# User keystroke authentication and recognition of emotions based on Convolutional Neural Network

Ihor Tereikovskyi[1] [0000-0003-4621-9668], Liudmyla Tereikovska[2] [0000-0002-8830-0790], Oleksandr Korystin[3] [0000-0001-9056-5475], Shynar Mussiraliyeva[4] [0000-0001-5794-3649], Aizhan Sambetbayeva[5] [0000-0003-0032-0533]

[1]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
[2]Kyiv National University of Construction and Architecture, Kyiv, Ukraine
[3]Scientifically Research Institute of the Ministry of Internal Affairs, Kyiv, Ukraine
[4,5]Al-Farabi Kazakh National University, Almaty, Kazakhstan
terejkowski@ukr.net, tereikovskal@ukr.net, alex@korystin.pro, mussiraliyevash@gmail.com, aizhan.sambetbaeva@kaznu.kz

**Abstract.** The article is devoted to the problem of improving Biometric identification systems based on Keystroke Dynamics for recognizing emotions and authenticating users of information systems through the implementation of modern neural network solutions based on Convolutional Neural Network (CNN). It is established that the difficulties of such implementation are associated with coding the keystroke parameters to a form suitable for CNN processing. A coding procedure based on the presentation of fixed-size keystroke parameters in the form of a color square image is proposed. Each encoded text symbol corresponds to a separate point of the image and is characterized using the corresponding ASCII code and keystroke parameters such as the key hold time and the time between keystrokes. Experimental studies showed that the proposed coding procedure made it possible to use CNN with the LeNet architecture for analyzing Keystroke Dynamics, which provided an error in recognizing the user's emotions and personality at the level of the best modern recognition systems.

**Keywords:** recognition of emotions, biometric authentication, keystroke dynamics, convolutional neural network.

## 1. Introduction

Currently, one of the most urgent tasks in the field of information technology is the development of effective emotion recognition tools (ERT) for users of information systems (IS) for various purposes. These tools are necessary, for example, for the operational monitoring of operators of critical infrastructure, where a significant number of accidents and emergency incidents are associated with a violation of their emotional state [1,2]. Another example is the use of ERT in distance learning IS for automatic control of the perception of educational materials.

ERT on the basis of Keystroke Dynamics has a quite broad prospects because of using standard peripheral equipment to obtain biometric characteristics, widespread using of text information in information systems, inalienability of the owner's identity, complexity of biometric information falsification, the possibility of covert monitoring during of professional activity [1-4].

## 2. Analysis of Literature Sources in the Field of Research

The concept of Keystroke Dynamics can be understood as user's individual biometric behavioral characteristic, which determines the features of typing text from the keyboard [4,5]. It is believed that each person has his own keystroke dynamics, which is described by the speed and dynamics of typing, the transition times between several (two or more) keys and typing errors….

When using universal keyboard input tools, the key code, the character corresponding to the key, the time the key is pressed and the time it is released can be registered for determining the keystroke parameters. In specialized systems of keyboard input for this, parameters that characterize the user's pressure on the key and the speed / acceleration of pressing can additionally be used. In the future, before entering the recognition system, these parameters are subject to filtering and primary processing, during which some general keystroke parameters are calculated.

A typical filtering procedure is given in [3-6] and consists in applying the so-called frequency, temporal, and keyboard filters. Analysis of sources [4-12] suggests that today there is no generally accepted methodology for forming a set of input data of the Keystroke Dynamics recognition module. In this case, a common feature of the most well-known Keystroke Dynamics recognition systems is the use of filtered values of the hold time of individual keys (TIK) and the time between pressing two and three separate keys (TMK) as input data. It is also possible to determine the dependence of the input data nomenclature on the mathematical support of the recognition process, which reduces to a comparison of the input sequence with the Keystroke Dynamics reference of a certain emotion of the IS user. There are two types of models of recognition: by a predetermined text and by the text of arbitrary content. In both cases, to determine Keystroke Dynamics patterns, user must enter one or several fragments of the same text several times.

Most of the known patterns are statistical models of Keystroke Dynamics parameters, for example, based on the normal or bimodal distribution [4-6]. In the case of Keystroke Dynamics recognition on the basis of a certain text fragment, the basis of the patterns, as a rule, is made up of the indicators of TIK and TMK concerning the sequential order of keystrokes. When recognizing an arbitrary content on the basis of a fragment of text, the patterens, for the most part, are formed on the basis of the statistical indicators of the TIK and TMK of individual stable sequences of keyboard events, reflecting the features of the typing dynamics of an individual user. However, in [4,7], on the basis of experimental studies, the unsatisfactory adaptation of statistical models to changes in the user's emotional state and the complexity of their formation in the case of analyzing a text fragment of arbitrary content was proved. It is also shown that the complexity of the Keystroke Dynamics recognition problem is explained by the need to analyze rather large volumes of multidimensional data. At the same time, just from the point of view of the proven efficiency of analyzing multidimensional data, a promising direction for improving Keystroke Dynamics recognition systems is the use of neural network models [8-10]. This is confirmed by the data of works [6,11], in which, along with the positive results of the application of neural network models, their limitations associated with the difficulties of forming the nomenclature of input parameters when analyzing the Keystroke Dynamics of an arbitrary text. You can also note the obsolescence of the used neural network models of the multilayer perceptron type, probabilistic neural networks, Kohonen maps, Hopfield networks, RBF networks and fully connected deep neural networks with direct propagation of the signal [4,11,12]. It is also worth noting the possibility of Keystroke Dynamics recognizing with the help of neural network models of type LSTM [13,14], which allows processing texts of arbitrary length. However, the construction of such neural network models is associated with the complexity of the formation of an academic sample.

At the same time, based on the methods of developing neural network information protection tools, we can assume the expediency of using the neural network models type of convolutional neural network (CNN) in ERT. So in [14-17], a method for converting keystroke data into an image was proposed for subsequent use in CNN for user authentication. There is an opportunity on fixed texts to achieve recognition accuracy of 96.8%. Similar results were also obtained in [18], in which CNN was used to analyze Keystroke Dynamics in order to increase the resilience of user password protection. At the same time, in the works [14-16] listed above, the problems of the influence of the user's emotional state on the authentication accuracy, as well as the coding of CNN input parameters, are not fully covered. Therefore, the main objective of this study is to develop a procedure for coding the input parameters of a convolutional neural network to recognize the emotions and personality of the user using keyboard handwriting.

## 3. Coding procedure

A feature of the classic version of CNN is the need to present the input information in the form of a square image. The basic version uses a black and white image. More sophisticated options involve the use of three-dimensional gray and color images. This feature imposes a significant limitation on the use of CNN - the ability to analyze Keystroke Dynamics on text fragments with a fixed number of characters. In this case, the general formulation of the problem of recognizing the emotions of the user of the IS implies the need to analyze both a predetermined fragment of the text and a fragment of the text of arbitrary content. The first case can be correlated with monitoring the emotional state of a user when he enters password data. The second case correlates with the current monitoring of the user's personal / emotional state when he enters text information from the keyboard. In accordance with the data of

[4, 6, 15], such monitoring is feasible due to the analysis of Keystroke Dynamics with the introduction of text of a fixed length even when using relatively simplified statistical models. Thus, the above restriction on a fixed number of characters practically does not adversely affect the functionality of CNN. In the first case, the number of characters in the password data is already a fixed value. In the second case, the volume of a fragment of text of arbitrary content to be analyzed is limited to a predetermined number of characters.

Let us consider the principle of coding Keystroke Dynamics parameters into a black and white square image on which the proposed coding procedure is based. It is proposed to correlate the ordinate axis with the keyboard layout — ASCII codes of keys or characters corresponding to keys. X-axis  proposed to correlate with the entered text. Thus, a single character of the entered text will correspond to one separate image point. On the axis of the co-ordinate of the encoded symbol corresponds to the position (number) of this symbol in the text. The coordinate of the ordinate corresponds to the position of the character on the keyboard / ASCII code of the previous character in the text. It is assumed that the first character on the keyboard corresponds to the space character. In the case when the number of characters of the text will be more than the number of characters on the keyboard / ASCII codes, then to save the square shape, the figure above the axis of ordinates is complemented by strings that correspond to the space character. If the number of characters of the text is less than the number of characters on the keyboard, then to save the form, the picture on the right will be supplemented with columns with space characters.

An illustration of the proposed method is Figure 1, which shows a two-dimensional black-and-white image of the encoded text "AUTOMOBILE". To simplify the demonstration, the coding adopted the assumption that it is necessary to analyze the text, which consists exclusively of the capital letters of the English alphabet and the space character. In this figure, auxiliary fragments are highlighted in gray on which the characters / character number on the keyboard and the character number / character of the text to be encoded are displayed along the vertical axis. For clarity, the individual points of the image are separated by straight lines.

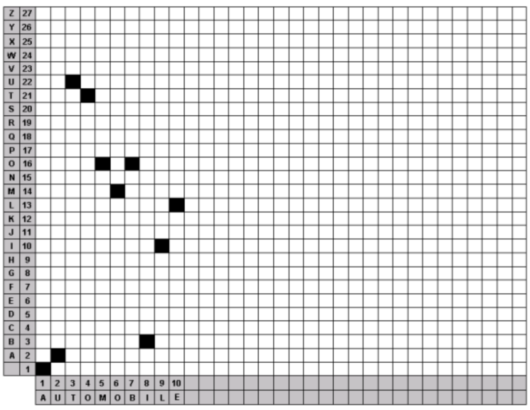


**Fig. 1.** Black and white display of the coded word "AUTOMOBILE"

Each dot of the image filled with black corresponds to the coded value of the text symbol. For example, the symbol “T” corresponds to the black point of the image located at the intersection of the vertical, drawn from position No. 3 (“T”) along the axis of the X-axis, and horizontal, drawn from position No. 22 (“U”) to the axis of ordinates. This is explained by the fact that in the word “AUTOMOBILE” the letter “T” follows the letter “U”. In numerical form, this figure represents a square matrix. The black spots of the pattern correspond to the elements of the matrix equal to 1, and the bright points to the elements of the matrix equal to 0.

The coding of the entered text in the form of a black and white image does not allow to realize the recognition of Keystroke Dynamics, which involves analysis, as the minimum of one of the main parameters of Keystroke Dynamics. Therefore, the proposed coding procedure involves presenting the entered text as a colored square image with a multichannel raster. In the base case, the procedure assumes two raster channels. Each point of such an image should characterize the entered symbol and one of the Keystroke Dynamics parameters relating to this symbol.

Based on the results [3-6], the expediency of the use of indicators of a TIK or TMK is determined, the calculation of which is implemented using the expressions:

$$y_r(i) = t_d(i) - t_u(i), \tag{1}$$

$$y_b(i, i+1) = t_d(i) - t_d(i+1), \tag{2}$$

where $y_r$ − key hold time; $t_d$ − key pressed time; $t_u$ − key release time; $y_b$ − time between successive pressing of two keys; $i$ − keystroke number when entering text.

We note the complexity of using and, which is explained by the necessity of their registration in universal computer systems with an accuracy of millisecond. Therefore, in accordance with the recommendations of [3, 6], a Windows-oriented program was developed that allows registering values with an error equal to the duration of 50 process cycles. For an example in fig. 2 and fig. 3, it is shown the histograms of the TIK and TMK values for the text “HEY THERE”, the input of which was recorded using the specified program.

Using of this program allowed to establish that the values of the TIK and TMK do not exceed 500 ms. Values of TIK and TMK which exceed this value should be filtered. It should be noted a significant dependence of TIK and TMK on the keyboard type. So, according to [6, 8], on the keyboard shortcuts (laptop keyboard), on average, $y_r$ = 100 ms, $y_d$ = 150 ms, and for a keyboard with a long key stroke (standard keyboard), $y_r$ = 150 ms, $y_d$ = 200 ms. The $y_r$ and $y_d$ values obtained in the course of this study are approximately 1.5–2 times smaller. Therefore, the normalization of TIK and TMK to average values may entail errors associated with the characteristics of the keyboard used. Thus, at the entrance to CNN it is advisable to submit the absolute filtered values of the TIK and TMK. In the first approximation, the introduced symbol is proposed to be represented as a number corresponding to the ASCII code.

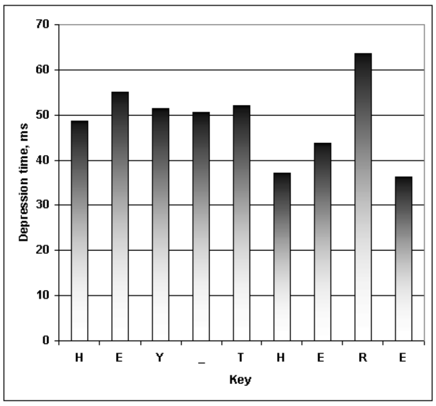


**Fig. 2.** Histogram of values of TIK for the text "HEY THERE"

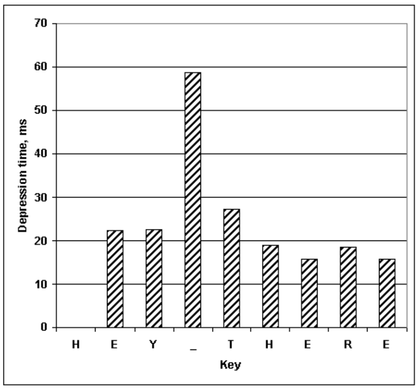


**Fig. 3.** Histogram of TMK values for the text "HEY THERE"

To illustrate the results of the coding procedure described in fig. 4 is a fragmentary representation of the display of the text "HEY THERE", encoded using TIK. Unlike fig. 1 each point of the image corresponding to the encoded value of the character of the text, is characterized by two numbers recorded in the corresponding cell. The first digit is the ASCII – code of the character entered, and the second is the TIK. For example, the "Y" symbol corresponds to a dot in an image located at the intersection of the third column with the top row. The values of 121 (ASCII – code) and 51 (TIK) are presented in the corresponding cell.

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Z | 27 | | | | | | | | | | | | |
| Y | 26 | | | | 32;51 | | | | | | | | |
| X | 25 | | | | | | | | | | | | |
| W | 24 | | | | | | | | | | | | |
| V | 23 | | | | | | | | | | | | |
| U | 22 | | | | | | | | | | | | |
| T | 21 | | | | | | 104;37 | | | | | | |
| S | 20 | | | | | | | | | | | | |
| R | 19 | | | | | | | | | 101;36 | | | |
| Q | 18 | | | | | | | | | | | | |
| P | 17 | | | | | | | | | | | | |
| O | 16 | | | | | | | | | | | | |
| N | 15 | | | | | | | | | | | | |
| M | 14 | | | | | | | | | | | | |
| L | 13 | | | | | | | | | | | | |
| K | 12 | | | | | | | | | | | | |
| J | 11 | | | | | | | | | | | | |
| I | 10 | | | | | | | | | | | | |
| H | 9 | | 101;55 | | | | | 101;44 | | | | | |
| G | 8 | | | | | | | | | | | | |
| F | 7 | | | | | | | | | | | | |
| E | 6 | | | 121;51 | | | | | 114;64 | | | | |
| D | 5 | | | | | | | | | | | | |
| C | 4 | | | | | | | | | | | | |
| B | 3 | | | | | | | | | | | | |
| A | 2 | | | | | | | | | | | | |
| | 1 | 104;49 | | | | 116;52 | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| | | H | E | Y | | T | H | E | R | E | | | |

**Fig. 4.** Display of the text "HEY THERE" encoded using the TIK parameter as an image with a two-channel raster

A variant of the coding procedure for the entered text in the form of a colored square image with a multichannel raster is the development of the basic case in the direction of adding channels that correspond to the parameters of the CP. The coding result in numerical form is a multidimensional matrix, the depth of which is equal to the number of parameters used in the Keystroke Dynamics recognition. In order to verify the proposed coding procedure, we performed experiment for user keystroke authentication and recognition of his emotional state. The following terms and conditions apply to CNN: the number of recognizable users is 8; three type of emotions are to be recognized - neutral, joy and fear; text may consist of capital letters of the English alphabet and 6 punctuation marks (a total of 32 characters); the length of analyzed texts is 28 characters; keystroke dynamics is described by the TIK and TMK parameters. These conditions are determined from the standpoint of the estimated nature of experimental studies, simplifying the formation of a training sample and the possibility of a correct representation of the CNN input field. Based on the recommendations [9, 10, 19] the well-known LeNet architecture with the following parameters was used for the experiments: input field size - 32x32, number of output neurons - 11, number of convolution layers - 2, number of subsampling layers - 2, the number of fully connected layers is 2, the size of the convolution kernels is 5x5, the number of convolution maps in the first layer is 6, the number of convolution maps in the second layer is 16, the number of neurons in the first fully connected layer is 120, the number of neurons in the second fully connected layer is 84.

The neural network model was implemented using the MATLAB 2018 application software package. For its training, a database of filtered keyboarding samples was used, corresponding to the three indicated emotions for 8 people. 80 records of keyboarding (10 entries per person) for the same text was used to represent one emotion. Watching appropriate multimedia provided a certain emotion of tested users. 90% of the database records were used to form the training sample, the remaining 10% were used for the test sample. The magnitude of errors in recognition of the user's identity and emotional state, calculated from the results of the experiments performed are shown in fig. 5.

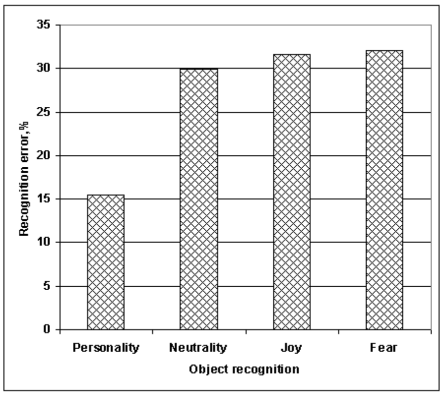


**Fig. 5.** Recognition Error Histograms

Note that although the experiments in this study used rather outdated neural network solutions, the errors obtained correlate with the errors of the best modern systems [3, 6, 11-18]. At the same time, the recognition accuracy of the neural system used in the experiments was negatively affected by the relatively small amount of the training set, the formation of which is a relatively resource-intensive process. This suggests that a fairly low recognition error was achieved due to the use of the proposed Keystroke Dynamics coding procedure. In addition, the experiments conducted showed the possibility of integral recognition of the emotions and personality of the operator. Thus, the results of the experiments confirm the promise of using the proposed method of coding, both in emotion recognition systems, and in user authentication systems. It can also be argued that the development of the Keystroke Dynamics coding procedure allows us to correlate the paths of further research with the development of a method for adapting the architectural parameters of modern CNN types to the problem of integral recognition of emotions and user authentication on the base of Keystroke Dynamics.

## 4. Conclusions

The article is devoted to solving the problem of improving the analysis systems on the base of Keystroke Dynamics for recognizing emotions and authenticating users of information systems through the introduction of modern neural network solutions based on convolutional neural networks. It has been established that the difficulties of such an implementation are connected with the need to encode the parameters of keyboard typing to a form suitable for processing a convolutional neural network. The coding procedure based on the representation of the parameters of the keyboard typing of the entered text of a fixed size in the form of a color image is substantiated. Each encoded character of the text is related to a separate point of the image and is characterized using the corresponding ASCII code and Keystroke Dynamics parameters. At the same time, the coordinate of the encoded symbol corresponds to the position of the given symbol in the text along the x-axis. The coordinate on the vertical axis corresponds to the position of the character on the keyboard of the previous character in the text. Experimental studies have shown the promise of using the proposed coding procedure, both in emotion recognition systems, and in user authentication systems. It is proposed to correlate the paths of further research with the development of a method for adapting the architectural parameters of modern types of convolutional neural networks to the problem of integral recognition of emotions and user keystroke authentication.

## References


1. S. Gnatyuk, Critical Aviation Information Systems Cybersecurity, Meeting Security Challenges Through Data Analytics and Decision Support, NATO Science for Peace and Security Series, D: Information and Communication Security. - IOS Press Ebooks, Vol.47, №3, pp. 308-316, 2016.
2. S. Gnatyuk, V. Sydorenko, M. Aleksander, Unified data model for defining state critical information infrastructure in civil aviation, Proceedings of the 2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT), Kyiv, Ukraine, May 24-27, 2018, pp. 37-42.
3. Paweł Kobojek and Khalid Saeed, Application of Recurrent Neural Networks for User Verification based on Keystroke Dynamics Journal of telecommunication and information technology, N3 2016 pp. 80-90.
4. Ivanov A.I. Nejrosetevye algoritmy biometricheskoj identifikacii lichnosti. Kn. 15: Monografiya / A.I. Ivanov. – M.: Radiotehnika, 2004. – 144 s.
5. N.A. Koshevaya, N.I. Maznichenko Podhod k povysheniyu nadezhnosti identifikacii polzovatelej kompyuternyh sistem po dinamike napisaniya parolej Sistemi obrobki informaciyi, 2014, vipusk 6 (122) c. 140-146
6. Savinov, A.N. Matematicheskaya model mehanizma raspoznavaniya klaviaturnogo pocherka na osnove Gaussovskogo raspredeleniya / A.N. Savinov, I.G. Sidorkina // Izvestiya Kabardino-Balkarskogo nauchnogo centra RAN. Vyp. I. - Nalchik: Kabardino-Balkarskij nauchnyj centr RAN, 2013. - S. 26-32.
7. Aitchanov, B., Korchenko, A., Tereykovskiy, I., Bapiyev, I. Perspectives for using classical neural network models and methods of counteracting attacks on network resources of information systems. (2017). // News of the national academy of sciences of the republic of Kazakhstan series of geology and technical sciences. Volume 5, Number 425 (2017), 202 – 212.
8. Saket Maheshwary, Soumyajit Ganguly, Vikram Pudi, Deep Secure: A Fast and Simple Neural Network based approach for User Authentication and Identification

via Keystroke Dynamics Conference: IWAISe, International Joint Conference on Artificial Intelligence (IJCAI) - 2017At: Melbourne, Australia pp. 34-40.

9. Dychka, I., Tereikovskyi, I., Tereikovska, L., Pogorelov, V., Mussiraliyeva, S. De-obfuscation of computer virus malware code with value state dependence graph // Advances in Intelligent Systems and Computing. 2018. Volume 754, pp 370-379.
10. Tereikovskyi I, Chernyshev D., Tereikovska L.A., Mussiraliyeva S, Akhmed G. The Procedure for the Determination of Structural Parameters of a Convolutional Neural Network to Fingerprint Recognition. Journal of Theoretical and Applied Information Technology, No 8, Volume 97, 2019, pp. 2381-2392.
11. Berik Akhmetov, Igor Tereykovsky, Aliya Doszhanova, Lyudmila Tereykovskaya (2018) Determination of input parameters of the neural network model, intended for phoneme recognition of a voice signal in the systems of distance learning. International Journal of Electronics and Telecommunications. Vol 64, No 4 (2018), 425-432. DOI: 10.24425/123541.
12. S. J. Alghamdi and L. A. Elrefaei. Dynamic user verification using touch keystroke based on medians vector proximity. In Computational Intelligence, Communication Systems and Networks (CICSyN), 2015 7th International Conference on, pages 121–126. IEEE, 2015.
13. C. Bo, L. Zhang, T. Jung, J. Han, X.-Y. Li, and Y. Wang. Continuous user identification via touch and movement behavioral biometrics. In 2014 IEEE 33rd International Performance Computing and Communications Conference (IPCCC), pages 1–8. IEEE, 2014.
14. Yunbin Deng and Yu Zhong Keystroke Dynamics Advances for Mobile Devices Using Deep Neural Network GCSR Vol. 2, pp. 59-70, 2015 DOI: 10.15579/gcsr.vol2.ch4.
15. Xiaofeng, L., Shengfei, Z., Shengwei, Y. Continuous authentication by free-text keystroke based on CNN plus RNN Procedia Computer Science 147, pp. 314-318 2019
16. Liu, M., Guan, J. User keystroke authentication based on convolutional neural network , Communications in Computer and Information Science 2019, 971, pp. 157-168.
17. Lin, C.-H., Liu, J.-C., Lee, K.-Y. On neural networks for biometric authentication based on keystroke dynamics. Sensors and Materials 2018 30(3), pp. 385-396
18. Hayreddin Çeker ; Shambhu Upadhyaya Sensitivity analysis in keystroke dynamics using convolutional neural networks 2017 IEEE Workshop on Information Forensics and Security (WIFS) 4-7 Dec. 2017 Page(s): 1 – 6 DOI: 10.1109/WIFS.2017.8267642.
19. Tereykovska L., Tereykovskiy I., Aytkhozhaeva E., Tynymbayev S., Imanbayev A. Encoding of neural network model exit signal, that is devoted for distinction of graphical images in biometric authenticate systems (2017). // News of the national academy of sciences of the republic of Kazakhstan series of geology and technical sciences. Volume 6, Number 426 (2017), 217 – 224.