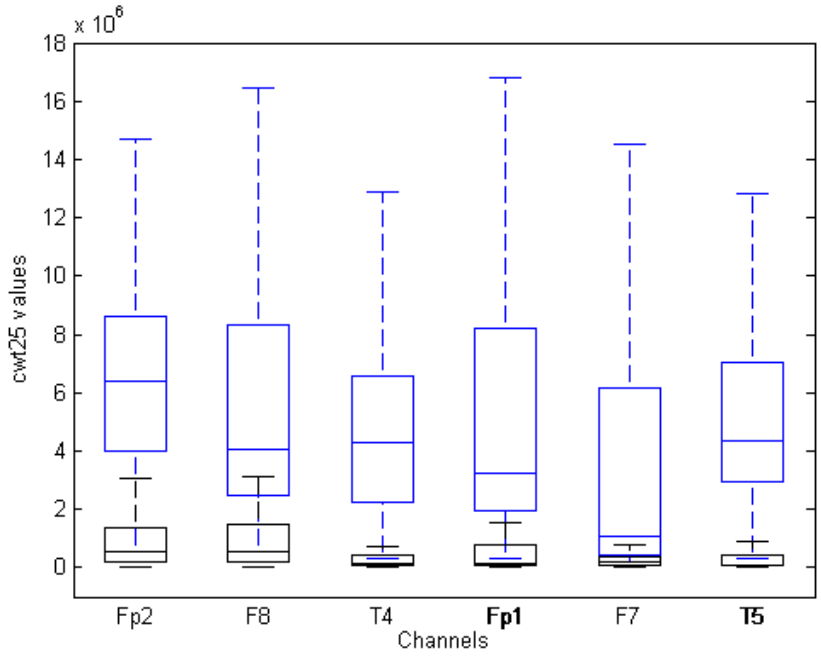


Figure 11 Values of the feature *cwt25* of epileptic (blue boxes) and non-epileptic (black boxes) EEG epochs.

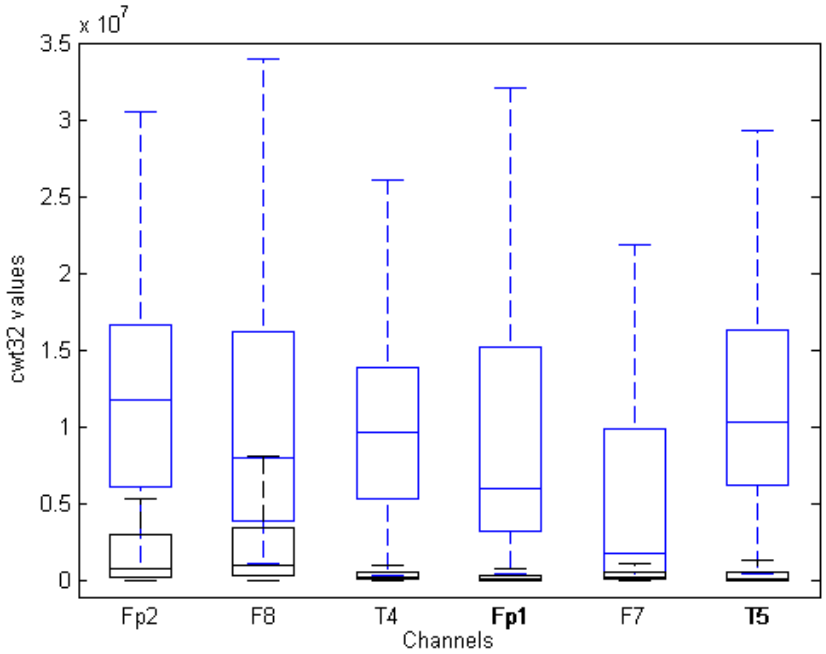


Figure 12 Values of the feature *cwt32* of epileptic (blue boxes) and non-epileptic (black boxes) EEG epochs.

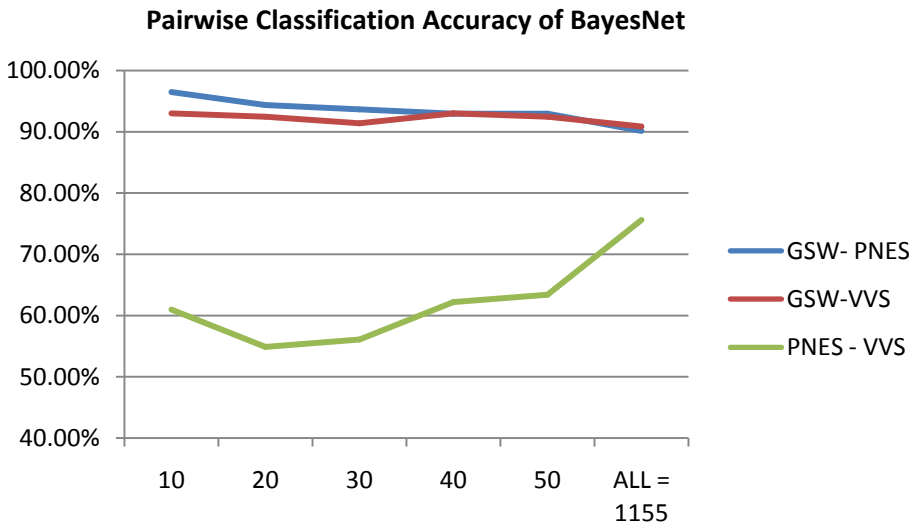


Figure 13 Pairwise classification accuracy of BayesNet classifier for different number of best features.

***Biography of the author(s)**

Click here to download Biography of the author(s): Biography of the authors.pdf

Evangelia Pippa, MSc. She received her BSc degree in Computer Engineering and Informatics and her MSc degree in Computer Science and Engineering from the University of Patras , in 2011 and 2013 respectively. She has participated as a researcher in a European R&D project in the area of data mining, pattern recognition and neuroinformatics. Currently she is a PhD student at the Dept. of Computer Engineering and Informatics of the University of Patras. Her research interests include data mining, pattern recognition and biomedical informatics.

Evangelia I. Zacharaki received the Diploma degree in Electrical and Computer Engineering and the Ph.D. degree from the National Technical University of Athens, Greece, in 1999 and 2004, respectively. She has been for 4 years a Postdoctoral Researcher at the Section of Biomedical Image Analysis, University of Pennsylvania, USA, and then joined for 3 years the Biosignal Processing Group at the University of Patras Medical School. She is currently working as a Postdoctoral researcher at the Dept. of Computer Engineering and Informatics of the University of Patras and as non-tenured Assist. Professor at the Technological Educational Institute of Western Greece. Her research interests include image and signal processing, pattern analysis, machine learning, and statistical modeling, with focus in biomedical applications. She has authored or co-authored 58 papers in refereed international journals and conference proceedings, receiving more than 700 citations until now.

Dr. Iosif Mporas was born in Athens in 1981. He received the Diploma degree in Electrical and Computer Engineering and the Ph.D. degree in signal processing from the University of Patras, Greece, in 2004 and 2009, respectively. He has participated as researcher in more than 5 European R&D Projects in the areas of signal processing and data-mining. He is currently working as a postdoctoral researcher at the Dept. of Computer Engineering and Informatics of the University of Patras and as non-tenured Assist. Professor at the Technological Educational Institute of Western Greece. His research interests include signal processing (speech, audio, bio-signals), data-mining and machine learning. He is author and co-author in more than 60 journal and international conference articles.

Vasilis Megalooikonomou (M'95) received a BSc in computer engineering and informatics from the University of Patras, Greece in 1991,and a M.S. and Ph.D. in computer science from the University of Maryland, Baltimore County, USA, in 1995 and 1997, respectively. He is currently a professor at the Department of Computer Engineering and Informatics at the University of Patras, Greece. Prior to this appointment he has been on the faculty of Johns Hopkins University, Dartmouth College and Temple University. He has co-authored over 160 refereed articles in journals and conference proceedings and three book chapters. His main research interests include medical image analysis, pattern recognition, data mining, data compression, biomedical informatics, and multimedia database systems. Prof. Megalooikonomou is a member of the ACM,IEEE, SIAM, and SPIE. In 2003 he received a CAREER award from the US

National Science Foundation for developing data mining methodsfor extracting patterns from medical image databases. He regularly serves as a program committee member or referee on several premier conferences and journals in his areas of research.

***Photo of the author(s)**

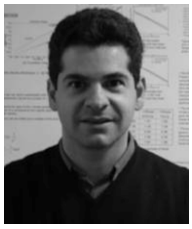
Click here to download Photo of the author(s): photos.pdf

Evangelia Pippa

Evangelia I. Zacharaki

Iosif Mporas

Vasileios Megalooikonomou



***Souce Files - Latex or Word**

Click here to download Souce Files - Latex or Word: Neurocomputing_review2_final.docx