Journal of
Bioengineering & Biomedical Science

Mechmeche et al., J Bioengineer & Biomedical Sci 2016, 6:2
DOI: 10.4172/2155-9538.1000183

Research Article

Open Access

# Two-Stage Feature Selection Algorithm Based on Supervised Classification Approach for Automated Epilepsy Diagnosis

**Mechmeche S[1]*, Salah RB[2] and Ellouze N[1]**

*[1]National Engineering School of Tunis, University of El Manar, UR SITI, Tunis, Tunisia*
*[2]PrinceSattam Bin AbdulazizUniversity, Biomedical Technology Department, Riyadh, Saudi Arabia*

## Abstract

Epileptic diagnosis is generally achieved by visual scanning of Interictal Epileptiform Discharges (IEDs) using EEG recordings. The main objective of this research is to select a smallest relevant feature subset from the original dataset in order to reduce the diagnosis time and increase classification accuracy by removing irrelevant and redundant features. For this purpose we suggest a two-stage feature selection algorithm based on supervised classification approach adopting successively a wrapper feature selection and a wrapper feature subset selection method. Matlab simulation results illustrate that through comparing the two classifiers, the high-dimensionality is reduced at only one relevant feature that showed classification metrics of 100%. The epilepsy diagnosis is successfully tested in the discriminant Fisher-space with the single-best relevant feature.

**Keywords:** Cross-validation; Classification metrics; EEG; Feature selection; IED's; LDA; Mahalanobis distance classifier; QDA; Supervised classification

## Introduction

Epileptic is a neurological disorder marked by sudden recurrent episodes of sensory disturbance, loss of conscience, convulsions, associated with abnormal electrical activity in the brain. The confirmation of the existence of an epileptic diseases is based on visual detection of isolated Interictal Epilepti form Discharges (IEDs) (spikes or spike-waves complex), using EEG (Electroencephalogram) signal recordings in certain brain areas , for example, the confirmation of the epileptic-absence type is based on presence of a spike-waves rhythmic at 3 Hz [1-3]. This technique is inaccurate, fastidious and too time consuming. The aim of our research is to establish an automated diagnosis of epileptic disease employing a supervisedclassification approach (Figure 1).

To create a training set, we need to build a knowledge database composed of normal EEG sample and epileptic EEG sample. Feature extraction is an essential pre-processing step to pattern recognition and machine learning problems. To build the training set, the signal pattern may be described by three field analysis: Time field [4-11], frequency field [11-13], and time-frequency field [4,7,11,14-16]. In this article, EEG-signal pattern is described in high dimensionality in the three previous fields. To reduce the dimensionality at a SRFS (Smallest Relevant Feature Subset), we have proposed two-stage feature selection algorithm using wrapper-based method in supervised classification [17]: The first stage uses the IFE (individual feature evaluation) method and the second stage uses the SBS (sequential Backward Selection) method.

A Mahalanobis Distance-based Classifier (MDC) is suggested to classify the unknown EEG signal into "Normal" or "Epileptic" classes. For an optimal visualization of both of them, the samples are projected in the linear Fisher space [18,19] using Fisher linear Discriminant Analysis (FDA) that consists of seeking the optimal directions that are efficient for discrimination [20,21].

## Methods

### Knowledge database

The population selected is composed of 20 labeled single-EEG signals (derived from the Neurology department of University Hospital of Sousse-Tunisia), sampled at a frequency F = 200Hz, segmented at 1 second epoch, and filtered from artifacts, divided into two groups: 10 normal signals for the first group and 10 epileptic signals for the second group. These signals will be modeled by a set of features to form the training set that will be used in the feature selection process.

### Feature extraction

In feature extraction process, we have adopted the statistical analysis approach from each single-EEG signal. Feature vector is

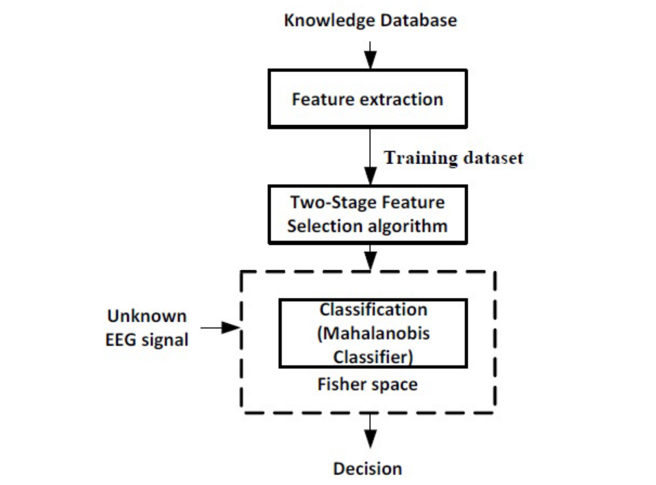

**Figure 1:** Block diagram of automatic diagnosis process.

***Corresponding author:** Mechmeche S, National Engineering School of Tunis, University of El Manar, UR SITI, Tunis, Tunisia, Tel: 21671872253; E-mail: s.biotech@laposte.net

**Received:** March 31, 2016; **Accepted:** April 13, 2016; **Published**: April 21, 2016

**Citation:** Mechmeche S, Salah RB, Ellouze N (2016) Two-Stage Feature Selection Algorithm Based on Supervised Classification Approach for Automated Epilepsy Diagnosis. J Bioengineer & Biomedical Sci 6: 183. doi:10.4172/2155-9538.1000183

composed of 48 features that are extracted from time, frequency and time-frequency fields (Table 1):

LPC: Linear Predictive Coefficients

DFTC: Discret Fourier Transformation Coefficients

CC: Cepstral Coefficients

DHTC: Discret Hilbert Transformation Coefficients

WC: Wavelet Coefficients

STFTC: Short Time Fourier Transformation Coefficients

## Training dataset

The training dataset is represented as (nxd) data pattern, it is defined as:

$$X^{TR} = \left[x_{i,k}\right], \tag{1}$$

$1 \leq i \leq n$ , $1 \leq k \leq d$

$n$: Total number of samples; d:dataset dimensionality

$x_{i,k}$: General term of training dataset

The signals are manually labeled and ordered into two groups, normal and epileptic, by an expert neurologist.

The “normal” group is defined by the following dataset:

$$X^{TR} = \left[x_{i,k}\right]_1^{TR}, \tag{2}$$

$1 \leq i \leq N_{X_1^{TR}}$

$N_{X_1^{TR}}$ : Samples number of first group

The “Epileptic” group is defined by the following dataset:

$$X_2^{TR} = \left[x_{i,k}\right]_2^{TR}, \tag{3}$$

$\left(N_{X_1^{TR}} + 1\right) \leq i \leq N_{X_2^{TR}}$

$N_{X_2^{TR}}$ : Samples number of second group

## Feature selection algorithm

For the classification difficulty, wrapper feature selection consists of selecting the features that maximize the classifier performance and capable of discriminating samples that belong to different classes. In this research, the classifier performance is evaluated from the confusion matrix that derives the important metrics, such as, Accuracy, Sensitivity and Specificity. The feature selection algorithm is composed of the two following stages (Figure 2):

| Analysis fields | Methods | Number of feature |
|---|---|---|
| Time | Min-Max | 2 |
| | Hjort parameters | 3 |
| | LPC | 4 |
| Frequency | DFTC | 3 |
| | CC | 4 |
| | DHTC | 8 |
| Time-Frequency | WC | 16 |
| | STFTC | 8 |
| | | Total: 48 |

**Table 1:** Analysis domains for feature extraction.

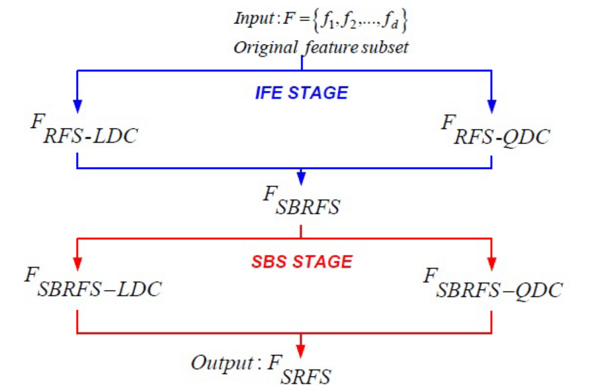

**Figure 2:** Block diagram of the relevant feature selection process.

-IFE (Individual Feature Evaluation) stage,

-SBS (Sequential Backward Selection) stage.

**Individual feature evaluation stage**: In the first algorithm stage, a wrapper feature selection method is used by applying the Individual Feature Evaluation technique. The choice of the features is accorded to the highest metrics that have been selected. Two classifiers have been evaluated for this process: LDC (Linear Discriminant Classifier) and QDC (Quadratic Discriminant Classifier) that provide the two following Relevant Feature Subsets (RFS):

$F_{RFS-\ LDC}$ : Relevant Feature Subset corresponding to the higher ranked-LDC metrics

$F_{RFS-\ QDC}$ : Relevant Feature Subset corresponding to the higher ranked-QDC metrics

In the output of the first stage, the algorithm compare between the higher ranked LDC metrics and the higher ranked QDC metrics in order to select the Smallest-Best Relevant Feature Subset $F_{SBREFS}$.

**Sequential backward selection stage:** In the second algorithm stage, to reduce the dimensionality of $F_{SBREFS}$ , a wrapper feature subset selection method is used applying Sequential Backward Selection method, consists of removing sequentially the features of the $F_{SBREFS}$ set until the removal of further features increase the classification metrics. The feature subsets according to the highest metrics have been selected to provide the two Smallest Best Relevant Feature Subsets (SBRFS):

$F_{SBRFS-\ LDC}$: Smallest-Best Relevant Feature Subset usingLDC classifier

$F_{SBRFS\ -QDC}$: Smallest-Best Relevant Feature Subset usingQDC classifier

The output of the second stage provides the smallest relevant feature subset $F_{SRFS}$ that is finally obtained by selecting the best smallest size between $F_{SBRFS-LDC}$and $F_{SBRFS-QDC}$.

**Mahalanobis distance classifier (MDC):** Mahalanobis Distance Classifier computes the distance d($x_{unk}$, $m_k$ )between unknown EEG feature vector and the two classes “Normal” and “Epileptic” as follow:

$$d(x_{unk}, m_c) = (x_{unk} - m_k)^T T^{-1} (x_{unk} - m_c) \tag{4}$$

$X_{unk}$: Unknown feature vector;

$m_k$: Mean of the kth class;

T: Covariance matrix of the learning dataset $X_{TR}$

## Results and Discussion

### First-stage experimental results

In the first part of individual feature evaluation (IFE) stage a 5-fold cross-validation procedure is used in LDC-classifier in order to estimate the metrics (Accuracy, Sensitivity and Specificity) of each feature (Figure 3). The algorithm chooses only the features having the higher metrics (Table 2 and Figure 3).

The feature subset deduced from the first stage using LDC-classifier will therefore be defined as:

$$\mathrm{F}_{FRFS-LDC}=\left\{f_{16}, f_{19}, f_{20}, f_{21}\right\}$$

In the second part of individual feature evaluation (IFE) stage, a 5-fold cross-validation procedure is applied in QDC-classifier in order to estimate the metrics of each feature (Figure 4) and the algorithm selects only the features having the higher metrics (Table 3).

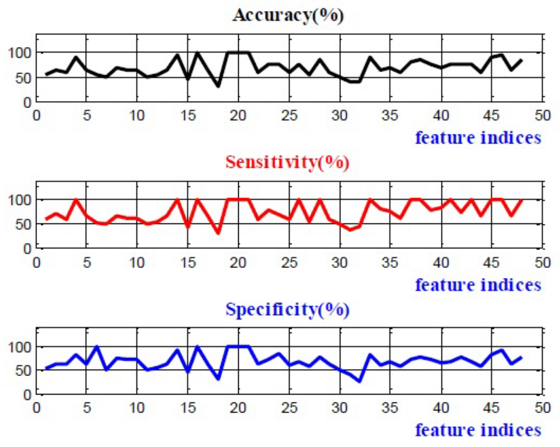

**Figure 3:** Metrics of individual features using LDC classifier.

| Top 4 feature indices | 16 | 19 | 20 | 21 |
|---|---|---|---|---|
| Accuracy | 100% | 100% | 100% | 100% |
| Sensitivity | 100% | 100% | 100% | 100% |
| Specificity | 100% | 100% | 100% | 100% |

**Table 2:** Top 4 feature indices using LDC-classifier.

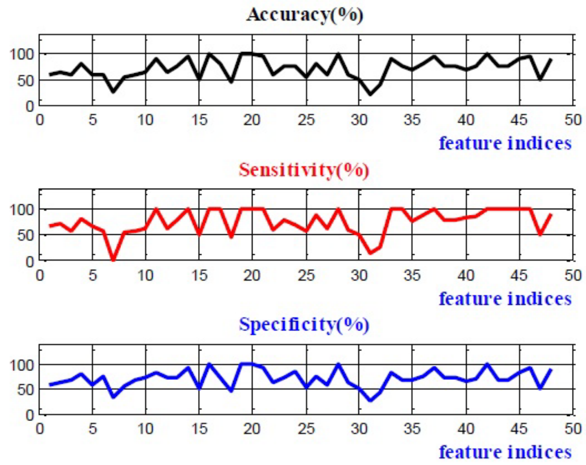

**Figure 4:** Metrics of individual features using QDC classifier.

| Feature indices | 16 | 19 | 20 | 21 | 28 | 42 |
|---|---|---|---|---|---|---|
| Accuracy | 100% | 100% | 100% | 100% | 100% | 100% |
| Sensitivity | 100% | 100% | 100% | 100% | 100% | 100% |
| Specificity | 100% | 100% | 100% | 100% | 100% | 100% |

**Table 3:** Top 6 feature indices using QDC-classifier.

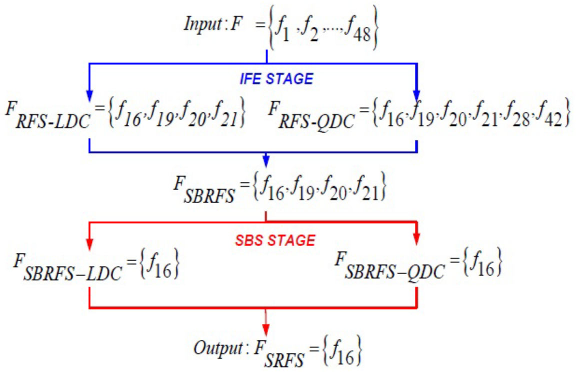
$Input: F = \{f_1, f_2, ..., f_{48}\}$
IFE STAGE
$F_{RFS\text{-}LDC} = \{f_{16}, f_{19}, f_{20}, f_{21}\}$
$F_{RFS\text{-}QDC} = \{f_{16}, f_{19}, f_{20}, f_{21}, f_{28}, f_{42}\}$
$F_{SBRFS} = \{f_{16}, f_{19}, f_{20}, f_{21}\}$
SBS STAGE
$F_{SBRFS-LDC} = \{f_{16}\}$
$F_{SBRFS-QDC} = \{f_{16}\}$
$Output: F_{SRFS} = \{f_{16}\}$

**Figure 5:** Experimental results of relevant feature selection process.

The feature subset deduced from the first stage using QDC-classifier will therefore be defined as:

$$\mathrm{F}_{FRFS-QDC}=\left\{f_{16}, f_{19}, f_{20}, f_{21}, f_{28}, f_{42}\right\}$$

At the end of the first algorithm stage the smallest best relevant feature subset (SBRFS) have been selected by comparing between the metrics and the size of both $F_{RFS-LDC}$and $F_{RFS-QDC}$subsets, the SBRFS will therefore be defined as:

$$\mathrm{F}_{SBRFS}=\left\{f_{16}, f_{19}, f_{20}, f_{21}\right\}$$

### Second-stage experimental results

To reduce the dimensionality of $F_{SBRFS}$ we have used the SBS (Sequential Backward Selection) search method that starts with all features and removes a single feature at each step until the desired dimension with the highest metrics is reached. For each step a 5-fold cross validation is applied for the feature subset selection process. In the first part of the second-algorithm stage, the experimental results using LDC-classifier illustrates that the SBRFS (Smallest Best Relevant Feature Subset) is composed of the 16th feature:

$$F_{SBRFS\text{ - }LDC}=\left\{f_{16}\right\}$$

In the second part of the second-algorithm stage, the experimental results using QDC-classifier illustrates that the SBRFS (Smallest Best Relevant Feature Subset) is also composed of the 16th feature:

$$F_{SBRFS\text{ - }QDC}=\left\{f_{16}\right\}$$

The output of the second stage selects the smallest relevant feature subset comparing both the metrics and thesize of $F_{SBRFS-LDC}$ and $F_{SBRFS-QDC}$,

The final SRFS (Smallest Relevant Feature Subset) is so deduced as: $F_{SRFS}= \{f_{16}\}$

The final experimental result of the two-stage algorithm feature selection is resumed in the following figure (Figure 5).

The combination of these two techniques (IFE and SBS) leads to reduce the dimensionality of the original feature set at only one best relevant feature that will be used in epilepsy diagnosis.

The diagnostic result was successfully tested (Figure 6) on EEG

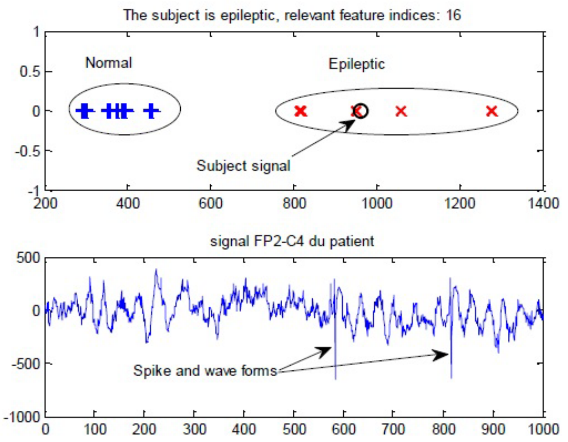

**Figure 6:** Automated diagnosis of the single-EEG signal.

| Classifier | Accuracy | Sensitivity | Specificity |
|---|---|---|---|
| AdaBoost | 93,9% | 95,5% | 92,4% |
| NN | 99% | __ | __ |
| LDC | __ | 82% | 90% |
| QDC | __ | 87% | 92% |
| LDC | 100% | 100% | 100% |
| QDC | 100% | 100% | 100% |

**Table 4:** Literature review of some classification metrics.

signals containing spikes and spike-waves, this figure gives an example of automatic affectation (using a Mahalanobis distance classifier) of an EEG signal that containing two spike-waves (Epileptic). This diagnostic is made using only one feature (16th feature) that has been selected from the dataset. Index 16 is accorded to the maximum of the DHTC magnitude of EEG signal that isdefined as: max ($|DHT(S(n)|$).

## Literature Review

Table 4 show a comparative study on IED's classification metrics in recent years, regardless of the number of features used: Our feature selection algorithm improves the classification metrics for both LDC and QDC classifier using the single-best relevant feature selected (Table 4).

## Conclusion

A two-stage feature selection algorithm has been proposed in this article in order to remove the redundancy and to reduce the dimensionality of the dataset at the relevant feature subset. The mRMR (Minimum-Redundancy Maximum-Relevance) approach was successfully confirmed and tested in the first algorithm stage using IFE method, and the dimensionality of the relevant feature subset selected was successfully reduced in the second stage using SBS method at only one single best relevant feature that may be reduce considerably the processing time of the diagnostic. The performance of the results can be improved by using other robust dataset features and other classifier types for validation, such as the ANN (Artificial Neural Network), SVM (Support Vector Machine) and GA (Genetic Algorithm) methods. Using the automated IED's diagnosis the doctor will no longer need to scan visually EEG signal leads.

### References

1. Hirsch E, Panayiotopoulos Cp (2005) Childhood absence epilepsy and related syndromes. Epileptic Syndromes in Infancy, Childhood and Adolescence. (4th edtn). Montrouge, France: John Libbey Eurotext.
2. Kakisaka Y, Alexopoulos AV, Gupta A, Wong ZI, Mosher JC, et al. (2011) Generalised 3Hz spike-and-wave complexes emanating from focal epileptic activity in pediatric patients. Epilepsy & Bihavior 20:103-106.
3. Fergus P, Hignett D, Hussain A, Jumeily DA, Aziz KA (2015) Automated Epileptic Seizure Detection Using Scalp EEG and Advanced Artificial Intelligence Techniques. BioMed Research International 2015.
4. Karlik B, Hayta SB (2014) Comparison Machine Learning Algorithm for Recognition of Epileptic Seizure in EE. Proceeding IWBBIO.
5. Guoham Z, Yan L, Peng W (2014) Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal. IEEE Journal of Biomedical and Health informatics 18: 1813-1821.
6. Guarnizo C, Delgado E (2010) EEG Single-channel seizure recording using Empirical Mode Decomposition and normalized mutual information. International Conference on signal Processing, Proceeding, ICSP, Beijing.
7. Srinivasan V, Eswaran C, Sriraam N (2005) Artificial neral network based epileptic detection using time domain and frequency domain features. Journal of Medical System 29: 647-660.
8. Cecchin T, Ranta R, Koessler L, Casparay O, Vespigani H, et al. (2016) Seizure Lateralisation in scalp EEG using Hjort Technology International. Clinical Neurophysiology 121: 290-300.
9. Khanwani P, Ridhar S, Vijaylakshmi K (2010) Automated Event Detection of Epileptic Spikes Using Neural Networks. International Journal of Computer Application 2.
10. Oikonomou VP, Tzallas AT, Fotiadis DI (2007) A Kalman filter based methodology for EEG Spike enhancement. Computer Methods and Programs in Biomedecine 85: 101-108.
11. Rasekhi J, Mollaei MR, Bandarabadi M, Teixeira CA, Dourado A (2015) Epileptic Seizure Prediction based on Ratio and Differential Linear Univariate features. Journal of Medical Signals and Sensors 5.
12. Polat K, Gunes S (2007) Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. Applied Mathematics and Computation 187: 1017-1026.
13. Kamath C (2013) Comparison of Baseline Cepstral Vector and Composite Vectors in the Automatic Seizure Detection Using Probabilistic Neural Networks. Hindawi Publishing Corporation, ISRN Biomedical Engineering 2013.
14. Mahajan K, Vargantwar MR, Rajput SM (2011) Classification of EEG using PCA, ICA and Neural Network. International Journal of Engineering and Advanced Technology 1: 80-83.
15. Laxman T, Warpe H (2011) Detection of Epilepsy Disorder Using Discret Wavelet Transform using Matlab. International Journal of advanced Science and Technology 28: 17-24.
16. Nijsen TME, Cluitmans PJM, Griep PAM, Aarts RM (2006) Short Time Fourier and Wavelet Transform for Accelerometric Detection of Myoclonic Seizures. Belgian Day on Biomedical Engineering IEEE/EMBS Benelux Symposium.
17. Guyon I and Elisseeff A (2003) An Introduction to Variable and feature Selection. Journal of Machine Learning Research 3: 1157-1182.
18. Belhumeur P, Hespanha JP, Kriegman DJ (1997) Eigenfaces vs Fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence 19: 711-720.
19. Zhao W, Chellappa R, Krishnaswamy A (1998) Discriminant Analysis of Principal Components for Face recognition. Proc of the 3th IEEE International Conference on Automatic Face and Gesture Recognition Japan.
20. Liu YC, Lin CCK, Tsai JJ, Sun YN (2013) Model-Based Spike Detection of Epileptic EEG Data. Sensors 13:12536–12547.
21. Nassim B (2016) Combined Odd Pair Autoregressive Coefficients for Epileptic EEG Signals Classification by Radial Basis Function. World Academy of Science, Engineering and Technology International Journal of Biomedical and Biological Engineering 3.