BULETINUL INSTITUTULUI POLITEHNIC DIN IAŞI
Publicat de
Universitatea Tehnică „Gheorghe Asachi” din Iaşi
Volumul 66 (70), Numărul 1, 2020
Secţia
ELECTROTEHNICĂ. ENERGETICĂ. ELECTRONICĂ

COMBINATION OF FEATURES EXTRACTED FROM ELECTROENCEPHALOGRAPHIC SIGNAL – A CASE STUDY

BY

OANA-DIANA HRIŞCĂ-EVA[*] and ANCA MIHAELA LAZĂR

"Grigore T. Popa" University of Medicine and Pharmacy, Iași,
Faculty of Medical Bioengineering

Received: September 3, 2020
Accepted for publication: October 5, 2020

Abstract. This work proves the usefulness of bringing together features extracted from electroencephalographic (EEG) data in several classifiers when a motor imagery task was performed. Autoregressive model, amplitude modulation and phase synchronization methods were used for selecting the features in order to form the hybrid combination for the feature vector. In the classification phase, linear discriminant analysis, quadratic discriminant analysis, Mahalanobis distance, *k* nearest neighbors and support vector machine as classifiers were verified. The results reveal that the combination of the phase locking value and phase lag index as measures for phase synchronization with support vector machine classifier can effectively improve the classification performance and outperforms all other tested combinations.

Keywords: motor imagery; brain computer interface; feature vector; classification algorithms.

1. Introduction

Brain-computer interfaces (BCIs) are devices that interface with the brain signals to enable interaction with the environment. BCIs have the potential to improve the quality of life for many people affected by disorders of

[*]Corresponding author; *e-mail*: oana.hrisca@umfiasi.ro

the brain, spine or limbs through direct interface with the nervous system (Hughes *et al*., 2020). BCIs have been used in rehabilitation, communication, in computer gaming and wheelchair locomotion (Lucas *et al.*, 2018). Electroencephalogram (EEG) records the brain activity from the scalp using electrodes. EEG-based BCI systems are widely used because they are simple and inexpensive compared to other systems developed for recording brain signals. The majority of the EEG based BCI systems handles the potential P300 (Guger *et al.*, 2009) and sensorimotor rhythms (SMR) (Yuan and He, 2014). Motor imagery (MI) is the cognitive process of imagining the movement of a body part without actually moving it (Aggarwal and Chugh, 2019). When motor information is processed, a decrease of sensorimotor rhythm amplitude can be observed (desynchronization). When there is no motor activity – the sensorimotor brain areas are at rest or inhibited – the amplitude of the SMR is high (synchronization). These pattern changes in SMR amplitude can be used to trigger an external device in order to display real-time sensory feedback or to execute the intended action (Van Dokkum *et al.*, 2015).

Motor imagery based BCI provides an interface for the patients with motor impairment or for those in completely locked-in-state to interact with the environment by controlling external devices.

The purpose of this research is to investigate which combination of features and which classifier are more appropriate to be used in a motor imagery paradigm with EEG signals recorded from a single subject during a period of eight consecutive years.

The paper is structured in five sections: the EEG dataset is detailed in Section 2, the used methods are described in Section 3, the results and conclusions in Section 4, and Section 5, respectively.

2. EEG Signals

The EEG signals were recorded from a single subject during the period 2012 – 2019. The brain activity was acquired with a sampling frequency of 256Hz, using gMobilab+ module (gtec, 2020) and BCI 2000 platform (Mellinger and Schalk, 2007) in The Signal Processing Lab from the Faculty of Medical Bioengineering. The EEG recordings were attained by positioning EEG electrodes over the brain scalp in a standardized 10–20 electrode system. The handled electrodes were CP3, CP4, P3, C3, Pz, C4, P4 and Cz and the reference electrode was placed on the right ear. The subject had to perform the following motor imagery paradigm: when a right/left arrow appeared on the computer screen the subject had to imagine the hand movement indicated by the arrow. The white screen indicated that the subject had to relax. 30 arrows appeared randomly. The subject was trained before each session by performing a session which involved the real movement of the hands. The sessions were performed in different days, months and years.

Session 1 was recorded in 2012, sessions 2 and 3 in 2013, session 4 in 2014, session 5 in 2015, sessions 6, 7, 8 in 2016, session 9 in 2017, session 10 in 2018 and session 11 in 2019.

3. Methods

Autoregressive (AR) process, phase synchronization (PS) and amplitude modulation (AM) were used as relevant methods for feature extraction. The measures for each used method were:

- Itakura Distance (ID) and symmetric Itakura Distance (SID) for AR process;
- phase locking value (PLV) and phase lag index (PLI) for PS;
- the amplitude modulation energy index (AMEI) for AM.

The methods are explained in detail in (Kong *et al.*, 1995; Estrada *et al.*, 2009; Gysels and Celka, 2004; Gonuguntla *et al*, 2013; Stam *et al.*, 2007; Eva and Lazăr, 2019; Hrişcă-Eva and Lazăr, 2020).

The frequency band 8 - 12 Hz corresponding to Alpha rhythm and all eight channels were taken into consideration for further processing. Three datasets were formed: $y_{RIGHT}(n)$ for EEG signal corresponding to right hand movement imagination, $y_{LEFT}(n)$ for EEG signal corresponding to left hand movement imagination and $y_{REST}(n)$ for EEG signal corresponding to relaxation state.

The EEG signal, $y(n)$, can be expressed as the output of an AR process:

$$y(n) = -\sum_{k=1}^{p} a_k y(n-k) + n(n), \tag{1}$$

where a_k are the parameters of the model, p is the model order and $n(n)$ the unpredictable part of the EEG signal $y(n)$.

It is shown that the minimum of mean square error (MSE_y) (Kong *et al.*, 1995):

$$MSE_y = a^T R_y(p) a, \tag{2}$$

where $a = \left[1\ a_1\ a_2 \dots a_p\right]^T$, $R_y(p)$ is the autocorrelation matrix of $y(n)$, and T the transpose of a matrix, leads to the optimum AR model.

If $y_{REST}(n)$ passes through $AR(p)$ model characterized by the a^{REST} parameters, the minimum MSE is defined by:

$$MSE_{y_{REST}, a^{REST}} = (a^{REST})^T R_{y_{REST}}(p) a^{REST}, \tag{3}$$

and if the same signal passes through AR(p) model, caracterized by a^{RIGHT} results the MSE as follows:

$$MSE_{y_{REST},a^{RIGHT}} = (a^{RIGHT})^T R_{y_{REST}}(p) a^{RIGHT}, \quad (4)$$

ID for right hand movement imagination is:

$$ID_{REST-RIGHT} = \log\left(\frac{MSE_{y_{REST},a^{RIGHT}}}{MSE_{y_{REST},a^{REST}}}\right). \quad (5)$$

In a similar manner, for $y_{LEFT}(n)$, the ID for left hand movement imagination $ID_{REST-LEFT}$ is defined.

The symmetric Itakura distance, SID_{RIGHT}, for right hand movement imagination (Estrada *et al.*, 2009) is:

$$SID_{RIGHT} = \frac{1}{2}(ID_{REST-RIGHT} + ID_{RIGHT-REST}), \quad (6)$$

SID_{LEFT} for left hand movement imagination can be obtaind in an identical way.

The model order *p=6* and model order *p=10* were used for IDs and SIDs.

For AM analysis, the temporal amplitude envelope is calculated using Hilbert transform $\mathcal{H}\{.\}$ (Gysels and Celka, 2004).

The analytic signal $y(n)_a$ is computed by:

$$y(n)_a = y(n) + j\mathcal{H}\{y(n)\} \quad (10)$$

The amplitude modulation (or the temporal amplitude envelope) for $y(n)_a$, expressed by $y_{AM}(n)$, is denoted:

$$y_{AM}(n) = \sqrt{y(n)^2 + \mathcal{H}\{y(n)\}^2}. \quad (11)$$

$y_{AM}(n)$ is multiplied by a 5 s Hamming window with 0.5 s delay and is obtained the temporal envelope for frame m, expressed by $y_{AM}(m,n)$.

The modulus of the Fourier transform of the amplitude modulation for frame m is computed:

$$y_{AM}(m,f) = |\mathcal{F}\{y_{AM}(m,n)\}|, \quad (12)$$

where f is the modulation frequency and $\mathcal{F}\{y_{AM}(m,n)\}$ is the discrete Fourier transform of $y_{AM}(m,n)$.

The energy of the j modulation band, denoted by $E_j(m,f)$, is computed as:

$$E_j(m,f) = y_{AM_j}(m,f)^2 \tag{13}$$

and the average of energies over all the frames, expressed by $\overline{E_j(m,f)}$.

The amplitude modulation energy index of the j modulation band (Eva and Lazăr, 2019), $AMEI_j(f)$, is defined by the following expression:

$$AMEI_j(f) = \frac{\overline{E_j(m,f)}}{\sum_{j=1}^{k} \overline{E_j(m,f)}}, \tag{14}$$

where k is the number of the modulation bands for the rhythm taken into account. For *Alpha* rhythm, k=3, because there are possible delta, theta, and alpha modulation bands. As the best results were obtained for *Theta* and *Alpha* modulation bands (Eva and Lazăr, 2019), only these cases were taken into account. The symbolization is *modulation band_EEG rhythm* (*e.g. Alpha_Alpha* and *Theta_Alpha*). The calculus was applied for $y_{RIGHT}(n)$ and for $y_{LEFT}(n)$ and two $AMEIs$ were achieved.

PLV value is expressed as:

$$PLV = \left|\langle e^{j\,\Delta\varphi(t)}\rangle\right|, \tag{15}$$

where $\varphi_x(t)$ and $\varphi_y(t)$ are instantaneous phases of $x(n)$ and $y(n)$ EEG signal, respectively, and $\Delta\varphi(t) = \varphi_y(t) - \varphi_x(t)$.

PLI (Stam *et al.*, 2007) is stated as:

$$PLI = |\langle sign[\Delta\varphi(t_k)\,]\rangle|\,, \tag{16}$$

$sign$ is the signum function and $\langle . \rangle$ denotes the average over the time.

In order to calculate the phase synchronization parameters, four combinations for EEG signals, Cz-C3, Cz-C4, Pz-C3 and Pz-C4, were taken into account. The above described steps were conducted for right and left hand movement imagination.

Discrimination between right and left motor imagery was performed using linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), Mahalanobis distance, k nearest neighbors (kNN) with k=5 and support vector machine (SVM) (Lotte *et al.*, 2018). The k fold cross validation (k=5) was used for estimating the classification rates.

4. Results

Eight features vectors (FV), denoted from A to H, were proposed using the combinations of the mentioned features:

- FV A - AMEIs for Alpha_Alpha and Theta_Alpha;
- FV B - PLV and PLI;
- FV C - IDs model orders 6 and 10 ;
- FV D - SIDs model orders 6 and 10;
- FV E - IDs model orders 6 and 10, AMEIs for Alpha_Alpha and Theta_Alpha;
- FV F - PLV, PLI, AMEIs for Alpha_Alpha and Theta_Alpha;
- FV G - PLV, PLI, IDs model orders 6 and 10;
- FV H - PLV, PLI, SIDs model orders 6 and 10.

In Table 1 to Table 8 are displayed the classification rates achieved with feature vector A to feature vector H, respectively. The best results are marked with dark blue.

Table 1
Classification Rates (%) for Feature Vector A

Session	LDA	QDA	MD	kNN	SVM
1	50	65.5	69.5	93.38	66.5
2	53.25	60	74	94.96	70.75
3	51.25	66.75	67.75	94	61.75
4	51.75	71.5	89.5	97.46	80
5	49.5	65	66.5	93.88	65.25
6	63.5	72.25	89.75	98.04	90
7	52.5	65.5	65	93.29	79.75
8	52.25	59	70.5	93.63	62.5
9	50.5	72	83	95.54	78.25
10	52.5	66.25	66.75	93.71	58.75
11	51.75	61.75	59.75	92.75	67.25

Table 2
Classification Rates (%) for Feature Vector B

Session	LDA	QDA	MD	kNN	SVM
1	66.24	82	81.87	96.9	99.03
2	62.41	82.6	81.14	96.75	99.21
3	68.13	77.8	77.01	95.78	99.33
4	57.66	82.91	82.12	96.68	98.97
5	68.92	77.68	77.01	95.95	99.15
6	58.03	80.41	80.23	96.5	98.78
7	64.96	82.24	82.24	97	99.51
8	66.36	83.33	82.42	96.84	98.72
9	66.73	81.02	80.72	96.7	99.39
10	61.01	75.79	74.15	95.49	98.84
11	68.07	79.87	79.74	96.54	99.76

Table 3
Classification Rates (%) for Feature Vector C

Session	LDA	QDA	MD	kNN	SVM
1	55.83	100	100	87.36	82.5
2	64.17	99.17	99.17	92.22	86.67
3	50	95.83	95.83	90.83	86.67
4	73.33	93.33	90.83	84.17	81.67
5	61.67	95	95	88.47	85.83
6	91.67	96.67	96.67	98.89	93.33
7	62.5	95	91.67	90.56	89.17
8	74.17	96.67	95	92.64	89.17
9	77.5	93.33	92.5	93.33	94.17
10	81.67	99.17	99.17	97.22	96.67
11	80	97.5	91.67	94.31	95

Table 4
Classification Rates (%) for Feature Vector D

Session	LDA	QDA	MD	kNN	SVM
1	54.17	100	100	89.44	84.17
2	66.67	100	100	94.72	93.33
3	49.17	94.17	94.17	89.86	89.17
4	75	86.67	86.67	85.14	86.67
5	63.33	96.67	96.67	91.67	90.83
6	86.67	98.33	98.33	95.83	88.33
7	61.67	96.67	95.83	92.5	85.83
8	67.5	97.5	97.5	95.28	84.17
9	83.33	96.67	96.67	95.97	93.33
10	75	98.33	98.33	95.69	99.17
11	90	96.67	95.83	95.56	95

The highest classification rate of 98.04% was obtained with classifier kNN (Table 1). SVM attained the best classification rate of 99.76% for feature vector B and for many sessions the classification rates were above 99% using SVM (Table 2). In Table 3 is shown that the maximum classification rates were achieved for feature vector C using QDA and MD classifiers. The same remarks as for feature vector C, are for the classification rates achieved with feature vector D (Table 4).

Table 5
Classification Rates (%) for Feature Vector E

Session	LDA	QDA	MD	kNN	SVM
1	52.88	56.35	67.12	91.06	64.23
2	52.69	55.77	65.77	91.6	64.81
3	52.88	58.27	66.35	91.35	63.27
4	52.88	58.27	66.54	90.45	61.15
5	52.88	55.19	67.69	91.54	60.58
6	52.69	56.35	66.35	93.01	63.65
7	52.5	53.65	67.88	91.6	63.65
8	52.5	56.15	64.23	90.96	62.12
9	52.88	53.46	67.5	91.96	63.85
10	52.88	55.96	66.92	92.88	60.96
11	52.31	56.54	65.77	92.21	62.88

Table 6
Classification Rates (%) for Feature Vector F

Session	LDA	QDA	MD	kNN	SVM
1	50.24	52.74	53.28	90.89	55.92
2	50.78	52.84	53.13	91.03	56.65
3	51.08	53.13	53.86	91.27	55.63
4	51.13	55.68	56.95	91.73	58.56
5	50.24	52.94	53.77	91.38	54.21
6	54.89	57.19	57.73	92.12	58.37
7	50.49	52.98	53.18	91.69	57.14
8	50.64	51.61	52.79	91.4	54.45
9	51.47	54.5	55.48	91.3	58.02
10	50.49	52.94	51.42	91.09	52.89
11	49.56	52.15	53.42	91.46	54.35

Table 7
Classification Rates (%) for Feature Vector G

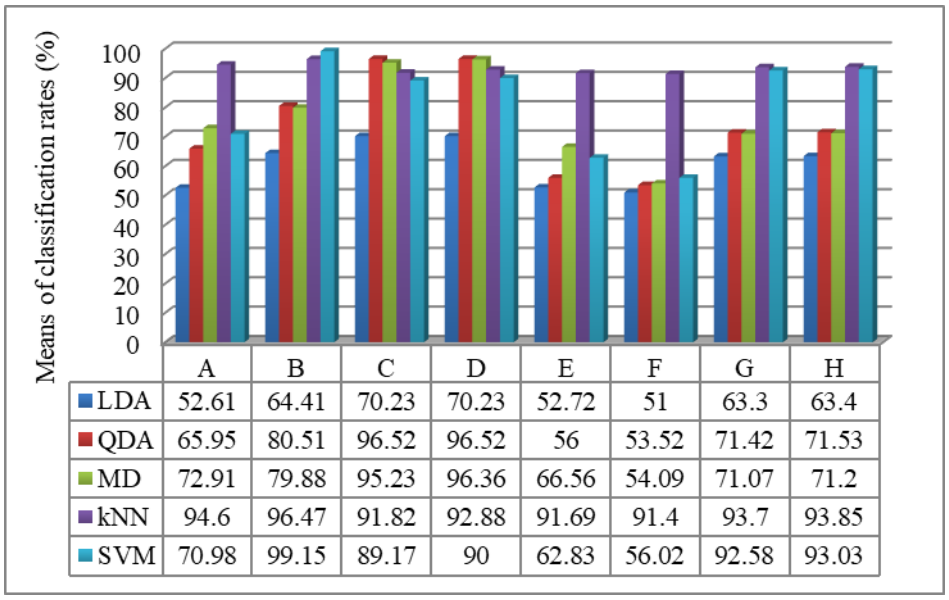
Session	LDA	QDA	MD	kNN	SVM
1	64.29	73.7	73.64	93.87	92.18
2	61.9	71.15	69.39	93.33	91.84
3	66.5	69.5	69.22	93.3	90.99
4	57.82	70.18	69.44	92.66	90.14
5	66.89	69.16	69.16	93.23	90.65
6	57.99	70.29	70.41	93.91	93.42
7	63.1	70.63	69.5	93.69	94.95
8	67.18	75.17	75.96	94.78	93.65
9	64.91	70.52	70.07	93.17	93.14
10	60.6	71.77	70.52	94.18	93.25
11	65.08	73.58	74.49	94.62	94.22

Table 8
Classification Rates (%) for Feature Vector H

Session	LDA	QDA	MD	kNN	SVM
1	64.34	72.68	73.13	93.77	92.57
2	62.07	71.15	68.99	93.45	92.12
3	65.7	69.84	69.16	93.14	91.55
4	58.11	71.37	70.41	93.36	92.12
5	66.44	69.16	68.31	93.24	91.27
6	58.56	69.61	70.69	94.03	94.33
7	62.81	70.46	69.73	94.12	95.12
8	67.35	75.51	76.36	94.87	93.42
9	64.74	71.26	70.18	93.67	93.37
10	61.05	71.15	71.03	94.11	92.86
11	66.21	74.66	75.17	94.6	94.61

For feature vectors E and F the highest classification rates were obtained with kNN classifier (Table 5 and Table 6). The highest classification rate using feature vector G was achieved with SVM classifier (Table 7), and for feature vector H for kNN classifier (Table 8).

In Fig. 1 are displayed the means of classification rate for each features vector.



	A	B	C	D	E	F	G	H
LDA	52.61	64.41	70.23	70.23	52.72	51	63.3	63.4
QDA	65.95	80.51	96.52	96.52	56	53.52	71.42	71.53
MD	72.91	79.88	95.23	96.36	66.56	54.09	71.07	71.2
kNN	94.6	96.47	91.82	92.88	91.69	91.4	93.7	93.85
SVM	70.98	99.15	89.17	90	62.83	56.02	92.58	93.03

Fig. 1 − The means of classification rates for each features vector from A to H.

The most consistent results for all feature vectors used were obtained using kNN classifier. The smallest mean of classification rates (51%) was

obtained with LDA classifier and feature vector F. The highest mean of classification rates was achieved for feature vector B and SVM classifier.

5. Conclusions

The classification rates have not been better in recent sessions than in the first ones. So, it can be concluded that the training of the subject did not determine an improvement of the performances.

The proposed combination of PLV and PLI phase synchronization extracted features with SVM classifier get the best mean of the classification rate from all the other cases. The mentioned hybrid combination significantly discriminates the two classes with the mean of classification rate of 96.47%, while in (Hrişcă-Eva and Lazăr, 2020), when one individual feature was employed, the maximum value of the mean of classification rate attained only 80% (using AMEIs for Alpha_Alpha with kNN classifier).

In (Hrişcă-Eva and Lazăr, 2020), the highest classification rate was 93.33% using ID (no matter the order 6 or 10) and MD classifier. In the present paper, the maximum classification rate of 100% was obtained with features vectors combining ID model orders 6 and ID model order 10 and SIDs model orders 6 and 10.

Overall, the classification rates using hybrid combinations of features are higher than the classification rates obtained by applying individual features from the sets of FVs.

Results show that the proposed approach enhances the classification rate and that the SVM classifier performs very well in classifying EEG signals using the combination strategy of feature extraction.

The future work implies recording EEG signals from other subjects for long period of times, searching for an adaptive method for selecting the suited feature combination and looking for other possible combinations of features.

REFERENCES

Aggarwal S., Chugh N., *Signal Processing Techniques for Motor Imagery Brain Computer Interface: A Review*, Array, 2019, 1, 100003.

Estrada E., Nazeran H., Ebrahimi F., Mikaeili M., *Symmetric Itakura Distance as an EEG Signal Feature for Sleep Depth Determination*, In Summer Bioengineering Conference, American Society of Mechanical Engineers, 2009, **48913**, 723-724.

Eva O.D., Lazăr A.M., *Amplitude Modulation Index as Feature in a Brain Computer Interface*, Traitement du Signal, 2019, **36**, *3*, 201-207.

Gonuguntla V., Wang Y., Veluvolu K.C., *Phase Synchrony in Subject-Specific Reactive Band of EEG for Classification of Motor Imagery Tasks*, 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2013, 2784-2787.

Guger C., Daban S., Sellers E., Holzner C., Krausz G., Carabalona R., Edlinger, G., *How Many People are Able to Control a P300-Based Brain–Computer Interface (BCI)?*, Neuroscience Letters, 2009, **462**, *1*, 94-98.

Gysels E., Celka P., *Phase Synchronization for the Recognition of Mental Tasks in a Brain-Computer Interfa*ce, IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2004, **12**, *4*, 406-415.

Hrișcă-Eva O.D., Lazăr A.M., *Feature Extraction and Classification Methods for EEG Signals in a Motor Imagery Task – A Study Using the Subject's Recordings from Several Years*, Traitement du Signal, 2020, Unpublished.

Hughes C., Herrera A., Gaunt R., Collinger J., *Bidirectional Brain-Computer Interfaces*, In Handbook of Clinical Neurology, Elsevier, 2020, **168**, 163-181.

Kong X., Thakor N., Goel V., *Characterization of EEG Signal Changes via Itakura Distance*, In Proceedings of 17th International Conference of the Engineering in Medicine and Biology Society, 1995, 2, 873-874.

Lotte F., Bougrain L., Cichocki A., Clerc M., Congedo M., Rakotomamonjy A., Yger F., *A Review of Classification Algorithms for EEG-Based Brain–Computer Interfaces: A 10 Year Update*, Journal of Neural Engineering, **15**, *3*, 031005 (2018).

Mellinger J., Schalk G., *BCI2000: A General-Purpose Software Platform for BCI Research*, In: G. Dornhege, J. del R. Millán, T. Hinterberger, D.J. McFarland, K.-R. Müller (Eds.), Toward Brain-Computer Interfacing, MIT Press, 2007.

Stam C.J., Nolte G., Daffertshofer A., *Phase Lag Index: Assessment of Functional Connectivity from multi Channel EEG and MEG with Diminished Bias from Common Sources*, Human Brain Mapping, 2007, **28**, *11*, 1178-1193.

Trambaiolli L.R., Falk T.H., *Hybrid Brain–Computer Interfaces for Wheelchair Control: A Review of Existing Solutions, their Advantages and Open Challenges*, Smart Wheelchairs and Brain-Computer Interfaces, Academic Press, 2018, 229-256.

Van Dokkum L.E.H., Ward T., Laffont I., *Brain Computer Interfaces for Neurorehabilitation–its Current Status as a Rehabilitation Strategy Post-Stroke*, Annals of Physical and Rehabilitation Medicine, 2015, **58**, *1*, 3-8.

Yuan H., He B., *Brain–Computer Interfaces Using Sensorimotor Rhythms: Current State and Future Perspectives*, IEEE Transactions on Biomedical Engineering, 2014, **61**, *5*, 1425-1435.

www.gtec.at (accessed on 1.08.2020).

COMBINAREA DE TRĂSĂTURI EXTRASE DIN SEMNALUL ELECTROENCEFALOGRAFIC – STUDIU DE CAZ

(Rezumat)

Prezentul studiu dovedeşte eficienţa combinării trăsăturilor relevante extrase din semnalele electroencefalografice într-o paradigmă creier-calculator ce implică imaginarea motorie. Metodele utilizate pentru extragerea trăsăturilor şi pentru formarea vectorului hibrid de trăsături sunt modelul autoregresiv, modulaţia în amplitudine şi sincronizarea de fază. Pentru clasificarea trăsăturilor s-a utilizat analiza discriminantă

liniară, analiza discriminantă pătratică, clasificatorul bazat pe calcularea distanţei Mahalanobis, algoritmul celui mai apropiat k vecin şi a clasificatorului cu support vectorial. Rezultatele obţinute indică faptul că indicele de întârziere al fazei şi indicele de blocare al fazei, folosiţi pentru sincronizarea de fază împreună cu clasificatorul cu suport vectorial, îmbunătăţesc performanţa ratelor de clasificare şi surclasează celelalte combinaţii testate.