D.S. RESHETNYKOV, graduate student,
International Research and Training Center
for Information Technologies and Systems of the NAS and MES of Ukraine,
Glushkov ave., 40, Kyiv, 03187, Ukraine,
denis.reshetnykov@gmail.com

EEG ANALYSIS OF PERSON FAMILIARITY WITH AUDIO-VIDEO DATA ASSESSING TASK

To solve the problem of assessing a person's familiarity with audio-video data, various methods of machine learning were compared. The feature space has been optimized for the best way to make such an assessment. The high efficiency of the genetic algorithm in the problem of optimizing the space of attributes is shown.

Keywords: machine learning, genetic algorithm, electroencephalogram, DEAP dataset, emotion recognition

Introduction

Assessment of the emotional state of students in the process of studying the lecture material is an important task in both scientific and practical terms [1]. However, such studies are rather complicated due to the fact that the indicators of the emotional state differ considerably, depending on whether the studied material is familiar to the student or not [2].

Studies [3, 4] show that the assessment of the degree of familiarity with the presented material allows much more accurate assessment of the emotional state. Dynamic evaluation of familiarity with the materials, also allows us to estimate the speed of the skills mastering, which require repetition of the material, as shown in [5] using the example of evaluating familiarity with integrated development environments.

Thus, the control of the educational process would be much more effective if it were possible to objectively evaluate, by the totality of signs, the familiarity of a particular student with the materials presented to him/her.

Evaluation of the level of familiarity with the presented audiovisual data is also needed in other areas, in particular, in the developing area of neuromarketing research [6—8].

Purpose

The purpose of the article is to conduct a comparative analysis and experimental study of the effectiveness of various machine learning methods, to build a model for determining familiarity with the presented audiovisual materials, based on the analysis of the user's EEG signal, and to determine the set of signs that best classify this signal.

Using the EEG signal in the analysis of cognitive activity

Due to technical progress in the methods of extraction and filtering of EEG, as well as the spread of EEG monitors with a reasonably simple connection to a PC, it is theoretically possible to use EEG monitoring to determine whether a person is familiar with the presented audio-video data.

It is known that the cognitive activity of the brain in many states correlates with EEG indices [9]. So in recent studies, precise correlates of EEG have

been found with orientation in an imaginary maze [10], solving verbal and spatial problems [11, 12].

To solve the general problem of analyzing cognitive activity, in addition, to pick up a signal, it is necessary to solve several types of problems.

The first type is directly related to the physical nature and methods of the EEG signal extraction. EEG sensors record the total activity of many independent sources (individual neurons) from the surface of the skull, which leads to the fact that the received signal is noisy and has recording artifacts. These artifacts can reveal themselves either as a result of external (pick-ups from electrical equipment, poorly fixed electrodes) and for internal reasons (for example, changes in potential due to eye movement or electrical muscle activity). Given the chaotic nature of the EEG signal, even a small noise can significantly change the analysis results. Fortunately, there are well-established methods of automatic filtering and error correction. These include noise suppressing filters of high and low frequencies, filters with a finite impulse response, the common average reference [13].

The second type of the tasks is associated with the difficulties of analyzing and interpreting obtained data. In the absence of objective features of cognitive activity in the original signal of an electroencephalogram, one or several methods of identifying characteristic features are traditionally used in EEG analysis:

- statistical analysis;
- spectral analysis;
- signal analysis in the time-frequency domain.

The statistical analysis considers various statistical characteristics of the signal, such as dispersion and an average signal value, it is the least complex (of the above) methods and is often used in addition to other methods [14].

Spectral analysis basically contains a signal conversion from the amplitude-time to a frequency domain. Within this conversion, signal power is calculated in various frequency ranges (commonly known as alpha, beta, delta, and others). This type of analysis has been widely used in various medical and non-medical tasks of EEG analysis, and the characteristics obtained through its use correlate with such cognitive processes as attention [15], a solution of spatial and verbal problems [16], reading and mental calculation [9].

The third type of task is the selection of significant features. It is known that different areas of the brain are responsible for different types of cognitive activity [17]. Although currently there are only few unambiguous indicators of the participation of the certain brain areas in a specific cognitive task, it is possible to note that not all areas of the brain, and not all characteristics of the EEG signal will contribute equally. This fact creates the need for feature selection to improve classification accuracy. In addition, reducing the dimension of the feature space significantly reduces the computational complexity of many mathematical analysis algorithms, which is especially useful in real-time execution applications.

The last, fourth task is the selection and optimization of a mathematical classification model. This class of tasks is a subclass of the currently popular class of supervised machine learning problems, and has many solutions, such as logistic regression, a naive Bayes classifier, decision trees (as well as ensemble models based on this method), the k-nearest neighbors method, linear support vector machine (and its nuclear version) and various neural network models.

It is important to note that the third (the choice of significant features) and the fourth (selection and optimization of the model) tasks are strongly related (the choice of significant features in practice may largely depend on the model), and for this reason are performed in an arbitrary order.

Analysis of existing approaches to solving the classification problem

Before proceeding to the solution of the problem, we will briefly describe the existing approaches to machine classification, which will be used in experimental studies.

Logistic regression. This model considers the probability of occurrence of an event as a dependent variable

$$y = \frac{1}{1 + \exp^{-(\theta_0 + \theta_1 x_1 + \ldots + \theta_n x_n)}}, \qquad (1)$$

where $x_1, \ldots, x_n$ — is a set of independent variables, and $\theta_0, \ldots, \theta_n$ — regression coefficients.

Naive Bayes classifier. This model is based on the maximum aposterior probability and treats each input parameter as an independent variable. Thus, the probability of occurrence of the desired event

is defined as

$$P(c \mid X) = \frac{P(X \mid c)P(c)}{P(X)} \quad , \qquad (2)$$

where P(c|X) is the a posteriori probability of class c with a given set of independent variables X, P(c) — is the a priori probability of a given class, P(X|c) — is the probability of a given characteristic values for a given class, P(X) — is the probability of a given characteristic value.

Decision trees. The method of the decision trees is based on the idea of recursive division of the set of elements of X into such subsets in order to maximize entropy. To do this, first calculate the entropy of the original set:

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i \qquad (3)$$

and then at each step of the algorithm, the information increment function is calculated for each of the attributes of the set:

$$Gain(X, A) = H(X) - \sum_{a \in values(A)} \frac{|X_a|}{|X|} H(X_a) \quad (4)$$

where values(A) are all received attribute values of A, X_a is a subset of the data set in which A = a, |X| — the number of elements in the set.

The attribute with the largest information increase is selected as separating for a given step, and the algorithm proceeds to the next step. The number of steps of the algorithm (called tree depth) is a configurable parameter.

Method k-neighbors. The method is based on the principle of similarity of the object with the majority of the nearest neighbors. For this, for each object from the test sample, the distance to all objects of the training sample is calculated, and the object class is determined as the most frequent among the k closest objects.

$$y = \arg\max \sum_{k=1}^{k} I(y_k == y) , \qquad (5)$$

The parameter k is a configurable parameter of this method.

Linear method of support vector machines (SVM). This method occupies an important place in the modern theory of the pattern recognition and successfully competes with the multilayer neural network pattern recognition systems [18].

Let there be a training set:: (x_1, y_1), ..., (x_m, y_m), $x_i \in R^n$, $y_i \in \{-1,1\}$. The idea of the method for the binary classification problem is to divide all the learning points by the classifier hyperplane so that the separation and the difference between the 2 classes are maximum. The classifier hyperplane is defined by the function:

$$\langle w, x \rangle + b = 0, \qquad (6)$$

where w is a normal vector to the separating hyperplane, b — is an auxiliary parameter. To do this, the optimization problem is solved:

$$\begin{cases} \arg\min_{w,b} \|w\|^2 \\ y_i(\langle w, x_i \rangle + b) \ge 1, i = 1, ..., m \,. \end{cases} \qquad (7)$$

The kernel method of support vector machines. This method is a generalization of the SVM method for linearly inseparable cases and is based on the mapping of the original features of X into a space of higher dimension φ: $R^n \to X$ where the set becomes linearly separable. Thus (6) takes the form

$$F(x) = sign(\langle w, x \rangle + b) \qquad (8)$$

The function $K : X \times X \to R$ is called the kernel, if it can be represented in the form of $K(x,x') = \langle \phi(x), \phi(x') \rangle_H$ with a mapping $\phi: X \to H$, where H — space with scalar product.

The task of building a model using the kernel method of SVM is reduced to choosing / building the most optimal kernel for this task. Building an adequate kernel is an art and, as a common practice, relies on a priori knowledge of the subject domain [19].

Multilayer perceptron. A multilayer perceptron is a neural network with forward propagation of an error and has the following features:

▪ consists of neurons with a continuously differentiated activation function (usually logistic):

$$f(x) = \frac{1}{1 + \exp^{-2\theta x}}. \qquad (9)$$

▪ contains several hidden layers of computational neurons. These layers allow you to consistently extract important features from the input image.

▪ provides a fully connected network architecture [20].

In addition to the above methods, we will also investigate the following ensemble models:
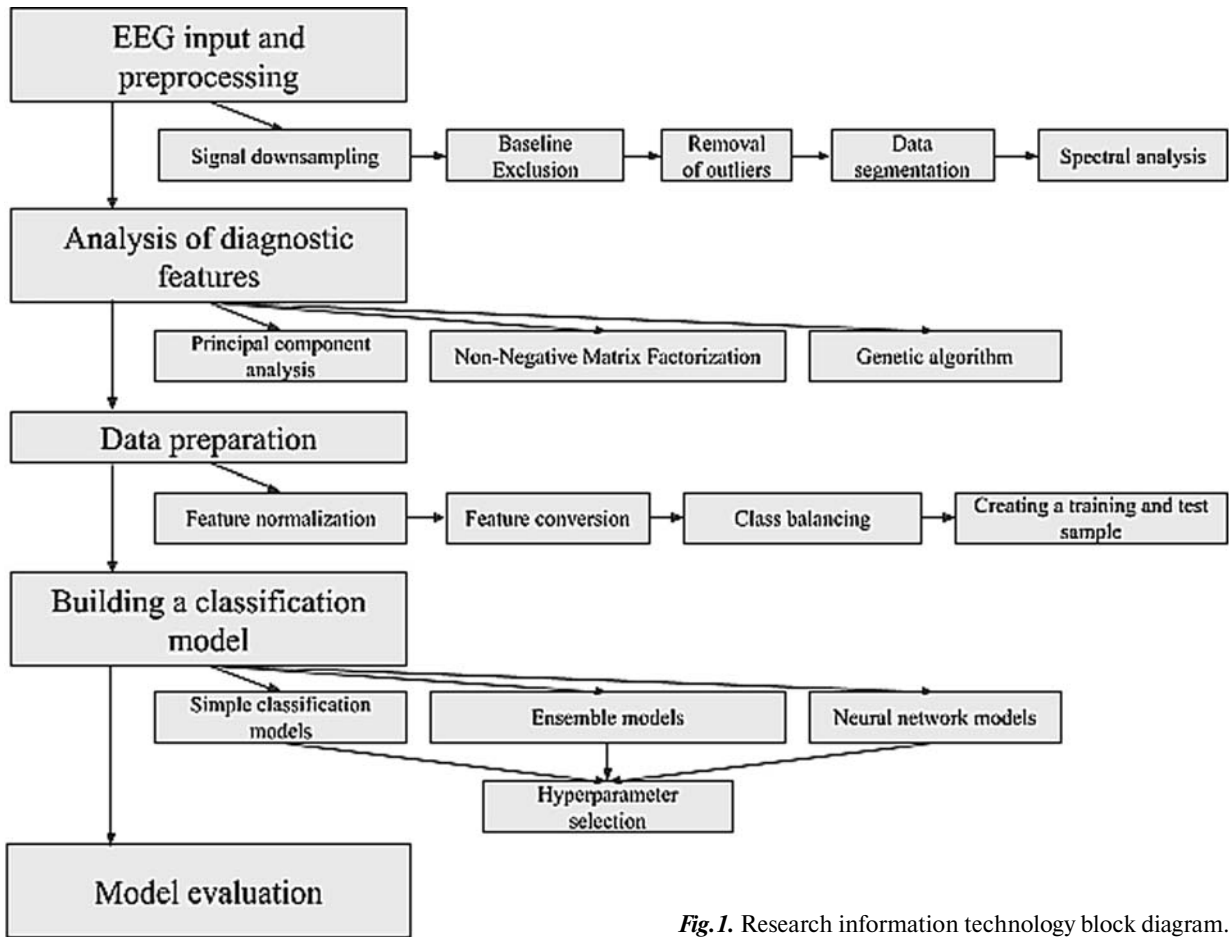
Fig. 1. Research information technology block diagram.

Random forest. This algorithm is to build an ensemble of decision trees. Each of the trees is constructed from a fixed sample with a return, from the original training set. The number of trees and the size of a fixed sample are parameters of this method in addition to the parameters of the decision trees themselves [21].

Gradient boosting decision trees. This method considers the decision tree algorithm:

$$b(x) = \sum_j \beta_j I \left[x \in X_j \right] \qquad (10)$$

in which the feature space is divided into a finite family of domains $\{X_j\}$, in each of which β_j is chosen independently, in order to minimize the error. The error function, for a fixed family $\{X_j\}$ takes the form:

$$\sum_{x_i \in X_j} L\left(y_i, a(x_i) + \beta_j \right) \to \min_{\beta_j} \qquad (11)$$

regulating parameter for the gradient boosting is the

speed of the gradient descent (also called the learning speed), and the configuration of the regions, given by the depth of the trees and the number of objects in trees [22].

Experimental design

The general block diagram of information technology for research is presented in Fig. 1. In the "Building a classification model" stage, simple classification models imply logistic regression models, naive Bayes, decision trees, k-neighbors, linear and kernel SVM.

A set of real EEG from DEAP database [23] was used for experiment. It represents multichannel recordings of physiological signals (PS) of 29 volunteers. For each volunteer, 40 tests were conducted, consisting of:
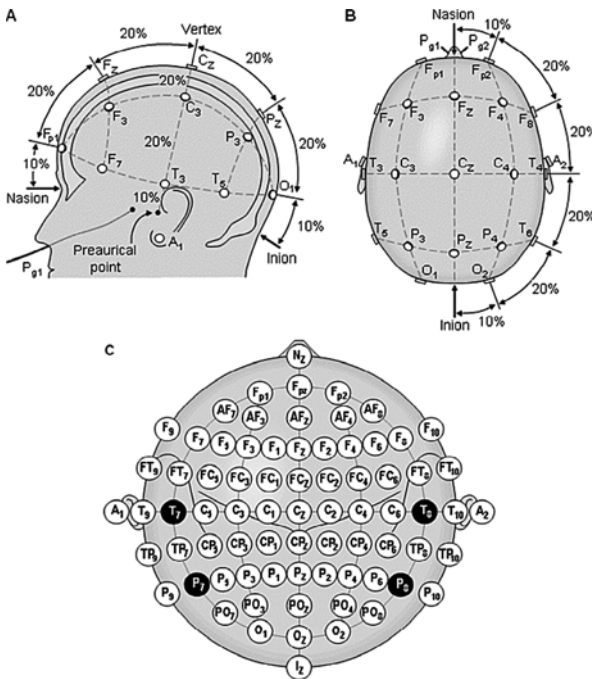
Fig.2. The standard arrangement of electrodes "10—20": *a* — side projection; *b* — top view; *c* — extended "10—20" scheme

1. Recording PS in the absence of a stimulus (baseline) with a duration of 3 s.

2. Recording PS in the presence of a stimulus (display of an audio-video fragment) with a duration of 60 s.

3. Evaluation of an audio-video fragment by several parameters, including the "Familiarity" parameter.

The EEG signal was recorded from 32 leads, with a sampling frequency of 512 Hz, according to the standard "10—20" (extended) electrode arrangement (Fig. 2).

Each subject signal records, subsampled to 128 Hz, was saved to individual data files, and translated into the numpy data format of the python programming language. Each data file contains 40 segments corresponding to 63s records (3s of the baseline and 60s of the stimulus). Therefore, overall dataset consists of $29 * 40 = 1160$ data segments.

Before the beginning of the analysis, each data segment was aligned with the mean value of the baseline, and the corresponding 3s records were excluded from further analysis. Since many ma-

chine learning methods are sensitive to data outliers, emissions of more than 2δ have been reduced to a mean value:

$$X_i = \begin{cases} X_i, X_i \in \left[\dfrac{1}{M}\sum_{m=1}^{M}X_m - 2\delta, \dfrac{1}{M}\sum_{m=1}^{M}X_m + 2\delta \right] \\ \dfrac{1}{M}\sum_{m=1}^{M}X_m, X_i \notin \left[\dfrac{1}{M}\sum_{m=1}^{M}X_m - 2\delta, \dfrac{1}{M}\sum_{m=1}^{M}X_m + 2\delta \right] \end{cases} \quad (12)$$

To construct the initial feature space, each of the 1160 data segments was divided into 15 parts with a duration of 4 s. Next, for each part, in each of the 32 leads, the signal power was calculated in 4 frequency bands: theta (4—8 Hz), alpha (8—12 Hz), beta (12—30 Hz), gamma (30—64 Hz) . Thus, for each of the stimulus, we received $32 * 15 * 4 = 1920$ characteristic features. Since most machine learning methods require normalized data, the obtained spectral power values were normalized.

The initial target parameter "Familiarity" is presented in the form of values on an ordinal 5-point scale. It was converted to a binary feature ("Familiarity" $>= 3$) and all data segments were divided into 2 derived classes. The ratio of classes in the original sample is 0: 725 (62%), 1: 435 (38%). For further use, the classes were balanced by dropping random segments of a larger class. Thus, in a further study, 870 data segments, equally divided between the two classes, were used.

For training and model testing, the resulting data set was randomly divided into training and test samples at a ratio of 70%/30%.

The model was measured using the accuracy measure:

$$\text{Accuracy} = \frac{P}{N} \quad (13)$$

where P is the number of correctly classified data points from test sample, N is the size of the test sample.

Since the majority of the above models have a number of hyperparameters, responsible for learning speed and regularization of the model, the values choice was carried out using the method of exhaustive search from the logarithmic scale

[0.001, 0.003, 0.01, 0.03, 0.1, 0.3, 1, 3, 10, 30, 100, 300, 1000, 3000, 10000].

To implement the machine learning algorithms, the Python programming language scikit-learn li-
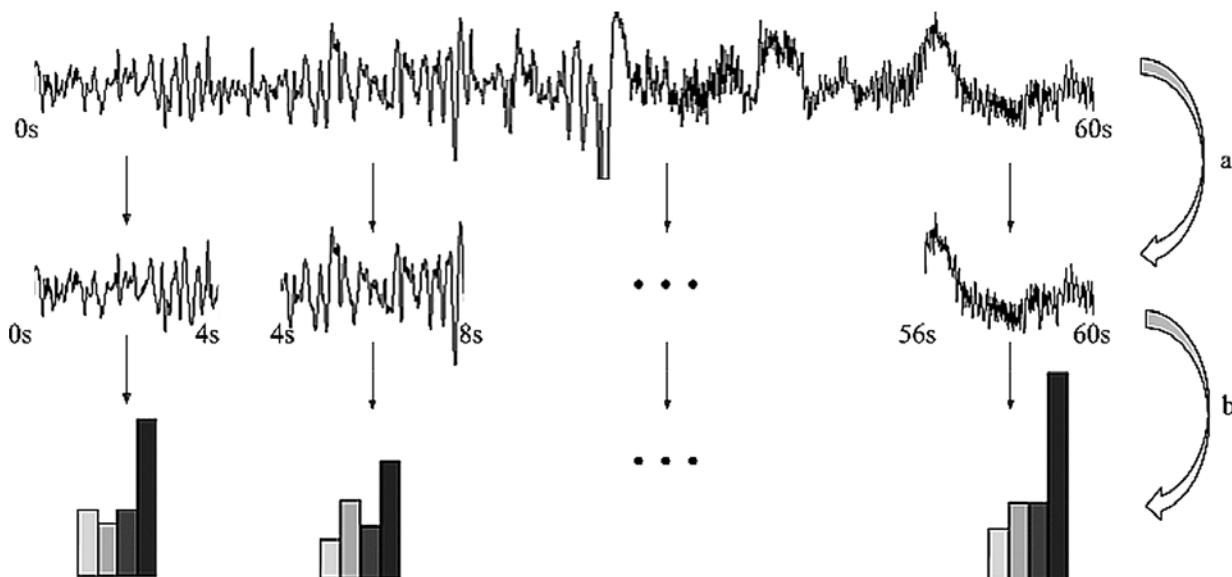
Fig. 3. Schematic representation of the data conversion to obtain the space of characteristic features: a — segmentation of the signal into 4s parts; b — obtaining the spectral characteristics of the signal

brary [24] was used. For evaluation of models, selection of hyperparameters and genetic algorithm the corresponding functions and modules in the Python programming language were implemented.

The selection of the significant features was carried out for a model built on the basis of the linear SVM, since this model has a smaller number of the regularization parameters, which made it possible to compare the different methods for selecting features.

Results of experimental studies

Using the above-described procedure for selecting model parameters, the following parameters were selected and the following model accuracy results were obtained (table 1):

Models in the table marked with an asterisk provide the accuracy of the above statistically significant. Ensemble and neural network models achieved the greatest accuracy.

Experiments have shown that further improvement in the quality of models can be achieved using the feature selection method. To compare such methods, a model of the linear method of support vector machines was chosen, as one of the models with the least influence of the hyperparameters

random selection, and a high dependence on the dimension of the input data.

The following algorithms were used for the selection of the significant features:

Principal component analysis. The essence of the principal component analysis (*PCA*) method is a significant reduction in the data dimension. The original matrix X is replaced by two new matrices T and P, the dimension of which is less than the number of variables (columns) of the original matrix X [25].

Let there be a matrix of variables X of dimension (M×N), where M is the number of samples (rows), and N is the number of independent variables (columns), which, generally, are many ($N >> 1$). The PCA is a method with minimal loss of useful information, since it uses new, formal variables t_a ($a = 1, ..., A$), which are a linear combination of the initial variables x_n ($n = 1,..., N$)

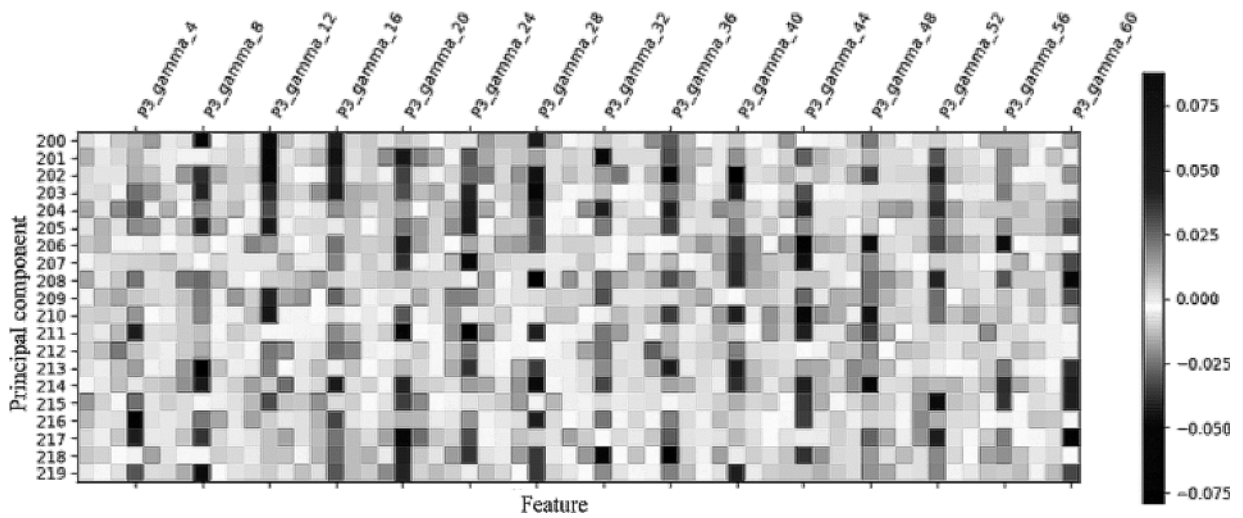$$t_a = p_{a1}x_1 + ... + p_{aN} x_N,\tag{14}$$

where p_{a1} — are the coefficients applied to the initial variables. With these new variables, the matrix X is decomposed into the product of two matrices T and P:

$$X = TP^t + E = \sum_{a=1}^{A} t_a p_a^t + E.\tag{15}$$
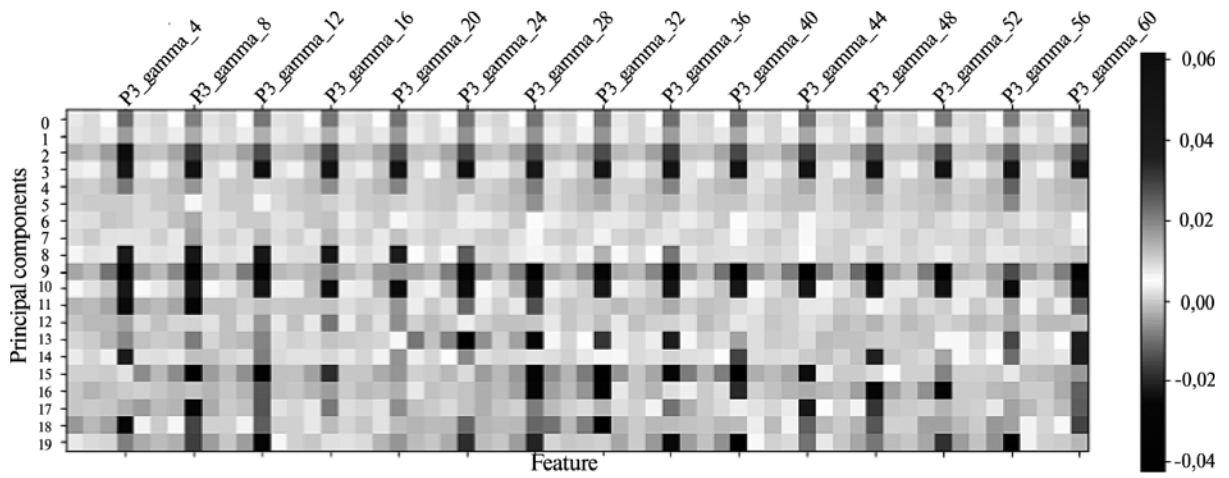
a

b

c

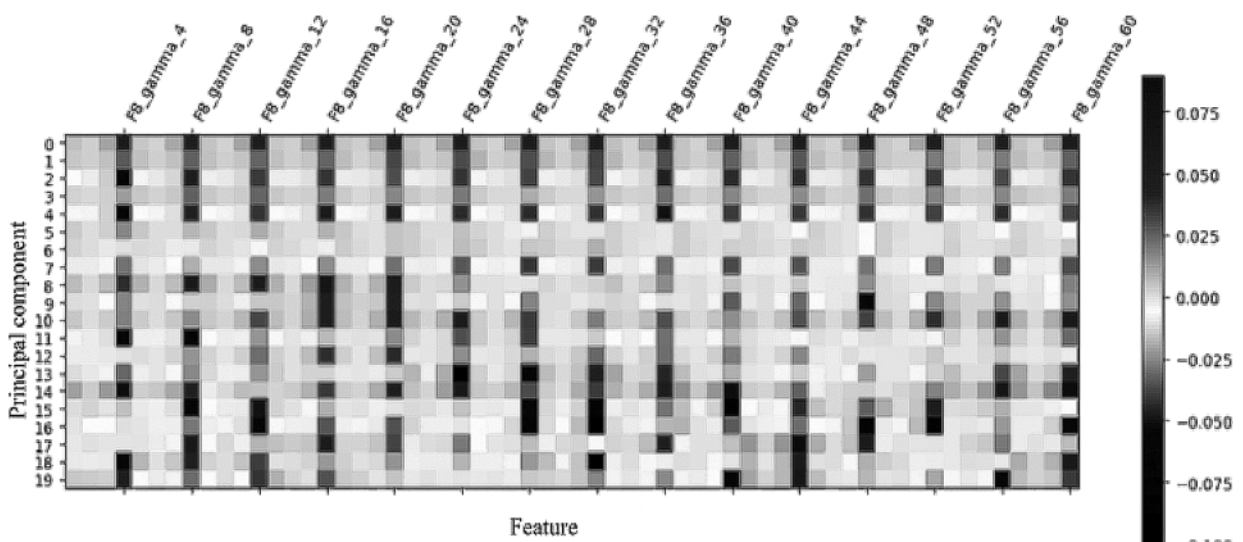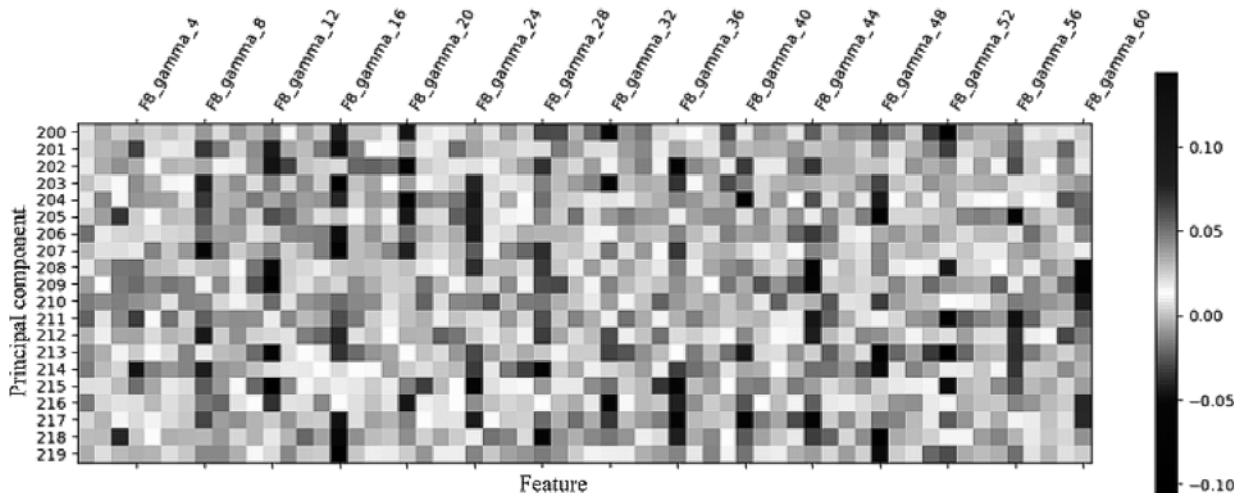**Fig. 4.** The coefficients of the linear combination applied to the original characteristics: *a* — derived from reference point P3 and applied to components № 0-19; *b* — from P3 to № 200-219; *c* — from F8 to № 0-19; *d* — from F8 to № 200-219; (darker areas correspond to the characteristics that have made a greater contribution to the corresponding principal component. The names of the characteristics are formed as reference point + spectral range + the end timepoint of the segment)

Fig. 4 shows partial examples of the matrix of coefficients applied to the initial characteristics when the number of significant features decrease to 447. The total size of this matrix is 447 * 1920 values.

Experiments have shown that the use of the *PCA* has made it possible to increase the accuracy of the *SVM* model from 55.9 to 58.2%, while reducing the number of significant features to 447.

Non-negative matrix factorization. This method is based on the representation of a matrix $Y \in R^{m \times n}$ as a product of matrices $W \in R^{m \times k}$ and $H \in R^{k \times n}$, in which all elements of the three matrices are non-

Table 1. **Accuracy and appropriate hyperparameters for different models without feature selection**

Model	Hyperparameters	Accuracy
Logistic regression *	C = 300	54,4%
Linear support vector machines*	C = 1000	55,9%
Kernel support vector machines*	C = 10000, γ = 0.001	57,1%
K-neighbors Classification	number of neighbors = 151	49,4%
Naive Bayes method	—	47,1%
Decision trees method	tree depth= 5	49,4%
Random forest ensemble model*	the maximum depths of trees = 2, number of features = 9	56,7%
Random forest ensemble model*	the maximum depths of trees = 2, number of features = 9	56,7%
Gradient boosting decision trees ensemble model*	the maximum depths of trees = 3, learning rate= 0.1, number of estimators = 100	57,5%
Multilayer perceptron neural network model with 1 hidden layer*	number of neurons in the hidden layer = 10, α = 1	57,1%
-‖- with 3 hidden layers*	number of neurons in the hidden layers = [10, 10, 30], α = 0.01	58,2%
-‖- with 5 hidden layers*	number of neurons in the hidden layers = [20, 20, 10, 10, 20], α = 0.01	59,0%
-‖- with 7 hidden layers*	number of neurons in the hidden layers = [10, 10, 30, 10, 50, 30, 70], α = 0.01	60,2%

a

b

c
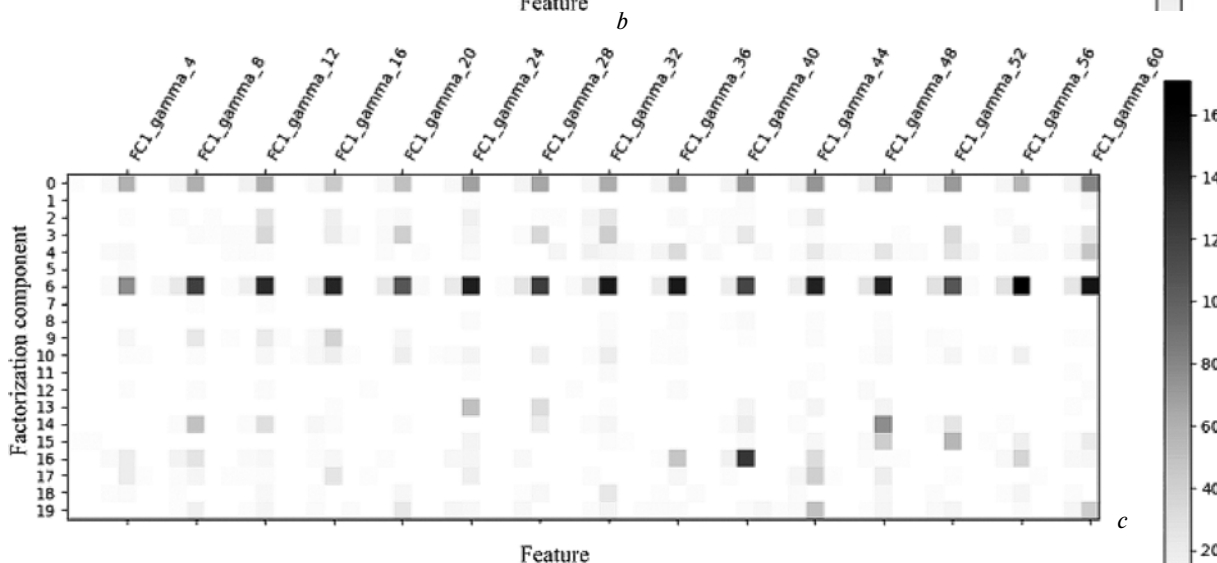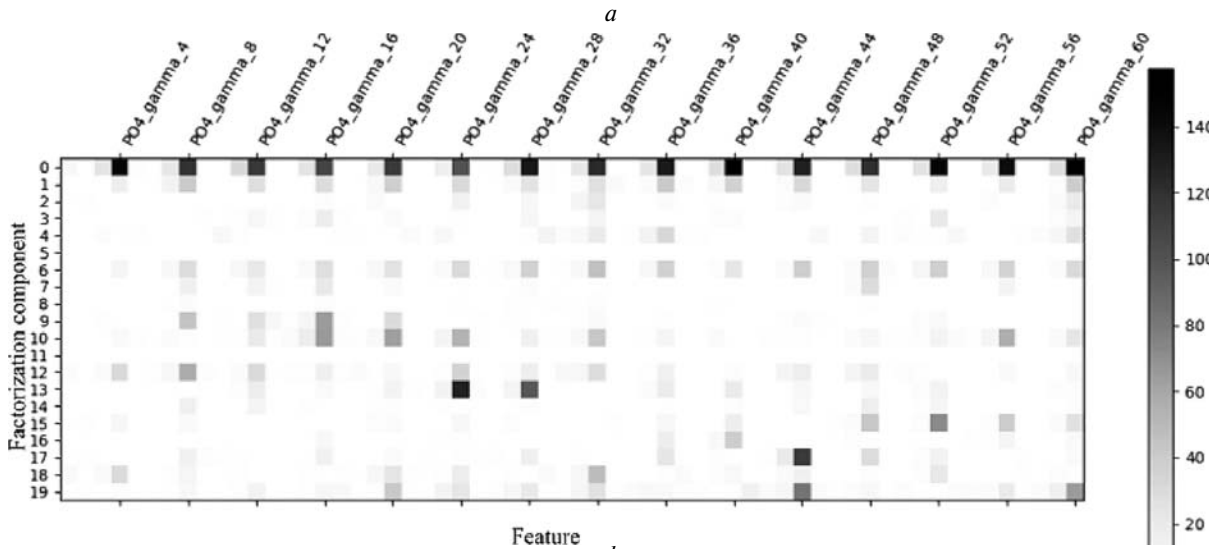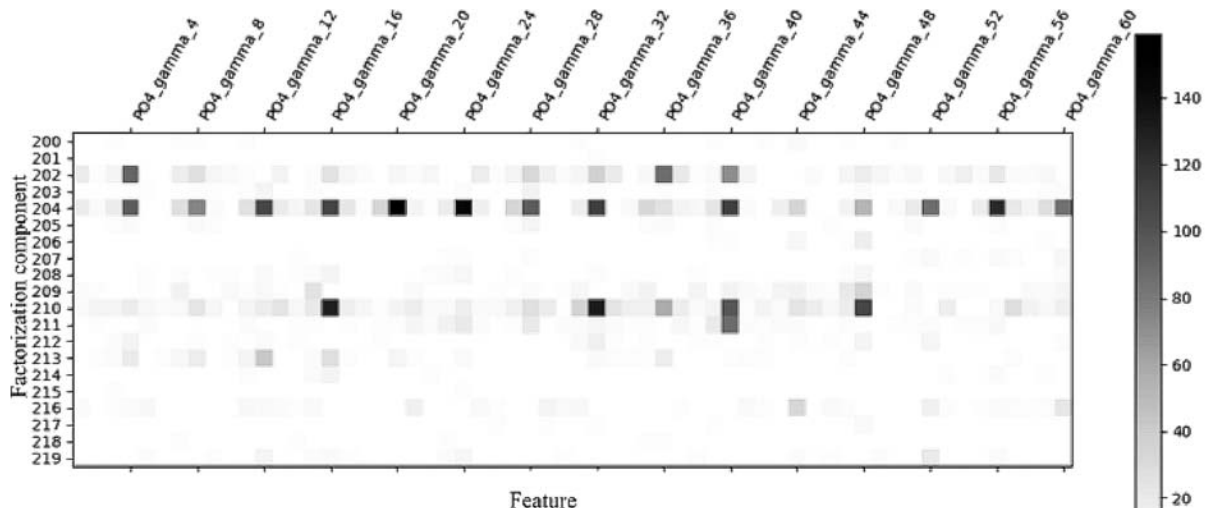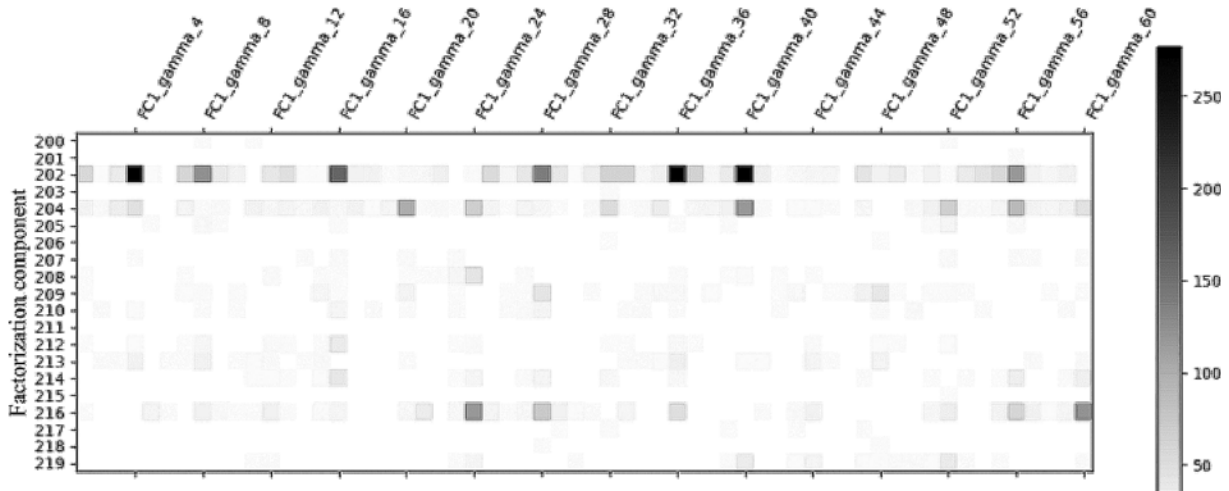
Fig. 5. The covariance coefficients applied to the initial characteristics: *a* — obtained from the reference point PO4 and applied to the components № 0-19; *b* — from PO4 to № 200-219; *c* — from FC1 to №0-19; *d* — from FC1 to № 200-219; (darker areas correspond to characteristics that have a higher covariance coefficient to the corresponding decomposition component. The names of the characteristics are formed as reference point + spectral range + the end timepoint of the segment)

negative [26]. At the same time, it is necessary to minimize the difference between the original matrix and the product of the resulting matrix:

$$f(W, H) = \frac{1}{2}\|Y - WH\|^2. \qquad (16)$$

This method resembles the *PCA* method. However, if in the *PCA* we need to obtain the orthogonal components explaining the maximum possible part of data variance, then in non-negative matrix factorization (*NMF*) we need non-negative decomposition components and covariance coefficients. Since the data in this study were previously normalized to the range [0, 1], this condition is satisfied. Fig. 5 shows partial examples of the matrix of covariance coefficients applied to the initial characteristics with a decrease in the number of significant features to 521. The total size of this matrix is 521 * 1920 values.

The process of decomposing data into a nonnegative weighted sum is especially useful for data created as a result of combining (or overlaying) several independent sources, which in our case is consistent with the nature of the data (the total electrical activity of many neurons).

Experiments have shown that the use of the method of non-negative matrix factorization made

it possible to increase the accuracy of the *SVM* model from 55.9 to 58.6%, while reducing the number of significant features to 521.

Genetic algorithm. This method refers to heuristic search algorithms, and is used to solve optimization and modeling problems by sequentially selecting, combining and varying the desired parameters using mechanisms resembling biological evolution. In general, the application of this algorithm can be divided into the following steps:

1. The choice of optimized parameters, and the generation of the initial genome. In this experiment — the genome consisted of 1923 binary genes $[X_1, ..., X_{1923}]$, where X_1, X_2, X_3 encoded one of the 8 variants of hyperparameter C from the logarithmic sequence [0.01, 0.03, 0.1, 0.3, 1 , 3, 10, 30] + 1920 genes responsible for the use of the corresponding feature in the model.

2. The choice of the fitness function $f(X_1, ..., ..., X_{1923})$. In this experiment — the accuracy function of the model for the linear SVM from hyperparameter C and the set of input features. In the case of the same accuracy of the model, the gene that had fewer features was considered as more fitted.

3. Cyclic selection of parameters of the function.

1.1. Generation of a "child" by mutation of the "parent" (inversion of m random genes). The pa-
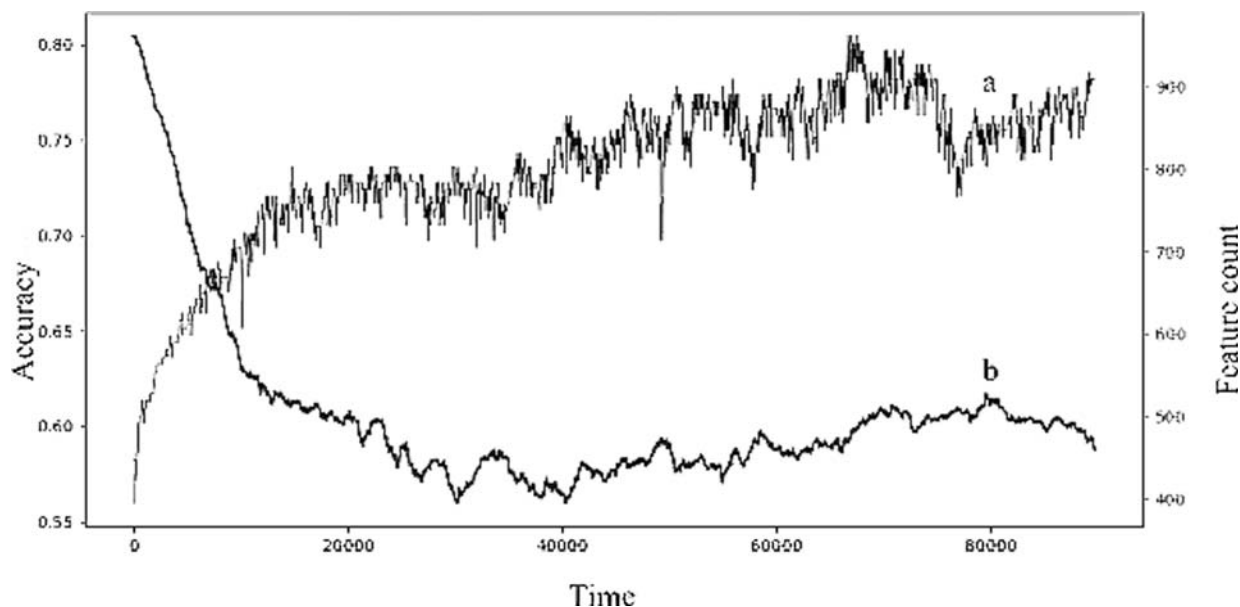
Fig. 6. Dependence of the number of significant features and the achieved accuracy of the SVM model on the total run time of the genetic algorithm: *a* — model accuracy; *b* — number of features.

rameter m affects the speed of movement along the slope of the fixture function.

1.2. The calculation of the fitness function of the device $f(X_1, ..., X_{1923})$ for the resulting "child".

1.3. Selection (selection of the most adapted individual from a descendant-ancestor pair).

1.4. Estimation of the age of the genome (the number of iterations in which a parent genome has the result of the fitness function better than a child genome). In case of reaching the limit — the destruction of the ancestor and its replacement by a descendant. This parameter affects the chance of the "stuck" of the optimizing function in local minima.

1.5. If the condition for stopping the cycle is satisfied, then the end of the cycle, otherwise the beginning of the cycle. Since the genetic algorithm refers to stochastic optimization methods, the condition for stopping is often the impossibility of finding a more adapted individual in k iterations or in a certain time interval. In this case, this parameter affects the total running time of the algorithm. In this experiment, the operation time of the algorithm was limited to one day.

The use of the genetic algorithm made it possible to increase the accuracy of the SVM model from 55.9 to 80.7%, while reducing the number of significant features to 478 (fig. 6).

The graphs of the achieved accuracy and the number of significant features show that the genetic algorithm quite quickly achieves high accuracy and reduces the set of significant features to ~ 1/4 of the initial number, after which further improvement slows down and becomes more random.

It is important to note that in the process of multiple repetition of the experiment, in a relatively short time (within 7200s), the genetic algorithm consistently found a set of parameters for the SVM that allows to obtain accuracy higher than any other previously used algorithm, including multilayer neural network models.

Conclusions

Conducted experiments on the analysis of the EEG signal to determine familiarity with the presented audiovisual data have shown that some of the machine learning methods do not allow achieving any significant accuracy. So the basic models of decision trees, naive Bayes and k-nearest neighbors did not show significant results, which is most likely due to the lack of clear distribution rules for features.

However, among many models without selection of features, ensemble and neural network models

have reached a significant accuracy. Neural network models accuracy increases with an increase in the number of the hidden layers. These results may indicate that the complexity of such process as thinking in general, and in particular the recognition of what had been seen, has the complex internal dependencies, which can only be seen by observing the whole process, and not its individual features. The high result of the ensemble models is also presumably related to the flexibility of these models in relation to non-uniformly distributed data.

Given the fact that the recognition of video-audio images is a complex process distributed in space (over brain regions) and time, increasing the accuracy of its determination is possible by reducing the dimensionality of the data and highlighting significant features.

Experiments have shown that the choice of the optimal (from a practical point of view) subspace of features can be achieved in a relatively short time using a genetic algorithm. When solving the problem posed, the genetic algorithm made it possible to isolate the signs with the greatest information and to increase the accuracy of the linear SVM model from 55.9 to 80.7%.

Further improvement of the result is possible by combining neural network models and a genetic algorithm as a search algorithm for building the model architecture. Strengthening the genetic algorithm for selecting significant features is possible by improving the mutation algorithms to speed up the search for a global maximum of the accuracy function from a set of features.

REFERENCES

1. Aleksandrov Y.I., 2012. Psihofiziologicheskie zakonomernosti naucheniya i metodyi obucheniya. Psihologicheskiy zhurnal. tom 33, #6, pp. 5—19
2. Liping Shen, Minjuan Wang and Ruimin Shen. 2013. Affective e-Learning: Using "Emotional" Data to Improve Learning in Pervasive Learning Environment. Journal of Educational Technology & Society Vol. 12, No. 2 (April 2009), pp. 176—189
3. Hadjidimitriou, S.K.; Hadjileontiadis, L.J. EEG-based classification of music appraisal responses using time-frequency analysis and familiarity ratings. IEEE Trans. Affect. Comput. 4, pp. 161—172.
4. Nattapong Thammasan, Koichi Moriyama, Ken-ichi Fukui, Masayuki Numao. 2017.Familiarity effects in EEG-based emotion recognition. Brain Informatics. pp. 4:39—50
5. Rostami Shahin, Shenfield Alex, Sigurnjak Steve, Fakorede Oluwatoyin, 2015. "Evaluation of mental workload and familiarity in human computer interaction with integrated development environments using single-channel EEG". In: 26th Annual Workshop of PPIG, Bournemouth University, July 2015.
6. Cheng Yang, Fang An, Chen Chen, Bin Zhu, 2017. The effect of product characteristic familiarity on product recognition. IOP Conf. Series: Materials Science and Engineering, 231.
   doi:10.1088/1757-899X/231/1/012016.
7. Meng-Hsien (Jenny) Lin, Samantha N.N. Cross, William J. Jones, Terry L. Childers, 2018. "Applying EEG in consumer neuroscience", European Journal of Marketing, 52 (1/2).
8. Jose Paulo Marques dos Santos, Marisa Martins, Hugo Alexandre Ferreira, Joana Ramalho, Daniela Seixas, 2016. "Neural imprints of national brands versus own-label brands", Journal of Product & Brand Management, 25 (2), pp. 184—195.
9. Novikova, S.I., 2015. "Ritmyi EEG i kognitivnyie protsessyi" [Elektronnyiy resurs]. Psihologicheskaya nauka i obrazovanie psyedu.ru, T. 4, 1, pp. 91—108, http://psyjournals.ru/jmfp/ 2015/n1/76178.shtml> [Accessed 20 Aug. 2018]. (In Russian).
10. Kahana, M.J., 2006. "The cognitive correlates of human brain oscillations". Journal of Neuroscience, 26 (6), pp. 1669—1672.
11. Naumov, R.A., 2010. Raspoznavanie tipov myislitelnoy deyatelnosti po EEG pri reshenii prostranstvennyih i verbalno-logicheskih zadach. Diss. kand. biol. nauk., 28 p. (In Russian).
12. Roik, A. O., Ivanitskiy, G. A., Ivanitskiy, A. M., 2012. "Kognitivnoe prostranstvo cheloveka: sovpadenie modeley, postroennyih na osnove analiza ritmov mozga i na psihometricheskih izmereniyah". Rossiyskiy fiziologicheskiy zhurnal im. I.M. Sechenova, 98 (11), pp. 1314—1328. (In Russian).
13. Sotnikov, P. I., 2014. Obzor metodov obrabotki signala elektroentsefalogrammyi v interfeysah mozg-kompyuter. Inzhenernyiy vestnik, pp. 13—24. (In Russian).
14. Bureeva, N.N., 2007. "Mnogomernyiy statisticheskiy analiz s ispolzovaniem PPP "STATISTICA"". Uchebno-metodicheskiy material po programme povyisheniya kvalifikatsii "Primenenie programmnyih sredstv v nauchnyih issledovaniyah i prepodavanii matematiki i mehaniki". Nizhniy Novgorod, 112 p. (In Russian).

15. Ermachenko, N.S., Ermachenko, A.A., Latanov, A.V., 2011. "Desinhronizatsiya EEG na chastote alfa-ritma kak otrazhenie protsessov zritelnogo selektivnogo vnimaniya". Fiziologiya cheloveka, 37(6), pp. 18—27 (In Russian).

16. Naumov R. A. 2010. "Raspoznavanie tipov myislitelnoy deyatelnosti po EEG pri reshenii prostranstvennyih i verbalno-logicheskih zadach". Diss. kand. biol. nauk. T. 3. — #. 01.

17. Casey, B.J., Giedd, J.N., Thomas, K.M., 2000. "Structural and functional brain development and its relation to cognitive development". Biological psychology, 54 (1—3), pp. 241—257.

18. Norkin, V.I., Kayzer, M.A., 2009. "Ob effektivnosti metodov klassifikatsii, osnovannyih na minimizatsii empiricheskogo riska". Kibernetika i sistemnyiy analiz, 5, pp. 93—105. (In Russian).

19. Bartlett, P., Shawe-Taylor, J., 1998. Generalization performance of support vector machines and other pattern classifiers. Advances in Kernel Methods. MIT Press, Cambridge, USA.

20. Voronov, I.V., Politov, E.A., Efremenko, V.M. "Obzor tipov iskustvennyih neyronnyih setey i metodov ih obucheniya". Elektrotehnicheskie kompleksyi i sistemyi, pp. 38—42. (In Russian).

21. Chistyakov, S.P., 2013. "Sluchaynyie lesa: obzor". Trudyi Karelskogo nauchnogo tsentra RAN, 1, pp. 117—136. (In Russian).

22. Dyakonov, A.G., 2010. "Analiz dannyih, obuchenie po pretsedentam, logicheskie igryi, sistemyi WEKA, RapidMiner i MatLab": Uchebnoe posobie. M.: Izdatelskiy otdel fakulteta VMK MGU imeni M.V. Lomonosova. (In Russian).

23. Koelstra, S., Muehl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., Patras, I., 2012. "DEAP: A Database for Emotion Analysis using Physiological Signals", IEEE Transactions on Affective Computing, 3 (1), pp. 18—31.

24. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., 2011. "Scikit-learn: Machine Learning in Python", JMLR, 12, pp. 2825—2830.

25. Kalinina, V.N., Solovev, V.I., Kolemaev, V.A., Shevelev, V.V., Mihalev, B.G., 2003. Vvedenie v mnogomernyiy statisticheskiy analiz, 66 p. (In Russian).

26. Renaud Gaujoux et al., 2010. A flexible R package for nonnegative matrix factorization. In: BMC Bioinformatics, p. 367.

Решетников Д.С., аспірант,
Міжнародний науково-навчальний центр
інформаційних технологій і систем НАН України та МОН України,
просп. Глушкова, 40, Київ 03187, Україна,
denis.reshetnykov@gmail.com

АНАЛІЗ ЕЛЕКТРОЕНЦЕФАЛОГРАМ У ЗАДАЧІ ОЦІНКИ
СТАНУ ОЗНАЙОМЛЕННЯ УЧНЯ З АУДІО ТА ВІДЕОДАНИМИ

Вступ. Оцінка емоційного стану учнів в цілому і динамічна оцінка ознайомлення з матеріалами зокрема, дозволяє краще оцінити засвоєння теоретичного матеріалу і швидкість оволодіння вміннями, які потребують багаторазового повторення матеріалу.

Контроль навчального процесу був би значно ефективніше, якби була можливість об'єктивно оцінити, за сукупністю ознак, знайомство конкретного учня з представленими йому матеріалами.

Мета статті. Виконати порівняльний аналіз і експериментальне дослідження ефективності різних методів машинного навчання для побудови моделі визначення знайомства з аудіовізуальними матеріалами, на основі аналізу сигналу електроенцефалограм і визначити набір ознак, які найкраще класифікують даний сигнал.

Методи. Запропоновано інформаційну технологію, яка на підставі набору реальних ЕЕГ бази даних DEAP виконує побудову, підбір гіперпараметрів і оцінку різних моделей класифікації стану ознайомлення учня з аудіовізуальними даними. Для лінійного методу опорних векторів реалізовано інформаційну технологію відбору діагностичних ознак на підставі методу головних компонент, невід'ємної матричної факторизації і генетичного алгоритму.

Результат. За використання запропонованої інформаційної технології підібрано параметри і отримано результати точності для різних моделей класифікації, що дозволило порівняти такі моделі і визначити найбільш адекватні до вирішення поставленої задачі. Застосування методів відбору ознак дозволило підвищити точність моделі лінійного методу опорних векторів з 55,9 до 80,7відсотків.

Висновок. Запропонована інформаційна технологія аналізу сигналу ЕЕГ для ознайомлення з представленими аудіовізуальними даними показала високу ефективність ансамблевих і нейромережевих моделей в даній задачі. Підвищення точності класифікації стану ознайомлення з поданими аудіо-відеоданими можливо шляхом знижен-

ня розмірності даних і виділення значимих ознак, що експериментально підтверджено на прикладі застосування генетичного алгоритму до відбору ознак методом опорних векторів.

Ключові слова: *машинне навчання, генетичний алгоритм, електроенцефалограма, набір данних DEAP, розпізнавання емоцій.*

Решетников Д.С., аспирант,
Международный научно-учебный центр
информационных технологий и систем НАН Украины и МОН Украины,
просп. Глушкова, 40, Киев 03187, Украина,
denis.reshetnykov@gmail.com

АНАЛИЗ ЭЛЕКТРОЭНЦЕФАЛОГРАМ В ЗАДАЧЕ ОЦЕНКИ СОСТОЯНИЯ ОЗНАКОМЛЕНИЯ УЧАЩИХСЯ С АУДИО-ВИДЕО ДАННЫМИ

Введение. Оценка эмоционального состояния учащихся в целом и динамическая оценка ознакомления с материалами в частности, позволяет лучше оценить усвоение теоретического материала и скорость овладения умениями, требующими многократного повторения.

Контроль учебного процесса был бы значительно эффективнее при возможности объективно оценить уровень ознакомления конкретного учащегося с материалами по совокупности признаков.

Цель статьи. Выполнить сравнительный анализ и экспериментальное исследование эффективности различных методов машинного обучения для построения модели определения знакомства с представленными аудиовизуальными материалами, на основе анализа сигнала электроэнцефалограмм и определить набор признаков, наилучшим образом классифицирующих данный сигнал.

Методы. Предложена информационная технология, которая на основании набора реальных ЭЭГ базы данных DEAP выполняет построение, подбор гиперпараметров и оценку различных моделей классификации состояния ознакомления учащегося с аудио-визуальными данными. Для линейного метода опорных векторов реализована информационная технология отбора диагностических признаков на основе метода главных компонент, неотрицательной матричной факторизации и генетического алгоритма.

Результат. С использованием предложенной информационной технологии, подобраны параметры и получены результаты точности для различных моделей классификации, что позволило сравнить такие модели и определить наиболее адекватные решению поставленной задачи. Применение методов отбора признаков позволило повысить точность модели линейного метода опорных векторов с 55,9 до 80,7 процентов.

Выводы. Предложенная информационная технология анализа сигнала ЭЭГ для определения степени ознакомления с представленными аудиовизуальными данным показала высокую эффективность ансамблевых и нейросетевых моделей в данной задаче. Повышение точности классификации состояния ознакомления с представленными аудио-видео данным возможно путем снижения их размерности и выделения значимых признаков, что экспериментально подтверждено на примере применения генетического алгоритма для отбора признаков методом опорных векторов.

Ключевые слова: *машинное обучение, генетический алгоритм, электроэнцефалограмма, набор данных DEAP, распознавание эмоций.*