Postprint of: Kurowski A., Mrozik K., Kostek B., Czyżewski A., Comparison of the effectiveness of automatic EEG signal class separation algorithms, JOURNAL OF INTELLIGENT & FUZZY SYSTEMS, Vol. 37, Iss. 6, (2019), pp. 7537-7543, DOI: 10.3233/JIFS-179360

Comparison of the effectiveness of and automatic EEG signal class separation

Adam Kurowski[a] , Katarzyna Mrozik[a], Bozena Kostek[b,*] and Andrzej Czyzewski[a]
[a]*Gdansk University of Technology, Faculty of Electronics, Telecommunications and Informatics, Multimedia Systems Department, G. Narutowicza 11/12, 80-233 Gdansk, Poland*
[b]*Gdansk University of Technology, Faculty of Electronics, Telecommunications and Informatics, Audio Acoustics Laboratory, G. Narutowicza 11/12, 80-233 Gdansk, Poland*

Abstract. In this paper an automatic brain activity class separation of EEG (electroencephalographic) signals is presented. The methodology consists of several steps, namely: signal acquisition, signal processing utilizing Independent Component Analysis (ICA), parametrization and data clustering. EEG signal is acquired from a headset containing 14 electrodes. ICA algorithm is performed to reflect brain activity generated by primary sources. For the parametrization two methods are used, i.e. Discrete Wavelet Transform (DWT) and autoencoder, i.e. the second order feature extractor. Then, the effectiveness of separation of the data into appropriate classes processed is observed and compared. Results show similar effectiveness, however they differ in the standard deviation values. Finally, supervised classification of signals is performed as a form of benchmarking.

Keywords: EEG Signal, Independent Component Analysis (ICA), Discrete Wavelet Transform (DWT), Autoencoder, Class Separation.

1. Introduction

The objective of this work was to assess whether it is possible to create a feature extraction algorithm which would be useful for data clustering and for unsupervised classification of EEG (electroencephalographic) signals with the use of the autoencoder neural network. Another aim was to compare separation effectiveness while applying DWT- and autoencoder-based feature extraction along with clustering methods.

Recording and monitoring EEG signals constituted the basis of research on brain activity for several decades as they form a valuable source of information for the purpose of biofeedback [20],[22]. Advances in EEG-headset hardware along with the development of signal analysis algorithms, clustering methods and machine learning allow for observation a person's mental state or anomalies that may be associated with certain cognitive problems. To that end, EEG signal *K*-means and fuzzy c-means (FCM) clustering are used to separate the input data set to be classified by various machine learning algorithms, such as for example: ANFIS (Adaptive Neuro-Fuzzy Inference System) [4], RBFNN (Radial Basis Function Neural Network) [16], RNN (Recurrent Neural Network) [3], SVM (Support Vector Machine) [19]. Recent Advances in Machine Learning and Soft Computing brought another approach, i.e. Kim and Yang [8] discussed an improvement in a classification effectiveness by utilizing AdaBoost instead of linear discriminant analysis (LDA). In order to confirm the classification improvement, the classification accuracy of both mentioned algorithms were analyzed [8]. To assess the abnormality in the brain activity, typically EEG signals are first decomposed into frequency subbands using discrete wavelet transform (DWT). Orhan *et al.* [14] followed this schema of analysis. They used the *K*-means clustering algorithm for each frequency sub-band wavelet coefficients. The probability distributions were computed according to distribution of wavelet coefficients to the clusters, and then used as inputs to the MLPNN (Multilayer Perceptron Neural Network) model [14]. Recently, an interesting study

*Corresponding author. E-mail: bokostek@audioakustyka.org.

was reported on the analysis of EEG microstate sequence properties, while employing five clustering algorithms. This study was performed in the context of neurobiological relevance of EEG microstates for physiological and pathological conditions [24]. Even though new outcomes in this research area are seen [6],[13],[24], automatization of such processes is still far from being fully achieved, thus more research is needed.

The study performed by the authors consisted of several stages. First, electroencephalographic (EEG) signals from six subjects performing three tasks were acquired through a headset containing 14 electrodes. Two types of parametrization were utilized, namely: wavelet- and autoencoder-based algorithms. Then, unsupervised clustering employing *K*-means, spectral and Ward hierarchical algorithm [5],[11],[12] was executed and the results of the autoencoder-based clustering are compared with the outcomes achieved with the DWT-based feature vectors. The EEG signal was pre-processed with the use of Independent Component Analysis (ICA) as a blind source separation algorithm in order to extract estimates of primary sources of EEG signals. Also, visualization of clustering performed by the autoencoder and *K*-means algorithm with the marked time periods associated with each performed task is shown. A comparison of results is shown in a form of statistical analysis as well as supervised neural network-based classification. Possible future directions were also depicted, e.g. taking into account influence of personal interests of subject in activities (e.g. specified kind of music) to be performed, fatigue and mood of the person participated in the test, or level of engagement in performed tasks on the assignment of data to different clusters.

The paper is a revised and extended version of the MISSI'2018 authors' conference paper [10].

2. Methodology proposed

2.1. Experimental Setup – Signal Acquisition

The experimental setup consists of several steps, which are shown schematically in Figure 1. For EEG signal acquiring a headset was used which contains 14 electrodes. It collects raw signals from a set of standard positions: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8 and AF4, according to the 10-20 electrode configuration on the scalp. Several scripts were written in the Python environment extended with libraries provided by the headset manufacturer to process the acquired raw signals. EEG raw data was sampled at 128 S/s. Data acquired from the headset were stored in the csv data format.

Six healthy subjects of age between 20 and 30 took part in tests consisted of three tasks. Each task lasted for approximately 5 minutes, the whole recording session took 15 minutes. Informed consent was obtained from all subjects. The subjects did not practice the tasks earlier. The conditions of recordings of the EEG signals were as follows:

– 1st task: resting state condition with closed eyes - during this task execution the subjects were asked to relax,
– 2nd task: playing back a music video with folk metal music genre and following it,
– 3rd task: playing a Netwalk logic game, in which a player has to rotate elements of the board in such manner, that there would be a connection between a single central element called "server" and multiple peripheral elements called "computers", thus observing the whole computer screen.

2.2. Parametrization

Two ways of raw signal parametrization were utilized, namely Discrete Wavelet Transform (DWT) and an autoencoder neural network, an unsupervised machine learning algorithm, a type of deep learning algorithms. However, the first step was to employ the ICA algorithm in order to perform blind source separation of signals gathered from the electrodes. The resulting parameters were as follows:

– wavelet transform-based parameters represent mean values and variances of all possible levels of discrete wavelet transform,
– the ICA algorithm provided the training data for the encoding part of the autoencoder neural network, its output created a set of features.

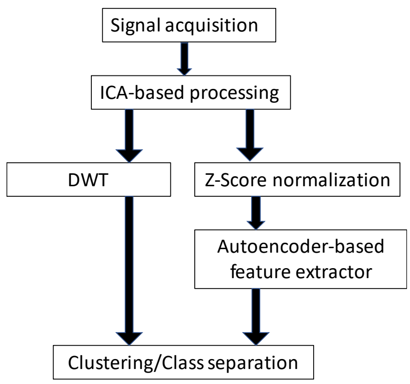

Fig. 1. The overview of the methodology proposed.

Signals were cut into a set of 15 seconds-long segments (frames), with the sampling rate of 128 S/s, entailing the length of 1920 samples per single channel of a single segment. Then, the ICA algorithm, which estimates primary sources of signals located on the surface of the brain was applied. The implementation of the algorithm requires the same number of estimated source signals as the number of the input signals. Therefore, each segment contains data associated with all 14 channels of the ICA-processed EEG data. These estimates were then used for calculating parameters for the data separation and clustering stages.

Data processing and separation/clustering schema is shown in Figure 2. For the visualization purpose the decision space was presented as if it has two dimensions associated with two features. However, it should be reminded that DTW-based feature vector consisted of 1513 elements and the autoencoder-based feature extraction resulted in 1120 parameters.

As already mentioned, all algorithms, including processing of signals gathered from the EEG acquisition headset, were implemented in the Python environment. Several libraries were employed: NumPy, SciPy and PyWavelets, Scikit-learn, Theano and Keras were used as machine and deep learning libraries [7],[9],[15],[21]. In the case of parameters extracted by autoencoder networks – an algorithm is capable of creating own features related with each channel of the input signal. Therefore the classification algorithm takes into consideration both spectral and spatial characteristics of the EEG signal acquired.

As seen in Figure 2, feature vectors resulted from the wavelet- and autoencoder-based processing are analyzed by three selected clustering algorithms, i.e.: *K*-means, spectral and Ward hierarchical algorithm, which are used for the unsupervised assignment of segment-related vectors of parameters to one of *k* clusters. The algorithms investigated perform an assignment of data into clusters based on different principles:

- *K*-means algorithm is a stochastic clustering method that minimizes the sum of the squared distances between *k* centroids of clusters and feature-related points in their neighborhood [5],
- spectral clustering is based on the extraction of eigenvalues of data before dividing them into clusters with another secondary algorithm (i.e. *K*-means) [12],
- Ward hierarchical clustering performs an assignment to classes by iterative merging of the most similar clusters of data. The similarity measure is a sum of squared distances between points of clusters, however the result of calculation is used for making decision to which clusters they should be merged. The algorithm execution is finished when there are only *k* clusters left [11],[25].

It was assumed that due to the differences in working principles of the algorithms investigated some of them would be more advantageous than others, that’s why three separation methods were used. The number of *k* clusters was set to 5 in relation to the number of the tasks performed by each subject during tests and associated with expected three states, i.e. resting, listening to music/watching a music video and playing a game, and two additional states, which may be associated with some unexpected events like distraction of subjects or noise present in the input data.

For the wavelet-based processing, coefficients of discrete wavelet transform were computed for the maximum possible level of decomposition. For each level of decomposition, mean and variance were calculated in order to decrease the number of generated features. Several wavelets were used, namely: Coiflets 1 and 2, Daubechies 1, 2 and 9 and Symlet 9. The basis mother wavelet and its order were chosen based on the critical literature review [2],[17].

In the case of the autoencoder neural network applied to feature extraction [3],[18], two parts: encoder function $h = f(x)$ and decoder function $r = g(h)$, where: x denotes an input, h refers to the hidden layer are employed. The aim of the decoder function is to produce a reconstruction of the original function $r = g(f(x)) = x$ with the least possible amount of deformation, but at the same time with compressed representation. The network may also be utilized for the extraction of n features associated with vector of N samples. In the autoencoder-based feature extraction, signals obtained from ICA were normalized using Z-score normalization, given by the following formula:

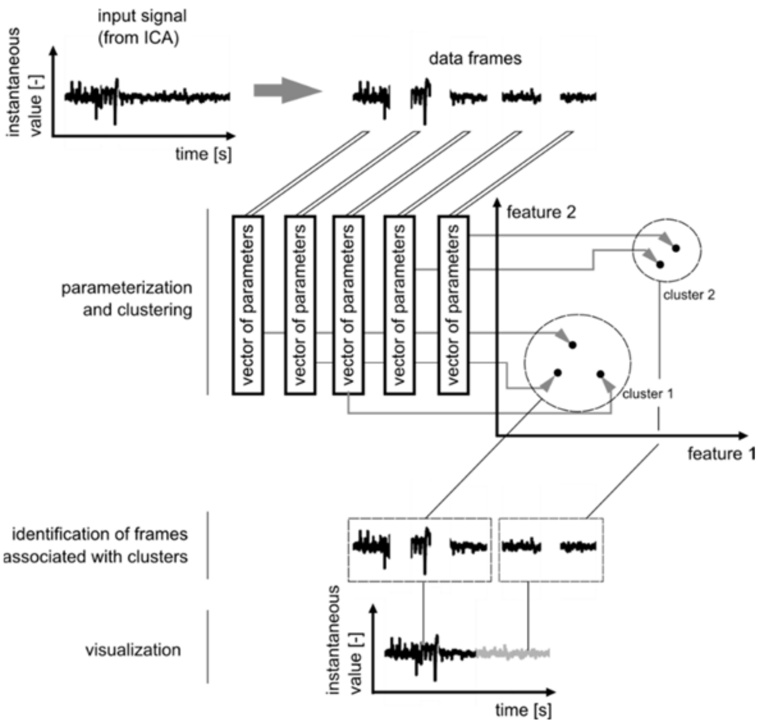

Fig. 2. Flowchart of the EEG-based data processing.

$$z = \frac{x - \acute{x}}{\delta_x}, \quad (1)$$

where z is normalized vector of features, x is unnormalized vector of features, $\acute{x}$ refers to the mean value of vector x, and δ_x is standard deviation of such vector.

Each single output signal from the ICA has its own autoencoder trained, therefore it was necessary to train 14 autoencoder networks in the process of feature extraction, as shown in Figure 3. Each vector is associated with one time segment of the EEG signal. This enables to connect a certain moment of time with a given feature vector. Therefore, after the clustering stage, each time segment was associated with one class assigned by the clustering algorithm. As seen in Figure 3, three types of clustering algorithms, mentioned earlier, were used in the final stage of data processing in order to compare their performance and ability to detect different types of data clusters present in the analyzed sets of feature vectors. The resulting features are a type of nonlinear dimensionality reduction performed on data returned by the ICA-based processing [23].

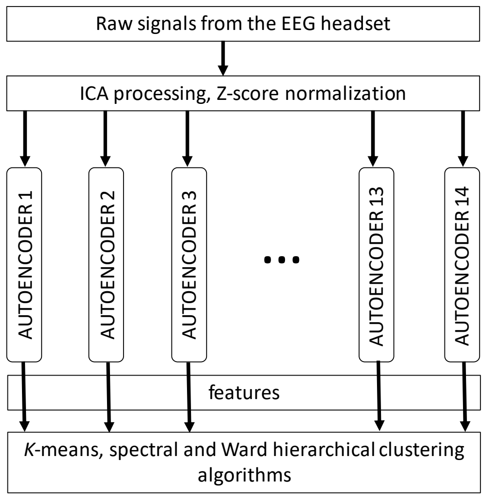

Fig. 3. Principles of the autoencoder-based feature extractor working and clustering performance.

The encoder consisted of six layers of neurons. As resulted from the pre-tests, i.e. the best performing set of hyperparameters, the following sizes of neuron layers: 640, 420, 320, 240, 160 and 80 were used. The decoder was designed analogously, with the reversed order of neuron layers. The hyperparameters were selected on the basis of the number of trials and selection of best results of clustering performed on their basis. The outcomes allowing for a better distinction between time periods associated with performing particular exercises were considered as the better ones. Resulting from this analysis, for each frame, 80 features were calculated by a single autoencoder instance. Next, the encoder parts of the trained networks were used for the computation of a feature vector associated with each segment of the analyzed EEG signal, which in turn can be used for data clustering. The resulting feature vector consisted of 14 concatenated *l* vectors calculated for each output signal generated by ICA.

3. Results and discussion

The results can be presented in several forms, e.g. a graphic representation of signals obtained by the ICA algorithm output or a tabular way of results obtained. However, visual analysis is not always suitable because of the complexity of the acquired results. Often, the result of classification obtained for a time segment associated with a task performed consists of a sequence of several mental states switched on and off. Therefore, each task performed by a subject is connected to the set of clusters, which are identified in corresponding time segments of an EEG signal captured during a test session. This kind of visualization is depicted in Figure 4. Time segments associated with each of three tasks performed by subjects are marked by rectangles. A shade of gray of waveforms corresponds to a cluster, to which it was assigned. Visually - each task is associated with the most frequently occurring cluster.

As already mentioned, though, a graphic representation of signals obtained by the ICA algorithm output makes it possible to evaluate roughly the performance of the clustering algorithm, there is a need for more objective evaluation of results. Therefore, an estimate of the probability density of states associated with each exercise was proposed. It allows for quantitative analysis of occurrence frequency of clusters as a result of the EEG data unsupervised classification. This estimate was then treated as a vector in a *k*-dimensional space. As the number of classes provided by clustering algorithms was set to 5, the *k* is also equal to 5. Therefore, each task performed by the subject was associated with one vector representing the estimate of probability of each cluster-related state. Then, distances between each pair of such vectors were calculated. The Euclidian distance was used in the process of assessing the mean spatial separation of clusters:

$$d = \sqrt[2]{\sum_{i=1}^{k} d_i^{\,2}}, \quad (2)$$

where *d* denotes distance between clusters and d_i is the *i*th distance from *k* distances, which are possible to be calculated.

The mean value was calculated for each of three distances obtained for each subject (see Table 1). The Ward clustering algorithm did not provide separation between classes, therefore, results for only *K*-means and spectral algorithms are presented in Table 1. Values were rounded up to two decimal places.

Results obtained from the unsupervised classification of the gathered EEG signals suggest that it is not only possible to associate each activity performed by the person involved in the experiment with a probability vector, but it should be presented that way. The probability vector is a vector in the *n*-dimensional decision space and can be treated as a kind of *fingerprint* of a state of an EEG signal gathered from the subject. Results from Table 1 were further processed to obtain four more metrics shown in Table 2, where $\overline{d}$ denotes the mean Euclidean distance between vectors associated with each exercise, δ_d refers to the standard deviation of the mean distance, d_{max} and d_{min} denote maximum and minimum distance, consecutively.

The mean distance and its variance between probability vectors vary depending on parameters used for the feature extraction process. In the case of the feature vectors extracted by an autoencoder network there is also dependence on the number of epochs of the training algorithms. The least value was associated with the largest separation distance provided by the *K*-means algorithm. In some cases - autoencoders allowed for obtaining a greater value of the mean distance and smaller standard deviation than ones obtained for the wavelet-based algorithm. It is also worth mentioning that in some cases *K*-means algorithm was not able to separate time segments associated with each task.

Finally, the parameters calculated for the four scenarios considered, i.e. obtained with the use of wavelets and autoencoders with an increasing number of training algorithm epochs, were used in supervised classification of tasks performed by subjects. Acquired signals were split into 339 frames assigned to three classes, that were connected to each of three tasks, 226 of them were randomly assigned to the training set, 113 were used for validation stage. Data were normalized according to Eq. (1).

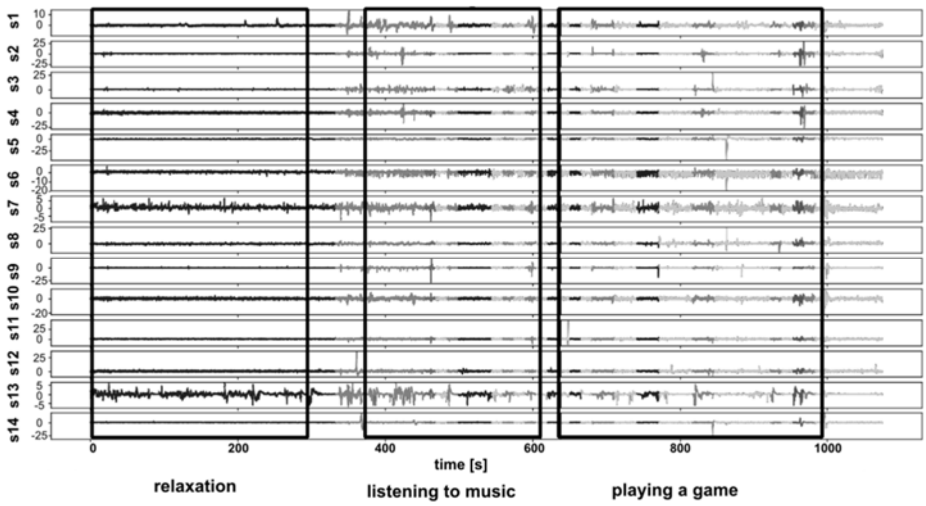

Fig. 4. Example of visualization of clustering performed by the autoencoder and *K*-means algorithm with the marked time periods associated with each performed task. Three time segments associated with each of performed tasks are marked below the *x* axis, on the *y* axis the instantaneous value of estimated primary source signals ($s1$ to $s14$) signals obtained from ICA is presented.

The neural network used for the purpose of classification consisted of 1120 input neurons, 100 neurons in the hidden layer, and 3 neurons in the output. Classes associated with exercises were encoded in the one-hot encoding manner. Therefore, *n* classes are associated with *n* vectors consisting of *n*-1 zeros and one value equals 1, which is placed in the unique position in the vector, allowing for identification of the associated class. A plain backpropagation algorithm was used for the training the neural network.

Table 1

Mean distance between vectors associated with clusters related to tasks performed by subjects

Subject	Parametrization/clustering method							
	wavelets		autoencoder-based (50 epochs)		autoencoder-based (100 epochs)		autoencoder-based (200 epochs)	
	K-means	spectral	*K*-means	spectral	*K*-means	spectral	*K*-means	spectral
1	0.05	0.22	0.04	0.17	0.2	0.15	0.09	0.08
2	0.4	0.53	0.89	0.2	0.41	0.47	0.53	0.38
3	1.06	0.77	0.63	0.46	0.76	0.56	0.06	0.56
4	0	0.14	0.73	0.19	0.7	0.29	0.17	0.29
5	0.02	0.12	0.1	0.14	0.18	0.28	0.11	0.08
6	0	0.3	0.66	0.39	0.12	0.16	0.13	0.18

Table 2

Metrics calculated from results contained in Table 1

	Parametrization/clustering method							
	wavelets		autoencoder-based (50 epochs)		autoencoder-based (100 epochs)		autoencoder-based (200 epochs)	
	K-means	spectral	*K*-means	spectral	*K*-means	spectral	*K*-means	spectral
$\overline{d}$	0.26	0.35	0.51	0.26	0.40	0.32	0.18	0.26
δ_d	0.42	0.26	0.35	0.13	0.28	0.17	0.17	0.19
d_{max}	1.06	0.77	0.89	0.46	0.76	0.56	0.53	0.56
d_{min}	0	0.12	0.04	0.14	0.12	0.15	0.06	0.08

Table 3

Results of EEG signal classification with the use of artificial neural network expressed by classification error rate

	wavelet-based	autoencoder 50 epochs	autoencoder 100 epochs	autoencoder 200 epochs
mean value	0.495	0.536	0.526	0.573
standard deviation	0.041	0.038	0.039	0.044

Vectors of features were randomly assigned to the training and validation sets. Also, the process of training and assessing the performance of the network was repeated 100 times and then repeated in order to evaluate mean efficiency of the classifier based on each of the feature sets investigated. Results of classification are shown in Table 3 in the form of mean efficiency and standard deviation of the classifier effectiveness.

No definitive answer as to the most suitable wavelet family is available at this stage. A short training with a maximum limit of epochs set to 75 and the target loss value set to 0.001 was first performed. In fact all wavelet-based training achieved the target loss in less than 75 epochs, however the best results were obtained for Symlet 9 mother wavelet. However, the number of epochs required to meet one of the stop conditions was larger and the target loss was expected to be smaller than 1e-7. Overall, the lowest mean value of the mean efficiency was obtained for the wavelet-based classification feature set. In the case of random classification, the efficiency of classification tended to be 1/3. Therefore, in each case the classifier performed better than random classification. All algorithms based on a set of autoencoders performed better than the one based on the wavelet-extracted parameters. Values of the standard deviations are similar in each case. The best performance was obtained for the system trained for 200 epochs. It should be noted that clustering algorithms performed worse for the last mentioned scenario, therefore additional experiments with the use of other clustering algorithms applied to a larger group of subjects should be done in order to further investigate the dependence between the distance of clusters obtained in the process of unsupervised clustering algorithms and the supervised classification methods with the use of artificial neural networks.

4. Conclusions

It was shown in this study, that it is possible to perform the clustering of the EEG signal with the use of unsupervised clustering algorithms and vectors of features obtained with the use of autoencoder neural networks. The proposed architecture of an autoencoder-based feature extractor allowed for obtaining similar Euclidean distances as the ones provided by the DWT-based feature extractor, however it should be noted, that the number of epochs of the training algorithm has a significant influence on the performance of this kind of the feature extraction algorithm.

In both cases the EEG signal was pre-processed with ICA algorithm in order to extract estimates of primary sources of EEG signals. Also, normalization is an additional prerequisite, if a feature extraction algorithm based on the autoencoder neural network is considered. That is why such a type of process was also used in the case of supervised classification. However, the satisfactory performance was achieved for only two algorithms: *K*-means and spectral. Achieved distances were significantly smaller for the Ward clustering algorithm and they were close to zero, providing almost no separation between classes. In addition, in some situations, a series of consecutive signal frames were not consistently associated with one cluster. Therefore, for each set of EEG signal frames associated with tasks performed by subjects, a vector of probabilities of classes assigned to frames by each algorithm was calculated. Such vectors, treated as vectors in the *n*-dimensional space can be used for further analyses and for the analysis of the state of the EEG signal gathered from the subject. The results obtained may imply that they are algorithm-dependent. This should be further investigated.

More research should also be performed in order to determine how an assignment to different clusters can be connected to such factors as personal interests of the subject in activities (e.g. specified kind of music) to be performed, fatigue and mood of the person participated in the test, or level of engagement in performed tasks. Such a goal may be achieved by extending the range of performed activities, through the increasing the number of participants and by the use of some additional signal analysis techniques. An example of such a technique may be the use of convolutional neural networks and data augmentation methods to obtain features better suited for the type of the processed data.

Potentially, monitoring such features of the EEG signal may allow for a distinction between brain activities, in which subjects were involved. Information about changes of the mental state of the subject may be utilized for investigating the mood of patients with brain injuries for whom BCI (Brain Computer Interface) often may be the only modality allowing for communication with other people. Each task performed by subjects could be assigned to a vectors of probabilities of occurrence of cluster and such a vector may be used in further process of classification. However, performance of such a method may depend on the type of the clustering algorithm used. Also and interesting future research could be associated with investigation of precedence of cluster-related states identified by clustering algorithms. Analysis tools like Hidden Markov chains could be used as a starting point for such a study.

Acknowledgments

The project was funded by the National Science Centre on the basis of the decision number DEC-2014/15/B/ST7/04724.

References

[1] A. Alhagry, A.A. Fahmy, R. A. El-Khoribi, EEG-Based Emotion Recognition using 3D Convolutional Neural Networks, (IJACSA) *International Journal of Advanced Computer Science and Applications*, Vol. 8, No. 10, (2017), 329-337.

[2] N.K. Al-Qazzaz, S. Hamid Bin Mohd Ali, S.A. Ahmad, M.S. Islam, J. Escudero, Selection of mother wavelet functions for multi-channel EEG signal analysis during a working memory task, *Sensors*, 15, (2015), 29015-29035.

[3] S. Gluge, R. Bock, A. Wendemuth, *Auto-encoder pre-training of segmented-memory recurrent neural networks.* in: ESANN 2013, Bruges, Belgium 2013.

[4] I. Güler, E.D. Ubeyli, Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients, *Journal of Neuroscience Methods* 148 (2005) 113–121, doi:10.1016/j.jneumeth.2005.04.013.

[5] P. Harrington, *Machine Learning in Action*, Manning Publications Co., Greenwich, Ct, USA 2012.

[6] M.Z. Ilyas, P. Saad, M.I. Ahmad, A.R.I. Ghani, *Classification of EEG signals for brain-computer interface applications: Performance comparison*, 2016 International Conference on Robotics, Automation and Sciences (ICORAS), 5-6 Nov. (2016). doi: 10.1109/ICORAS.2016.7872610.

[7] E. Jones, T. Oliphant, O. Peterson, et al., SciPy: Open source scientific tools for Python, Internet website of the Python library, http://www.scipy.org/ (accessed November 2018).

[8] J-H. Kim, Y.-M. Yang, An enhanced classification scheme with AdaBoost concept in BCI, *Journal of Intelligent & Fuzzy Systems*, vol. 35, 1 (2018), 63-68, doi: 10.3233/JIFS-169567.

[9] Keras, Internet website of the Python machine learning library, https://keras.io/ (accessed November 2018).

[10] A. Kurowski, K. Mrozik, B. Kostek, A. Czyzewski, Automatic Clustering of EEG-Based Data Associated with Brain Activity; in: 11th edition of International Conference on Multimedia & Network Information Systems MISSI (2018), 470 - 479, Wroclaw, Poland, 12.9.2018 - 14.9.2018, doi: 10.1007/978-3-319-98678-4_47.

[11] F. Murtagh, P. Legendre, Ward's Hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *J. Classif.*, 31 (2014), 274-295, doi:10.1007/s00357-014-9161-z

[12] A. Ng, M. Jordan, Y. Weiss, *On spectral clustering: analysis and an algorithm. Advances in Neural Information Processing Systems*, 849-856, MIT Press 2001.

[13] L.F Nicolas-Alonso, J. Gomez-Gil, Brain Computer Interfaces, a Review, *Sensors* (Basel), 12(2), Published online Jan. 31 (2012), 1211–1279, doi: 10.3390/s120201211.

[14] U. Orhan, M. Hekim, M. Ozer, *Expert Systems with Applications*, 38 (2011) 13475–13481, Elsevier 2011, doi:10.1016/j.eswa.2011.04.149.

[15] PyWavelets, Website of the Python library, http://pywavelets.readthedocs.io (accessed November 2018).

[16] S.K. Satapathy, S. Dehuri, A.K. Jagadev, EEG Signal Classification using PSO Trained RBF Neural Network for Epilepsy Identification, *Informatics in Medicine Unlocked*, 6, (2016), doi: 10.1016/j.imu.2016.12.001.

[17] S. Saeid, J.A. Chambers, *EEG Signal Processing*, John Wiley & Sons, Chicester, USA 2007.

[18] A. Saroff, *Musical audio synthesis using autoencoding neural networks*, in: Proceedings of the International Computer Music Conference, Athens, Greece (2014).

[19] K. Sercan Bayram, M.A. Kizrak, B. Bolat, *Classification of EEG Signals by using Support Vector Machines*, 2013/06/19, 2013 IEEE International Symposium on Innovations in Intelligent Systems and Applications, IEEE INISTA (2013), doi: 10.1109/INISTA.2013.6577636.

[20] J.C-Y. Sun, K.P-C. Yeh, The effects of attention monitoring with EEG biofeedback on university students' attention and self-efficacy: The case of anti-phishing instructional materials, *Computers & Education*, 106 March (2017) 73-82, https://doi.org/10.1016/j.compedu.2016.12.003.

[21] Theano, Internet website of the Theano Python machine learning library, http://deeplearning.net/software/theano (accessed November 2018).

[22] D.J. Vernon, Can neurofeedback training enhance performance? An evaluation of the evidence with implications for future research, *Applied Psychophysiology and Biofeedback*, 30, 4 Dec. (2005), 347-364.

[23] W. Wang, Y. Huang, Y. Wang, L. Wang, *Generalized autoencoder: a neural network framework for dimensionality reduction*, in: IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus (2014), 496-503, doi: 10.1109/CVPRW.2014.79

[24] F. von Wegner, P. Knaut, H. Laufs, EEG Microstate Sequences From Different Clustering Algorithms Are Information-Theoretically Invariant, *Front. Comput. Neurosci*., 27 August 2018, https://doi.org/10.3389/fncom.2018.00070.

[25] M. Zaki, W. Meira, Data mining and analysis: fundamental concepts and algorithms. Cambridge University Press, New York, USA 2014.