DOI:
10.1016/j.jneumeth.2016.10.004

Document Version
Peer reviewed version

Citation for published version (APA):
Pippa, E., Zacharaki, E. I., Koutroumanidis, M., & Megalooikonomou, V. (2016). Data fusion for paroxysmal events' classification from EEG. *Journal of Neuroscience Methods*.
https://doi.org/10.1016/j.jneumeth.2016.10.004

Accepted Manuscript

Title: Data fusion for paroxysmal events' classification from EEG

Author: Evangelia Pippa Evangelia I. Zacharaki Michael Koutroumanidis Vasileios Megalooikonomou

PII:	S0165-0270(16)30228-X
DOI:	http://dx.doi.org/doi:10.1016/j.jneumeth.2016.10.004
Reference:	NSM 7606
To appear in:	*Journal of Neuroscience Methods*
Received date:	23-6-2016
Revised date:	11-9-2016
Accepted date:	7-10-2016

Please cite this article as: Pippa Evangelia, Zacharaki Evangelia I, Koutroumanidis Michael, Megalooikonomou Vasileios.Data fusion for paroxysmal events' classification from EEG.*Journal of Neuroscience Methods* http://dx.doi.org/10.1016/j.jneumeth.2016.10.004

Data fusion for paroxysmal events' classification from EEG

[*]Evangelia Pippa[1], Evangelia I. Zacharaki[2],

Michael Koutroumanidis[3] 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]CVC, Department of Applied Mathematics, Centrale Supélec - Equipe GALEN, INRIA Saclay, France

[3]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

Highlights

The main contributions of this work are summarized as following:

- highly accurate classification of epileptic and non-epileptic EEG events
- comparison between EI and LI fusion schemes regarding this problem
- novel LI fusion to handle the high dimensionality and the limited number of samples
- study of the behavior of each scheme as a function of the dimensionality
- dimensionality reduction through PCA

Abstract.

Background: Spatiotemporal analysis of electroencephalography is commonly used for classification of events since it allows capturing dependencies across channels. The significant increase of feature vector dimensionality however introduce noise and thus it does not allow the classification models to be trained using a limited number of samples usually available in clinical studies.

New Method: Thus, we investigate the classification of epileptic and non-epileptic events based on temporal and spectral analysis through the application of three different fusion schemes for the combination of information across channels. We compare the commonly used early-integration (EI) scheme - in which features are fused from all channels prior to classification - with two late-integration (LI) schemes performing per channel classification when: (i) the temporal context varies significantly across channels, thus local spatial training models are required, and (ii) the spatial variations are negligible in comparison to the inter-subject variation, thus only the temporal variation is modeled using a single global spatial training model. Furthermore, we perform dimensionality reduction either by feature selection or by principal component analysis.

Results: The framework is applied on events that manifest across most channels, as generalized epileptic seizures, psychogenic non-epileptic seizures and vasovagal syncope. The three classification architectures were evaluated on EEG epochs from 11 subjects.

Comparison with Existing Methods: Although direct comparison with other studies is difficult due to the different characteristics of each dataset, the achieved recognition accuracy of the LI fusion schemes outperforms the performance reported in the literature.

Conclusions: The best scheme was the LI with global model which achieved 97% accuracy.

Keywords: epileptic seizures, PNES, vasovagal syncope, EEG classification, data fusion.

1. Introduction

In clinical practice, electroencephalography (EEG) is used for the diagnosis and classification of interictal and ictal events (epileptic seizures) as well as the differentiation of the latter from other non-epileptic clinical events that may occur during recording, that mostly include vasovagal syncope (VVS) and psychogenic non-epileptic seizures (PNES). Electrodes, which act as sensors to detect the electrical activity, are attached to the scalp and provide both spatial and temporal information. There are two main approaches for fusing data from different EEG channels: early-integration and late-integration [1,2]. In EI, which is commonly used to exploit the spatiotemporal variation of EEG [3,4,5,6,7,8] and the dependencies across channels, the data are fused directly after feature extraction. Feature vectors from each channel are combined and events are classified by one global classifier. On the other hand, in LI, events are classified for each channel by its local classifier and the results from these local classifiers are later fused in the decision layer [1,2,7].

Analysis of the electrical activity of the brain is very complex and difficult to summarize with a small number of variables extracted from EEG signals. As a result, analysis of EEG is usually accompanied by extraction of high dimensional feature vectors from the data. The dimensionality is further increased in EI approaches aiming to exploit the spatial information of EEG, where already high dimensional feature vectors from several channels are combined to a single large feature vector. The problem of high dimensionality coupled with the limited number of samples usually available in clinical studies, makes the analysis of multidimensional EEG signal a challenging task.

Thus in this paper, we compare the commonly used EI scheme and LI scheme and propose a new LI scheme to deal with the problem of high dimensionality in conjunction with limited number of samples. The proposed scheme combines information from all channels in order to train the classification model and thus is channel-independent. In general, the LI scheme keeps the dimensionality quite low, while the incorporation of a global training model allows the use of more training samples (by combining all channels). The performance of each scheme, as a function of the feature vector dimensionality, is also studied by performing feature ranking and selection prior to the classification using t-test as ranking method. The performance of the different schemes is investigated in relation to the problem of discrimination between clinical events of different nature, manifested by paroxysmal loss of consciousness. The differential diagnosis that a clinician usually faces is mainly that of a generalized epileptic seizure, a psychogenic non-epileptic seizure (PNES) and a vasovagal syncope (VVS). The diagnosis and management of paroxysmal loss of consciousness 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 [9].

An epileptic seizure is a transient occurrence of signs and/or symptoms due to abnormal excessive or synchronous neuronal activity in the brain [10], typically associated with EEG specific changes. The identification of epileptic events can be achieved by certain characteristic ictal neurophysiological patterns that appear during the episode. Psychogenic non-epileptic seizures (PNES) are paroxysmal events that result in loss of consciousness resembling epilepsy but without the characteristic electrical changes associated with the episodes of the latter [11]. Although historical information can sometimes support the discrimination between PNES and epileptic seizures, confident distinction on clinical grounds is frequently difficult. This is due to the insufficient event description by the patient and witnesses and the possible coexistence of epilepsy and PNES in the same patient.

Vasovagal syncope (VVS) is a common type of syncope [12]. During a VVS characteristic EEG changes may include progressive generalized theta slowing of background rhythms, followed by diffuse delta activity of high voltage and appearance of progressively lower voltage rhythms until isoelectric suppression [13, 14]. This pattern is progressively reversed after the end of the event when cerebral perfusion is restored. The changes captured by EEG recordings during a VVS do not include any ictal activity.

Despite such diagnostic uncertainty, the problem of automated discrimination between epileptic and non-epileptic pathological EEG events is rarely tackled in the literature. Relevant literature include an algorithm proposed in [15] that is based on the correlation between features extracted from an appropriately selected epileptic EEG segment and the unknown ones in order to classify the latter into epileptic on non-epileptic. The extracted features used consist of auto-correlation coefficients and the achieved sensitivity and specificity are 83% and 90%, respectively. Two years later, the authors in [16] used a set of auto-correlation coefficients to train an LVQ1 neural network. The evaluation of the LVQ1 model on testing EEG segments resulted in 86% accuracy. The feature extraction methods of the aforementioned classification frameworks, as well as the achieved results were subsequently reviewed in [17] and statistical analysis using a chi-square test revealed the superiority of the LVQ1 method.

In a previous work [5], both PNES and VVS events were examined in an attempt to extend the non-epileptic class. In order to automatically classify epileptic and non-epileptic (PNES and VVS) EEG epochs, a large set of temporal and spectral features was examined using an EI scheme for the combination of information across EEG channels using a dataset of 11 patients. Although such a spatiotemporal analysis captures holistically the change of the EEG signal, the limited number of the available samples was not enough to fully capture the spatiotemporal variation. This is a common problem in biomedical applications where the high dimensionality of the data hinders data modeling and representation [18].

Building upon our previous work [5] in which an EI scheme was implemented, we now investigate two LI schemes performing per channel classification. The first scheme is based on the assumption that the temporal context varies significantly across channels, thus local training models are built, while the second scheme is based on the assumption that the spatial variations are negligible in comparison to the inter-subject variation, thus global training models can be used. Obviously this type of analysis can only be performed on events that generalize across EEG channels, such as the ones used in this study.

The rest of this paper is organized as follows. In Section 2 the evaluation data and the different fusion schemes for classification of epileptic and non-epileptic events are presented. Section 3 provides details about the validation and the achieved results. Finally in Section 4 we conclude this work.

2. Material and methods

2.1 Data

In this paper, we use EEG recordings acquired by the Department of Clinical Neurophysiology and Epilepsies in St Thomas' Hospital, London, UK from 11 patients for the needs of the ARMOR project [19]. For investigation purposes the epileptic and non-epileptic events were manually annotated and isolated from the recordings. The epileptic events were derived from patients diagnosed with Idiopathic Generalized Epilepsy (IGE) with absence seizures. The isolated epileptic events consist of Generalized Spike Waves (GSW) derived from the epileptic group. The non-epileptic events were derived from 2 patients with VVS and 5 patients with PNES. For all the examined subjects, at least one typical epileptic or non-epileptic event appear during the recording. The recordings were performed using conventional AgCl EEG electrodes positioned according to the extended international 10-20 system. For the analysis we used EEG channels Fp2, F8, F4, T4, C4, A2, P4, T6, O2, Fp1, F7, F3, A1, C3, T3, P3, T5, O1, Fz, Cz, and Pz. Note that in this study,A1 and A2 are midtemporal active electrodes. The montage is referenced to C3+C4/2. During the training and test phases of our classification models we considered only epochs that contained the epileptic or non-epileptic events. Details on the examined subjects are shown in Table 1.

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

The presented methodology performs short time analysis in the multidimensional EEG data (one dimension per electrode) and binary classification between epileptic or non-epileptic (PNES or VVS) events using one of three investigated fusion schemes. The multidimensional EEG data are initially preprocessed using notch filtering, baseline correction, and segmentation of the incoming EEG signals to epochs of constant length w with constant time-shift and without time-overlap between successive epochs. Thus each data sample is represented by a $N \times w$ matrix, where N is the number of EEG electrodes. The epoch length was selected equal to 2 seconds.

After preprocessing, temporal and spectral analysis is performed for each epoch resulting to a feature vector of dimensionality equal to 55 for each of the N EEG channels, as described in more details in our previous work [5, 6]. For completeness the extracted features are summarized in Table2. The rationale and clinical basis for the features are explained in separate section (section 2.3).

During the training phase, training data including EEG recordings with manual time annotations for the onsets and offsets of the epileptic and the non-epileptic events are preprocessed, segmented and parameterized as described above. The produced feature vectors from each epoch (either concatenated from all channels for the EI or separately for the LI) are used to build a binary classification model. During the test phase the unknown multidimensional EEG signal is preprocessed and parameterized with the same features as in the training phase. Each produced feature vector is compared against the epileptic and non-epileptic classification model, and a class label is assigned to each corresponding epoch.

We previously evaluated [5] the ability of the extracted features to differentiate epileptic from non-epileptic epochs by examining several classification algorithms implemented by the WEKA machine learning toolkit software [24] including the BayesNet [24,26], RandomCommittee, RandomForest [27], IBk [28] and SMO [29,30] with RBF kernel. Since the overall highest accuracy was achieved by the BayesNet classifier, we now evaluate the examined fusion schemes with respect to the BayesNet.

In BayesNet algorithm, a bayesian belief network structure, which is a directed acyclic graph, is built [25,26]. In such a graph, features are represented by nodes and dependencies among features are represented by edges between nodes. If there is no direct probabilistic dependency between the features the corresponding edge is absent. In a bayesian belief network, a conditional probability function exists for each node that relates it to its parents. In order to find the most probable structure for the bayesian belief network, numerical probabilities are derived based on the training data. Thus, the probabilities in the belief network are used to compute the probability of any sample.

2.2.1 Early Integration

During the training phase, a set of EEG epochs with known class labels is used to estimate a model. Let as denote with M the number of epochs resulting from the segmentation of the training signals, N the number of channels (here, $N = 21$) and f the number of features extracted from each epoch of each channel (here, f = 55). In the EI scheme, each of the N available channels from each training epoch is processed in parallel by the feature extraction algorithm yielding to $M \times N$ feature vectors of dimensionality f: $[f_{mn1}, f_{mn2}, \dots, f_{mnf}]$, $m = 1,2, \dots M$ and $n = 1,2, \dots N$. For each epoch, the N estimated feature vectors derived from the N channels are concatenated into a single feature vector $[f_{m11}, f_{m12}, \dots, f_{m1f}, f_{m21}, f_{m22}, \dots, f_{m2f}, \dots, f_{mN1}, f_{mN2}, \dots, f_{mNf}]$, $m = 1,2, \dots M$. This high dimensional feature vector is used as a representative signature for the corresponding training epoch in the training set. Therefore, the training set is a data matrix M × (N × f). In particular, the feature matrix F is formulated as follows:

$$F = \begin{bmatrix} f_{111}, f_{112}, \dots, f_{11f}, f_{121}, f_{122}, \dots, f_{12f}, \dots, f_{1N1}, f_{1N2}, \dots, f_{1Nf} \\ f_{211}, f_{212}, \dots, f_{21f}, f_{221}, f_{222}, \dots, f_{22f}, \dots, f_{2N1}, f_{2N2}, \dots, f_{2Nf} \\ \vdots, \vdots, \ddots, \vdots, \vdots, \vdots, \ddots, \vdots, \ddots, \vdots, \vdots, \ddots, \vdots \\ f_{M11}, f_{M12}, \dots, f_{M1f}, f_{M21}, f_{M22}, \dots, f_{M2f}, \dots, f_{MN1}, f_{MN2}, \dots, f_{MNf} \end{bmatrix}$$

As a result, each row of the data matrix in the EI contains $N \times f = 21 \times 55 = 1155$ features.

Similarly, for each test epoch, the estimated feature vectors from each channel are concatenated into a single feature vector of dimensionality $N \times f$ and fed to the classification model trained beforehand. The EI scheme is illustrated in Figure 1. Although, such a scheme exploits the spatial information of the EEG data, it leads to a feature vector of high dimensionality imposing the need either for feature selection before classification, or the availability of a significant number of training samples.

2.2.2 Late Integration with Local training models

In the LI with local (channel dependent) training models scheme, a separate classification model is built for each channel. During the training phase, the N channels are processed in parallel by the feature extraction algorithm. Each channel produces M (one for each training epoch) feature vectors of dimensionality f: $[f_{m1}, f_{m2}, \dots, f_{mf}]$, $m = 1,2, \dots M$. These feature vectors are used as rows to form the training set for the corresponding channel n, $n = 1, \dots N$. As a result, we have N training sets $\{F_1, F_2, \dots F_N\}$, one for each channel formulated as follows:

$$F_n = \begin{bmatrix} f_{11} & f_{12} \cdots & f_{1f} \\ f_{21} & f_{22} \cdots & f_{2f} \\ \vdots & \vdots \ddots & \vdots \\ f_{M1} & f_{M2} \cdots & f_{Mf} \end{bmatrix}, n = 1, \dots N$$

The data matrix of each training set is M × f.

During the test phase, the estimated feature vector of dimensionality f from each one of the N channels is fed to the corresponding local classification model. For each epoch, N decisions are made by each one of the N local classifiers. A final decision is made by combining the N output class labels using majority vote. The LI with local training models fusion scheme is illustrated in Figure 2. In LI schemes the dimensionality of the feature vector is smaller than in EI schemes. However, this scheme uses training samples only of the corresponding channel.

2.2.3. Late Integration with Global training model

In the LI with global (channel independent) training model fusion scheme, a common classification model is used for the feature vectors extracted from the different channels. The data matrix F of the training set is $(N \times M) \times f$ and is constructed by merging all training sets from the LI local fusion scheme:

$$F = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_N \end{bmatrix}$$

In particular, the LI scheme with global model uses one global feature matrix F (the same for all of the N classifiers) formulated as follows:

$$F = \begin{bmatrix} f_{111} & f_{112} \cdots & f_{11f} \\ f_{211} & f_{212} \cdots & f_{21f} \\ \vdots & \vdots \ddots & \vdots \\ f_{M11} & f_{M12} \cdots & f_{M1f} \\ f_{121} & f_{122} \cdots & f_{12f} \\ f_{221} & f_{222} \cdots & f_{22f} \\ \vdots & \vdots \ddots & \vdots \\ f_{M21} & f_{M22} \cdots & f_{M2f} \\ \vdots & \vdots \ddots & \vdots \\ f_{1N1} & f_{1N2} \cdots & f_{1Nf} \\ f_{2N1} & f_{2N2} \cdots & f_{2Nf} \\ \vdots & \vdots \ddots & \vdots \\ f_{MN1} & f_{MN2} \cdots & f_{MNf} \end{bmatrix}$$

In this scheme the number of training samples is increased since each epoch appears in the training set N times, one time for each one of the available channels.

During the test phase, for each epoch, N decisions are made by feeding the signature from each channel to the global classification model. A final decision is made at a score level by combining the N output class labels using majority vote.

The LI with global training model scheme is illustrated in Figure 3. Although this scheme is less specific, it handles better both the high dimensionality (by keeping the size of the feature vector lower than the one of EI) and the problem of limited number of instances (by treating the epoch from different channels as independent samples in the training set).

2.3 Feature Extraction

The selection of features to be extracted was based on the existing literature in an attempt to include features that have been widely used for the analysis of EEG signals. In order to provide an insight into the extracted features we refer to the corresponding studies [1,6,20,21,22,23]. Statistical features such as minimum, maximum, mean, variance, standard deviation, percentiles, interquartile range, mean absolute deviation, range, skewness, kyrtosis and energy are used to capture the variations in the amplitude of the EEG signals that accompany the electroencephalographic seizure activity [20,21,22]. Entropy can be interpreted as a measure of signal complexity and so represents a potential feature for seizure classification [1]. The number of local maxima and local minima are used to capture the different amount of smoothness of EEG signal variance that are observed during different types of seizures [6].The zero crossing rate is related to changes in the frequency and thus it has been proposed to capture changes during seizure activity [20,23]. The autoregressive model expresses the signal with lagged terms of itself and thus specifies whether the EEG epoch depends linearly on its own previous values. The lower absolute values of the AR coefficients indicate that the signal is much more noisy and stochastic-like such as in case of PNES whereas the higher ones indicate more structured and deterministic-like signal such as in case of GSW [6]. The purpose of extracting the power spectral density was to find the most prominent rhythmic component of each epoch. The frequencies with maximum and minimum amplitude were derived from power spectral density for similar purposes. The wavelet transform expresses the signal as linear combination of the chosen wavelet basis functions and thus capture the frequency content of the signal on a localized area which seem to differentiate between the examined classes [6].

2.4 Dimensionality reduction, Feature Ranking and Feature Subsets Evaluation

Since the whole set of features was quite large (*55 features* for the LI schemes and *21channels* × *55 features = 1155 features* for the EI scheme) and the usefulness of each feature for differentiating epileptic from non-epileptic events is not similar, a dimensionality reduction was made in order to increase the classification performance. Two different strategies for dimensionality reduction were selected: feature selection by feature ranking and principal component analysis (PCA).

Concerning feature selection, for each fusion scheme we examined the discriminative power of the extracted features for the classification of epileptic and non-epileptic EEG events. The importance of each feature in binary classification was estimated by statistical analysis using the t-test. In order to perform ranking we followed a leave-one-out strategy on the available subjects. In particular, for each leave-one-out experiment, feature ranking was performed using the t-test in each training subset. As a result, the retained features were slightly different for each leave-one-out experiment. We examined the performance of the method, in terms of accuracy, sensitivity and specificity, for different number of N-best features *(N =1, 2, 3, ... 55)* when LI schemes are evaluated and *(N=10,20, 30, ... 1150)* when EI scheme is evaluated.

As an alternative strategy for dimensionality reduction, we employed PCA. PCA is a transformation that convert possibly correlated features to an orthogonal basis set of principal components that consists of linearly uncorrelated features. The linear combinations of the principal components can represent the data with the highest variance in a feature subspace and thus is considered as optimal. PCA sorts the eigenvalues of the covariance matrix of the feature vectors in descending order and retains the eigenvectors that corresponds to the largest ones since they capture a high percentage of the total variance (e.g. 99%). The selected eigenvectors form the transformation matrix and result in feature vectors with reduced dimensionality. PCA was performed for each fusion scheme and the performance in terms of accuracy, sensitivity and specificity, for different number of retained eigenvectors so as different amounts of variation are kept, was evaluated.

3. Results and Discussion

A leave-one-patient-out cross-validation strategy was employed for the evaluation. In particular, for each iteration, one subject was left-out for testing, while the rest of the subjects were used for training. For the left-out subject, as test samples we used all the epochs between seizure onset and offset. Table 3 shows the number of epochs (M) that were extracted for each subject during the seizure.

It is worth to note that the number of epochs for the subjects with PNES is very small. This is owed to the following facts. Since in the case of PNES there is no significant pathological change on the EEG it was difficult even for the experts to define the offset of the seizure. Furthermore, the clinical symptoms are so atypical that neither the patient can define the offset. As a result, only the onset of the seizure was annotated and for the analysis purposes only one epoch was isolated as sample for each PNES. However, we believe that the number of PNES samples although limited are sufficient given the lack of ictal EEG changes and the fact that their variability reflects only muscle and movement activities.

The fusion schemes described in the previous section were evaluated regarding the classification performance they obtained. Table 4 shows the classification performance in terms of accuracy, sensitivity and specificity, defined as:

$$\text{Accuracy} = (TP+TN)/(TP+FP+TN+FN) \quad (1)$$

$$\text{Sensitivity} = TP / (TP + FN) \quad (2)$$

$$\text{Specificity} = TN / (FP + TN) \quad (3)$$

where true positives are denoted as TP, true negatives as TN, false positives as FP and false negatives as FN. Here we consider the epileptic class as the positive and the non-epileptic class (PNES or VVS) as the negative.

As can be seen in Table 4, the overall highest accuracy for classification between epileptic and non-epileptic EEG events is 90% for the LI with global training model fusion scheme. The LI with local models and EI schemes follow with 89% and 86% accuracy, respectively. For the LI with global model scheme with the highest accuracy, the sensitivity (or recall), i.e. the proportion of epileptic events which are correctly classified as such, is 94% and the specificity, i.e. the fraction of non-epileptic events (either PNES or VVS) which are correctly identified as such, is 84%. Although a LI scheme with either channel dependent models or one channel independent model, does not exploit the spatial information of the EEG, it improves the classification performance. It seems that the high dimensionality of the training samples at the EI fusion scheme of [5] is not appropriate for rather limited datasets. The characteristics of the LI with global model fusion scheme (smaller dimensionality, more training samples), make it the most appropriate for our dataset.

In a further step, we applied feature ranking using t-test and the leave-one-out strategy as described in Section 2.4 The performance of the classification, in terms of accuracy, for each fusion scheme separately and for different number of N-best features, *N=10,20, 30, ... 1150* when EI scheme is evaluated and *N =1, 2, 3, ... 55* when LI schemes are evaluated are shown in Figure 4 and Figure 5 respectively.

As can be seen in the above figures for all fusion schemes the highest classification accuracy is achieved when a small subset of discriminative features is used. Specifically, when LI fusion schemes are used the highest accuracy is achieved for a subset of 2 best features (number of local minima and number of local maxima) with a percentage of 91,71% for both of them, which is sufficiently high in comparison to the accuracy achieved when all features are used (88,78% for the LI with local models and 90,24% for the LI with global model). Similarly, EI fusion scheme achieve its highest accuracy (90,73%) for a subset of 100 best features. The LI scheme with global training model present a more stable behavior as the number of best features increases while the one with local models present a slight but more clear drop. It seems that the larger number of samples in training set of the LI scheme with the global model does not allow the high dimensionality of the feature vector to affect the classification performance significantly. Furthermore, the LI fusion scheme with global model outweighs the EI scheme even when the latter is used with a small subset of best features. It is worth to note that the best performance for the EI scheme (90,73% accuracy) which is achieved for the 80 best features is very close to the performance of the LI scheme with global model even when the latter is used with its highest dimensionality (55 features). In general, the experimental evaluation of the fusion schemes under examination indicates that in datasets with rather limited number of samples, LI schemes with global models give better classification performance compared to EI schemes even after feature selection.

Finally, for each fusion scheme we performed PCA on the corresponding feature matrix. The feature matrices are formulated as described in Section 2. The feature matrix in the EI scheme is of dimensionality 205×1155. The dimensionality of each feature matrix in LI scheme with local models is 205×55 whereas in LI scheme with global model it is 4305×55.

After sorting the eigenvectors in descending order we compute the proportion of retained variance for different number of retained eigenvectors by

$$Variance = \frac{\sum_{i=1}^{r} \lambda_i}{\sum_{j=1}^{m} \lambda_j}$$

where λ_i is the eigenvalue for the i-th principal component, r is the number of retained eigenvectors and m is the total number of components. The retained variance as a function of the number of retained eigenvectors for the EI and LI with local and global models fusion schemes is shown in Figure 6. For the LI fusion scheme with local models the mean across channels variance is shown. Figure 7 shows the per channel retained variance for different number of PCA retained eigenvectors for the LI fusion scheme with local models. For the EI scheme 198 eigenvectors are required in order to achieve 100% variance while for the LI schemes 43 and 36 eigenvectors are required for the case of local and global models, respectively. However, in order to achieve 99,99% variance, only 11 retained eigenvectors are required for the EI scheme, 3 for the LI scheme with local models and 2 for the LI scheme with a global model. For the EI scheme the covariance matrix is of dimensionality *1155x1155* with *rank = 107*. The number of the non-zero eigenvalues is 204. For the LI local the 21 covariance matrices are of dimensionality *55x55*. The rank for each channel and the number of non-zero eigenvalues are shown in Table 5. Finally, for the LI scheme with global model the covariance matrix is of dimensionality *55x55* with *rank = 26*and non-zero eigenvalues =52.

Table 6 shows the classification performance in terms of accuracy, sensitivity and specificity for each fusion scheme when 99,99% and 100% of variation are kept with respect to the PCA. In this study, we do not standardize the data before applying PCA which means that different features are measured on different scale. As a result, a sorting of the features based on their variance is performed since the principal components are dominated by a single or a few features, the one(s) with the highest variance. In such a case, all the variance is explained by very few components. When the data are standardized using z-score before the application of PCA, more principal components contribute on the explanation of the data variance, since z-score implies that all features have similar importance. In particular, when standardizing the data, in order to achieve 99% variance, 115 retained eigenvectors are required for the EI scheme, 19 for the LI scheme with local models and 20 for the LI scheme with a global model. However, in this case, the classification accuracy is significantly reduced probably due to the introduced noise by the additional components.

As can be seen from Table 6, for all fusion schemes only a few retained eigenvectors (so as 99,99% of variance is kept) are enough to achieve a higher classification performance in comparison to the 100% retained variance case in which the additional eigenvectors introduce noise. Once again the LI scheme with a global training model outperforms the other two schemes and achieves the overall highest accuracy (96,59%) and sensitivity (100%) of our work with the burden of lower specificity (91,46%). However, the fact that PCA assists the LI scheme with global model in classifying more accurately the epileptic events but fails to improve the classification of the non-epileptic ones should not be considered as a general conclusion since it is possibly owed to our unbalanced dataset. The number of the epileptic epochs in our dataset (123 epochs) is slightly higher than the one of the non-epileptic (82). As a result, the few retained eigenvectors describe better the epileptic class resulting to high sensitivity. For the two other schemes, the classification accuracy when using PCA drops in comparison to both classifications with fully-dimensional or reduced by the t-test feature vectors. It seems that the variability of the feature matrix on those schemes is high (since each epoch appears once) and as a result it does not aid the PCA to reveal components of the data useful to discriminate between the two classes. Such a claim does not hold for the LI scheme with a global model. Each epoch contained in the feature matrix constructed by the LI scheme with a global model is represented multiple times (one for each available channel). As a result, the feature matrix of the LI scheme with a global model presents a relatively low variability which is owed to the potential correlation between the feature vectors that while they are representing the same epoch, they are considered as independent samples.

The classification performance for different number of retained eigenvectors with respect to each fusion scheme are shown to Tables 7, 8, and 9. For both LI schemes the classification accuracy increases with the growth of retained variance reaching its maximum when 3 and 2 components are retained for the LI scheme with local and global models respectively. For the EI scheme the maximum accuracy is achieved when 198 components are retained (100% retained variance) and such a steady increase of accuracy cannot be observed.

Although it is difficult to directly compare our results with other studies due to the different characteristics of each dataset (e.g., different seizure types, lack of PNES or VVS examples or single channel data), the achieved epileptic recognition accuracy of the LI fusion schemes outperforms the performance reported in the literature. In particular, the achieved accuracy in [16] is 86%, lower to the accuracy of both late-integration fusion schemes in our methodology. Furthermore, in [15] the reported sensitivity (83%) is lower than the sensitivity of all the fusion schemes evaluated in our work, while the specificity is 90%, lower than the specificity of all fusion schemes after feature selection.

4. Conclusions

In this paper, we investigated the problem of classification between epileptic (IGE) and non-epileptic (PNES and VVS) events from multi-channel EEG data using temporal and spectral analysis. We examined three different fusion schemes for the combination of information across channels, one EI scheme performing fusion of features per channel to reach a decision and two LI schemes performing fusion of channel based decisions using either channel dependent training models or a channel independent training model. The proposed methodology was evaluated in EEG data from 11 subjects and the average accuracy was 90% for the LI scheme with a global classification model, greater than the one of other relevant studies. The superiority of the LI with global model fusion scheme (which is the one with the smaller dimensionality and the more training samples) indicates that both feature vector dimensionality and size of the training set plays a crucial role for the classification performance. The classification performance was further improved by feature selection using a subset of 2-best features resulting to 92% accuracy for the best scheme (LI with global model). In a further step, we performed dimensionality reduction through PCA. Classification using only a few eigenvectors so as a high percentage of variance is retained (99,99%) was performed for each scheme. PCA helped the proposed LI scheme with global model reach the overall highest accuracy of our work (97% accuracy for a 100% sensitivity and 91% specificity). Although the dimensionality of the feature vector was examined in this study, the size of the dataset is still being one more parameter that should be considered when comparing the three fusion schemes. Under this scope we aim to evaluate our framework on datasets with different sizes and study the behavior of each scheme as function of the number of samples available in the training set.

5. Acknowledgement

The authors wish to thank Iosif Mporas from Multidimensional Data Analysis and Knowledge Management Laboratory, Dept. of Computer Engineering and Informatics, University of Patras for his help with data preprocessing, Dr Vasiliki Tsirka from, Division of Neuroscience at the Institute of Psychiatry, Psychology & Neuroscience, King’s College London for contributing to the data annotation, Prof Mark Richardson from, Division of Neuroscience at the Institute of Psychiatry, Psychology & Neuroscience, King’s College London and Konstantinos Deltouzos from Multidimensional Data Analysis and Knowledge Management Laboratory for valuable discussions. This study was partially funded by the EC under the FP7/2007-2013 with grant ARMOR, Agreement Number 287720. This research has been also 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] 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.

[2] B. 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, pp. 1157

[3] 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).

[4] I. Mporas, V. Tsirka, E.I. Zacharaki, M. Koutroumanidis, M. Richardson, V. Megalooikonomou, "Seizure Detection using EEG and ECG Signals for Computer-based Monitoring, Analysis and Management of Epileptic Patients," Expert Systems With Applications, vol.40, pp.3227-3222, 2015.

[5] E. Pippa, E.I. Zacharaki, I. Mporas, V. Tsirka, M. Richardson, M. Koutroumanidis and V. Megalooikonomou, “Classification of Epileptic and Non-Epileptic EEG Events”, 4th International Conference on Wireless Mobile Communication and Healthcare (MOBIHEALTH), 2014

[6] E. Pippa, E. Zacharaki, I. Mporas, V. Megalooikonomou, V. Tsirka, M. Richardson, et al., "Improving Classification of Epileptic and Non-Epileptic EEG Events using Feature Selection," Neurocomputing, vol. 171, pp. 576-585, 2015.

[7] 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.

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

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

[10] 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.

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

[12] T. Lewis, “Vasovagal syncope and the carotid sinus mechanism,” vol. 1, J. Brit. Med., 1932, pp. 873-876.

[13] 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 ClinElectrophysiol., 1998, pp. 2420-5.

[14] 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, NeurophysiolClin., 2004, pp. 203-7.

[15] 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.

[16] 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, 2005.

[17] 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.

[18] G. Erus, E.I. Zacharaki and Davatzikos, “Individualized statistical learning from medical image databases: Applixation to identification of brain lesions,” vol. 18, Medical Image Analysis, 2014, pp. 542-554.

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

[20] M.J. van Putten, T. Kind, F. Visser, V. Lagerburg. “Detecting temporal lobe seizures from scalp EEG recordings: a comparison of various features” Clinical neurophysiology, pp. 2480-2489

[21] P.C. Petrantonakis and L.J. Hadjileontiadis. Emotion recognition from EEG using higher order crossings. IEEE Transactions on Information Technology in Biomedicine, pp.186-197.

[22]M. Valderrama, S. Nikolopoulos, C. Adam, V. Navarro and M. Le Van Quyen. Patient-specific seizure prediction using multi-feature and multi-modal EEG-ECG classification. XII Mediterranean Conference on Medical and Biological Engineering and Computing, pp. 77-80

[23] M. Le Van Quyen, J. Martinerie, M. Baulac and F. Varela. Anticipating epileptic seizures in real time by a non-linear analysis of similarity between EEG recordings. Neuroreport, pp. 2149-2155.

[24] 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.

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

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

[27] L. Breiman, " Random Forests, " vol. 45, Machine Learning, 2001, pp.5-32.

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

[29] 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.

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

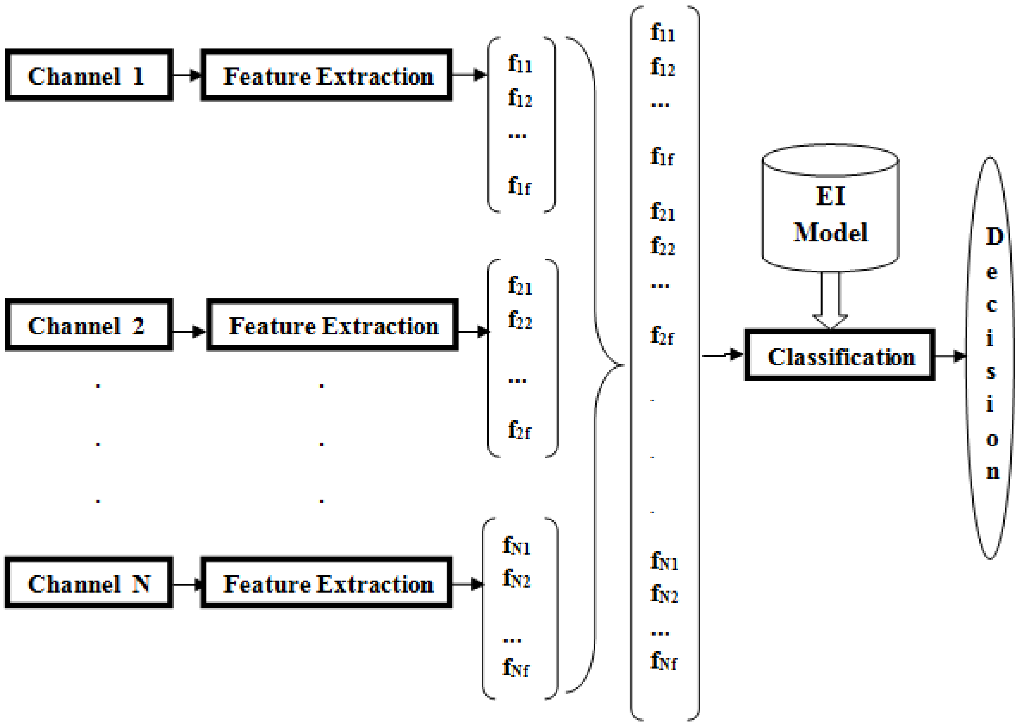

Figure 1Early Integration fusion scheme (applied to each epoch)

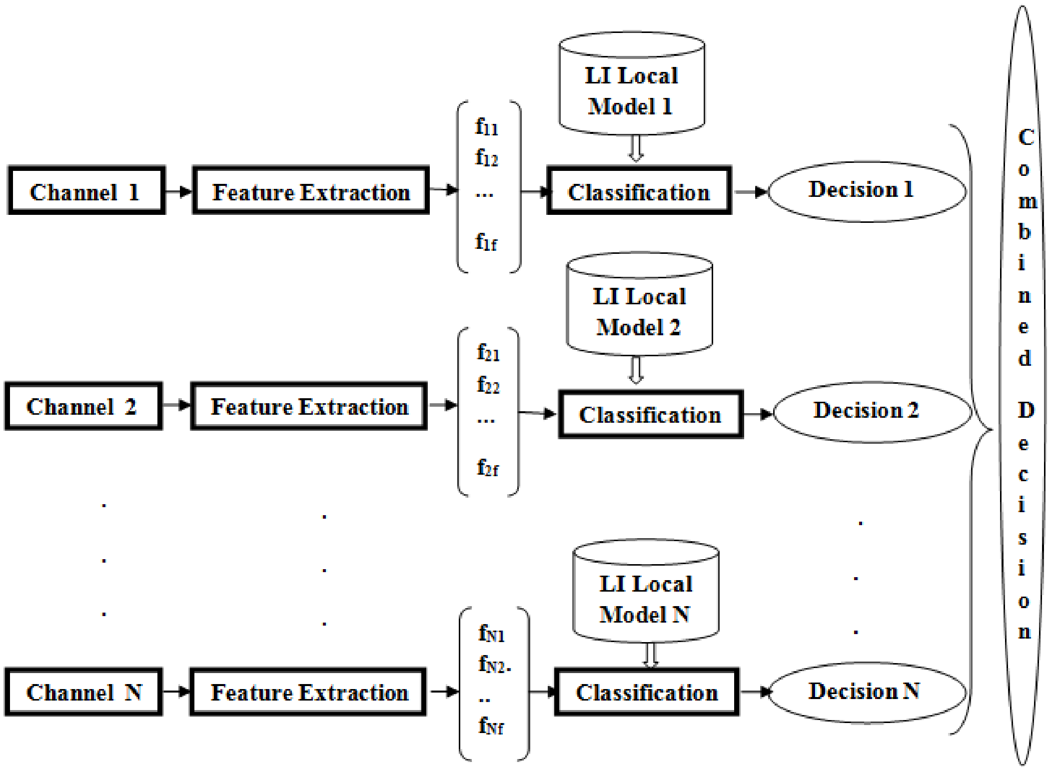

Figure 2 Late Integration with local (channel dependent) training models fusion scheme (applied to each epoch)

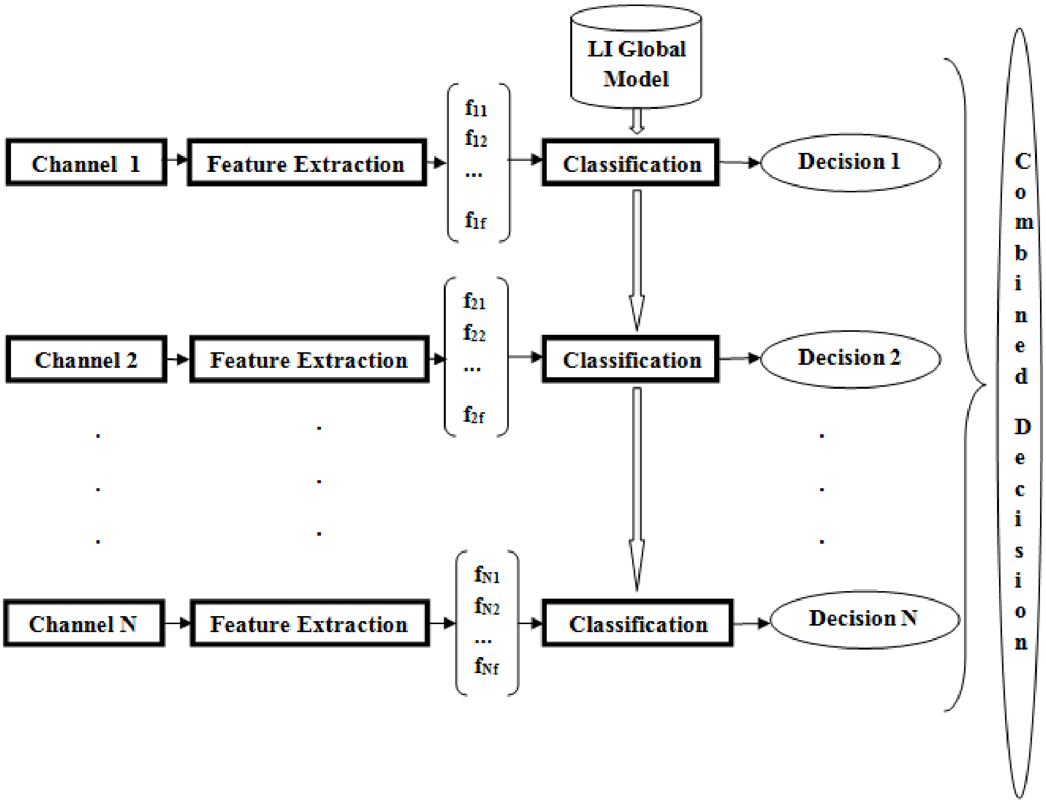

Figure 3 Late Integration with global (channel independent) training model fusion scheme

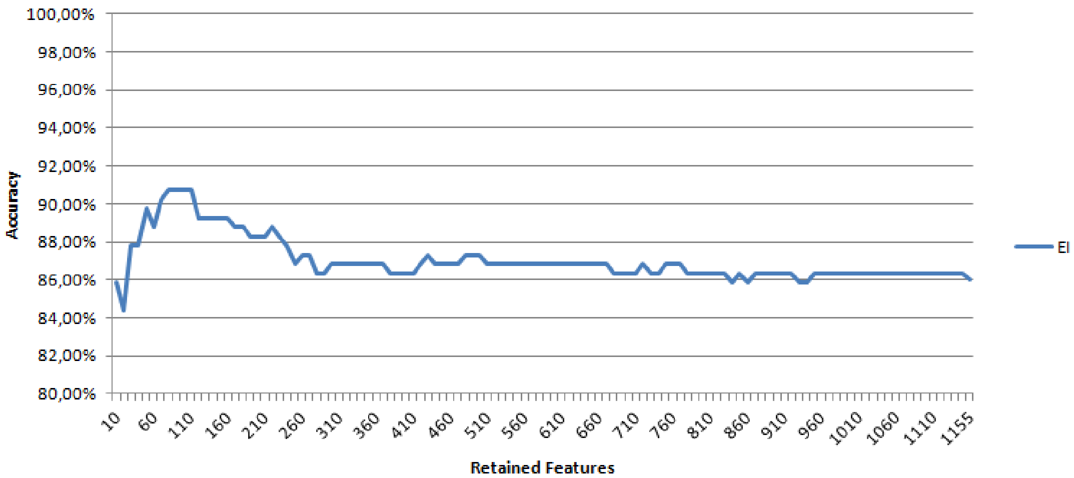

Figure 4. Classification Accuracy for the EI fusion scheme.

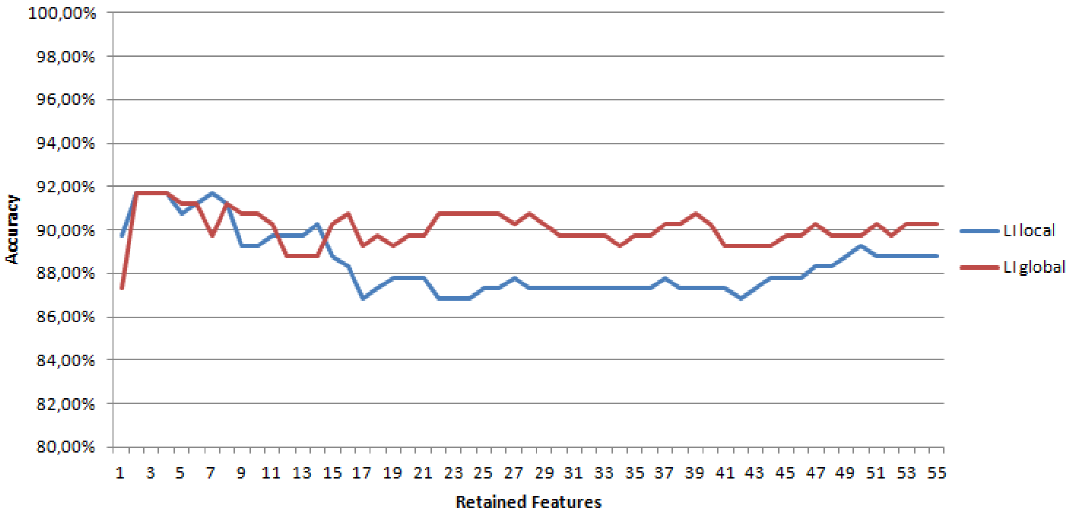

Figure 5 Classification Accuracy for LI fusion schemes.

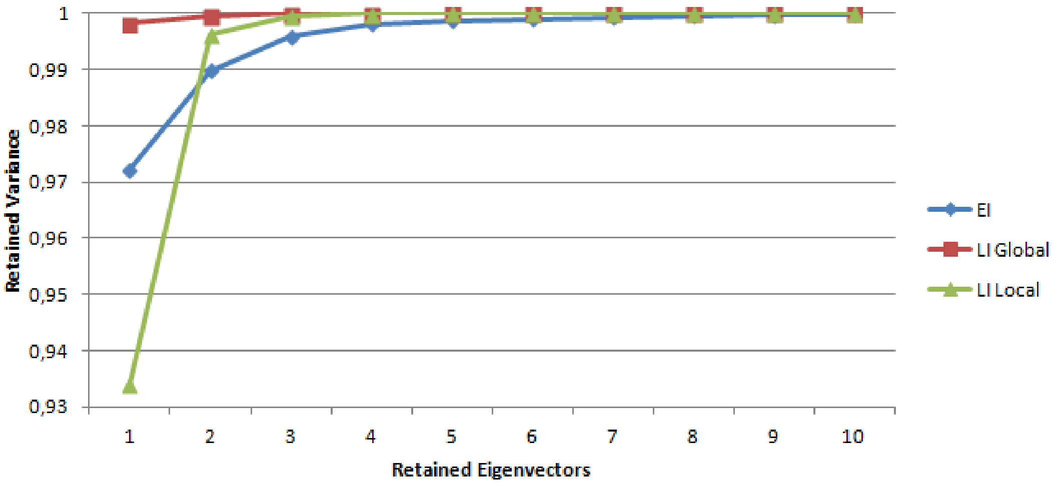

Figure 6 Retained Variance for different number of PCA retained eigenvectors. For the LI fusion scheme with local models the mean across channels variance is shown.

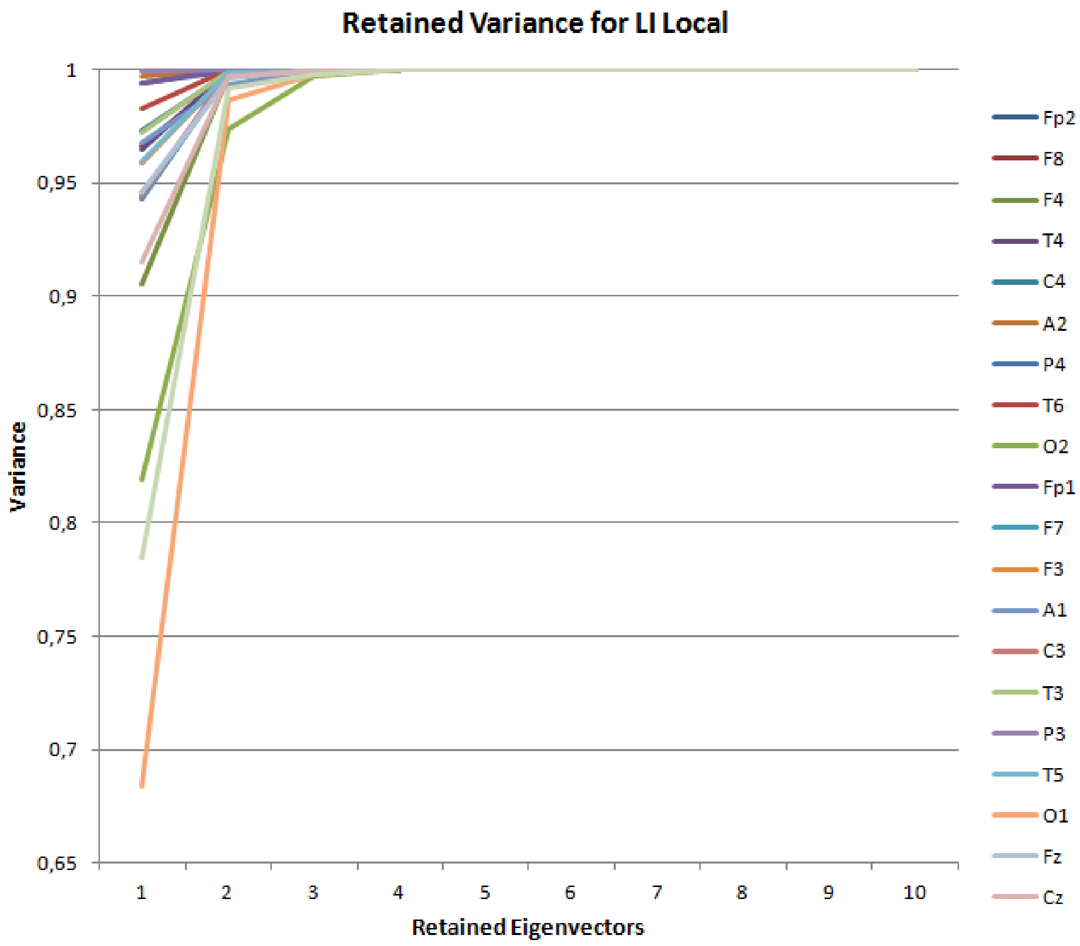

Figure 7 Retained Variance for different number of PCA retained eigenvectors for the LI fusion scheme with local models

Table 1Details on the examined subjects

Subject	Class	Sex	Date of birth	Date of test	Age	Sampling frequency
1	GSW	f	15/5/1972	8/3/2012	39	200Hz
2	GSW	f	14/9/1991	29/8/2012	20	200 Hz
3	GSW	f	21/2/1995	4/1/2012	16	500 Hz
4	GSW	m	28/7/1971	6/9/2012	40	500 Hz
5	PNES	f	22/12/1982	11/1/2012	29	500 Hz
6	PNES	f	1/1/1960	6/12/2011	51	500 Hz
7	PNES	f	8/5/1985	5/4/2012	26	500 Hz
8	PNES	f	13/10/1983	4/12/2012	29	500 Hz
9	PNES	f	10/6/1940	9/11/2009	69	500 Hz
10	VVS	f	7/6/1986	7/2/2013	26	500 Hz
11	VVS	f	19/2/1956	12/5/2013	57	500 Hz

Table 2 Extracted Features

Feature Category	Feature Description	Actual number of extracted features
Temporal Features	Minimum value	1
	Maximum value	1
	Mean	1
	Variance	1
	Standard deviation	1
	Percentiles (25%, 50%-median and 75%)	3
	Interquartile range	1
	Mean absolute deviation	1
	Range	1
	Skewness	1
	Kyrtosis	1
	Energy	1
	Shannon's entropy	1
	Logarithmic energy entropy	1
	Number of local maxima	1
	Number of local minima	1
	Zero-crossing rate	1
Spectral Features	6-th order autoregressive-filter (AR) coefficients	7
	Power spectral density	9
	Frequency with maximum amplitude	1
	Frequency with mimimum amplitude	1
	Power of continuous wavelet transform using symlet 5 mother wavelet of scale 25	1
	Power of continuous wavelet transform using symlet 5 mother wavelet of scale 32	1
	Power of discrete wavelet transform with mother wavelet function Daubechies 16 and decomposition level equal to 8	16
SUM		55

Table 3 Number of seizures and number of seizure epochs (2 seconds) per subject

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 4 Classification performance before and after feature selection by feature ranking with t-test

Fusion Scheme	Statistical Measures before Feature Selection			Statistical Measures after Feature Selection		
	Accuracy	*Sensitivity*	*Specificity*	*Accuracy*	*Sensitivity*	*Specificity*
EI	86%	92%	78%	90,73%	88,62%	93,90%
LI-local models	89%	**96%**	78%	**91,71%**	90,24%	**93,90%**
LI-global model	**90%**	94%	**84%**	**91,71%**	**91,06%**	92,68%

Table 5 Rank of the covariance matrices and number of non-zero eigenvalues for the LI scheme with local models

Channel	Rank	Non-zero eigenvalues
Fp2	35	52
F8	35	52
F4	35	52
T4	35	52
C4	36	52
A2	28	52
P4	30	52
T6	21	52
O2	36	52
Fp1	23	52
F7	33	52
F3	32	52
A1	34	52
C3	34	52
T3	30	52
P3	15	52
T5	35	52
O1	42	52
Fz	32	52
Cz	34	52
Pz	36	52
Average	32	52

Table 6 Classification performance using PCA with 100% and 99,99% retained variance

Fusion Scheme	100% variance			99,99% variance		
	Accuracy	*Sensitivity*	*Specificity*	*Accuracy*	*Sensitivity*	*Specificity*
EI	82,44%	86,99%	75,61%	80,49%	77,24%	85,37%
LI-local models	85,85%	92,68%	75,61%	89,27%	94,31%	81,71%
LI-global model	**87,80%**	**95,12%**	**76,83%**	**96,59%**	**100,00%**	**91,46%**

Table 7 Classification performance for different amounts of retained variance with respect to the EI scheme

Retained Components	*Accuracy*	*Sensitivity*	*Specificity*
1	60,98%	89,43%	18,29%
2	78,54%	82,93%	71,95%
3	75,61%	82,93%	64,63%
5	78,05%	83,74%	69,51%
11	80,49%	77,24%	**85,37%**
198	**82,44%**	**86,99%**	75,61%

Table 8 Classification performance for different amounts of retained variance with respect to the LI scheme with local model

Retained Components	*Accuracy*	*Sensitivity*	*Specificity*
1	84,39%	**94,31%**	69,51%
2	87,80%	94,31%	78,05%
3	**89,27%**	94,31%	**81,71%**
43	85,85%	92,68%	75,61%

Table 9 Classification performance for different amounts of retained variance with respect to the LI scheme with global model

Retained Components	*Accuracy*	*Sensitivity*	*Specificity*
1	85,85%	97,56%	68,29%
2	**96,59%**	**100,00%**	**91,46%**
36	87,80%	95,12%	76,83%