

## MEDICAL IMAGES OF BREAST TUMORS DIAGNOSTICS WITH APPLICATION OF HYBRID CNN–FNN NETWORK

YU. ZAYCHENKO, G. HAMIDOV, I. VARGA

**Abstract.** The problem of classification of breast tumors on medical images is considered. For its solution the new class of convolutional neural networks-hybrid CNN–FNN network is developed in which convolutional neural network VGG-16 is used as the feature extractor while fuzzy neural network NEFClass is used as the classifier. Training algorithms of FNN were implemented. The experimental investigations of the suggested hybrid network on the standard data set were carried out and comparison with known results was performed. The problem of data dimensionality reduction is considered and application of PCM method is investigated.

**Keywords:** medical diagnostics, breast cancer classification, FNN, CNN, hybrid network, dimensionality reduction, PCM.

### INTRODUCTION. STATE-OF-ART PROBLEM ANALYSIS

Now cancer constitute the great problem for health defense all over the world. Basing on the data of IARC (International Agency of Cancer Research) 8,2 million death cases were registered in year 2012, 27 million new cases of illness are expected till 2030 [1]. Among the different types of cancer breast cancer takes the second place by its occurrence in women . Besides, mortality of it is very high as compared with other cancer diseases [1].

Nowadays, in practice, at every stage of diagnostics information technologies are utilized. The main goal of medical automated systems are extension of spheres of practical tasks which may be solved with computers aid, raise of level intellectual decision support of doctors in particularly in process of express diagnostics based on processing and analysis of medical images of human tissue obtained by different source ( MRT, CT etc).

In medical diagnostics problems substantial amount of problem constitute the features extraction for further processing and the choice of features for classification method. With development and wide dissemination of decision-support systems the demands to training algorithms are increasing. Reliability and simplicity of application influence on speed and quality of decision- making which is very important for express medical diagnostics. The advantages of medical diagnostics systems are speed, automation and stability of work which make them very comfortable tools for express medical diagnostics. Despite of young age of medical informatics which doesn't exceed 30 years information technologies in a whole are fast penetrating in various spheres of medicine and health defence (family medicine, insurance medicine, building unified information space, integration in European medical space etc).

Despite of progress which was achieved by diagnostics technologies final diagnosis of breast cancer including classification of tumours and diagnosis still is

performed by pathologist-anatomists which use visual analysis of histological patterns by microscope. The latest achievements in images processing technologies and machine learning enable to construct systems of automatic detection and diagnostics that may help pathologist-anatomists to make true diagnosis and accelerate his work. Classification of images histopathology on different patterns which corresponds to cancer and not-cancer states of tissue is often first rank goal in images analysis systems for automatic cancer diagnostics.

Up to date several models and methods were developed for breast cancer detection using various machine learning algorithms. Using such methods and technologies of AI as neuron networks and SVM [2, 3] accuracy of diagnostics from 76% to 94% was attained at data set with 92 images.

Zhang and others [4] suggested cascade classifiers approach. At the first cascade level the classifiers reject easy cases (those which evidently don't pass test) and the others are transferred to the second level which uses more complex classification system and so on. This method was applied to data base of Israel technological Institute consisting of 361 images and accuracy results was 97%. The most of last papers refers to field of breast cancer classification oriented on integer images [3–6]. But wide implementation of breast image classification (BIC) and other forms of digital pathology faces with such disturbances as high cost of implementation, insufficient productivity for huge amount of clinic procedures, interior technologic problems, and opposition from pathologist-anatomists side. Till now the most of works based on histology breast cancer analysis were performed on not large datasets. Some improvement presents data set with 7909 breast images obtained from 82 patients [7]. In this research the authors estimated various texture descriptors and various classifiers and carried out the experiments with accuracy from 82% to 85%.

Based on results presented in [7] one can make the conclusion that texture descriptors may propose good solution for images processing. But some researchers believe that main weakness of modern machine learning methods occurs just at this stage. This means that machine learning algorithms should be less dependent on functional engineering and be able to extract and organize discriminating information directly from images, in other words be capable to learn presentations.

The idea of learning presentations isn't new one but it became implementable only now with appearance GPU( Graphic Processing Units) which are capable to provide high speed performance ( productivity) with relatively small cost due to their parallel architecture [8].

The alternative to this approach is the application of CNN for medical images processing and diagnostics, which is considered and developed in the present research. It was shown that CNN is able to overcome the conventional texture descriptors [9, 10]. Besides traditional approach to detection of features based on descriptors demands much efforts and high level knowledge of experts and usually is specific for every task that prevents its direct application for another similar tasks.

Therefore in our research we suggested and developed hybrid CNN–FNN medical images classification system in which CNN is utilized to extract informative features of images and FNN NEFClass is applied for classification of detected breast tumors on images in two classes: benign and malicious ones.

The main goal of this work is development and investigation of algorithmic and software tools for fast analysis of breast tissue images, detection of tumors and their classification into two classes: benign or malignant one. This will enable to provide express analysis of images and raise the quality medical diagnostics.

## DATA SET DESCRIPTION

For our investigation we used data set BreKHis specially created for estimation efficiency of different approaches and tools for medical images of breast tumor diagnostics.

Data set BreKHis [7] contains microscope biopsies from benign and malign tumors of breast. The images were obtained in clinic research since January 2014 till December 2014.

BreKHis consists of 7909 clinically representative microscopic images of breast tumors received from 82 patients with different scale augmentation (40×, 100×, 200×, 400×).

All patients which during this period were investigated in R&D medical lab with clinical conclusion of breast cancer were invited to take part in this investigation. All data were anonymized. The patterns are generated of biopsy breast slides colored with hematoxulin and eosin (HE). The patterns are collected by surgery biopsy prepared for histologic research and marked by pathologist-anatomists of R&D lab. The main goal is to preserve original structure of tissue and molecular composition which allows to observe it with optical microscope. For investigation all images were split into slides of size 3 mkm. The final conclusion of each case was made by experienced pathologist-anatomist which was confirmed by additional investigation such as immune histology-chemistry (IHC).

The microscope system Olympus BX-50 with augmentation 3.3 connected with digital camera Samsung SCC-131AN, is used for obtaining digitized images of breast tissue. Images were obtained in 3-channels color space True color (24 bits value, 8 bits color channels RGB) with magnification coefficients 40×, 100×, 200×, and 400×. In the fig. 1–4 four images are presented with four magnification coefficients: (a) 40 ×, (b) 100 ×, (c) 200 ×, (d) 400 × — obtained from one slide of breast tumor which contains malignant tumor (breast cancer) Separated rectangular ( added by hand for illustrative aims) — region of interest (ROI) which was chosen by pathologist-anatomist will be described in the next section. In the fig. 5 the image of benign tumor is presented.

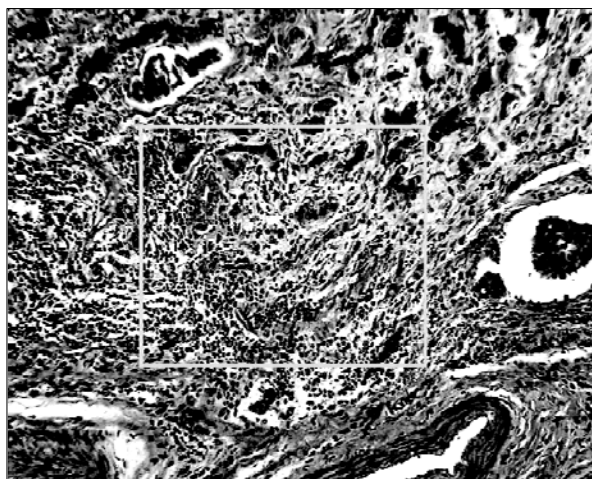
Up to date dataset BreakHis consists of 7909 images, divided into benign and malignant tumors. Table 1 presents the distribution of images [7].

**Table 1.** Distribution of images by magnification coefficients and class

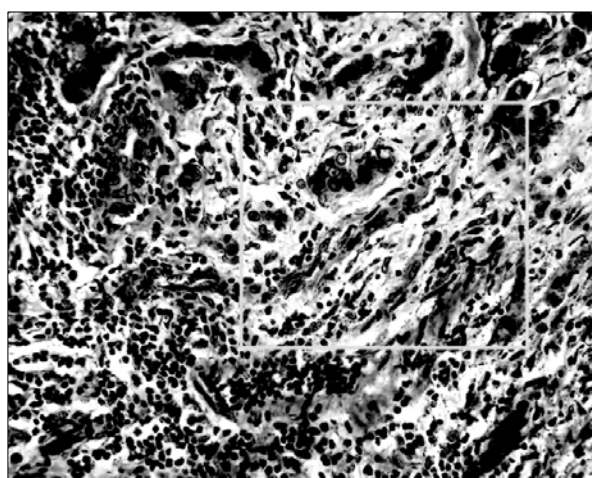
Magnification	Benign	Malignant	Total
40×	625	1370	1995
100×	644	1437	2081
200×	623	1390	2013
400×	588	1232	1820
Total	2480	5429	7909
Number of patients	24	58	82



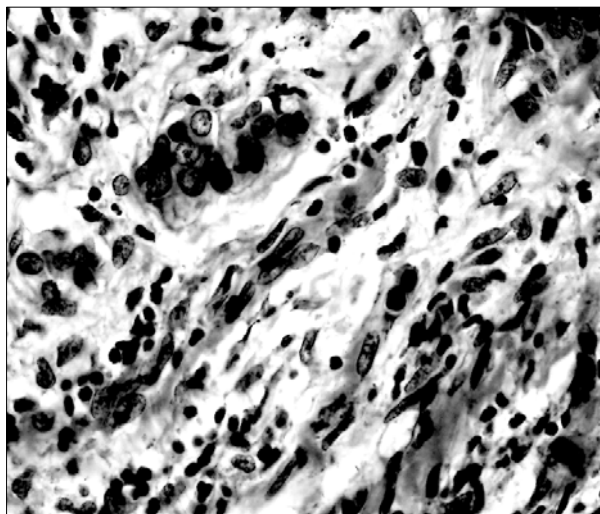
*Fig. 1.* Slide of malignant tumor with magnification 40×



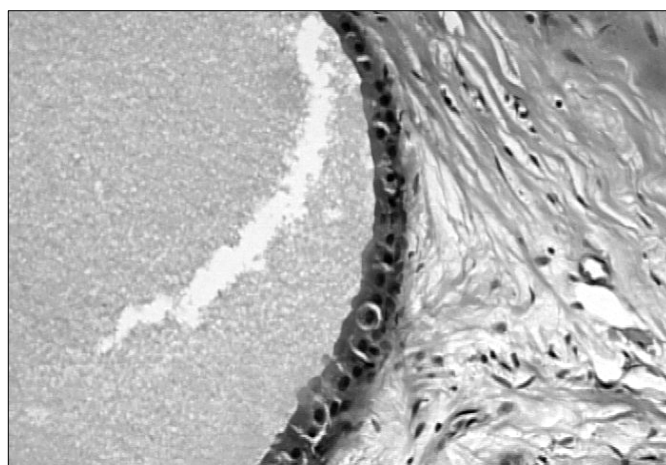
*Fig. 2.* Slide of malignant tumor with magnification 100×



*Fig. 3.* Slide of malignant tumor with magnification 200×



*Fig. 4. Slide of malignant tumor with magnification 400×*



*Fig. 5. Slide of benign tumor with magnification 100×*

#### **CONVOLUTIONAL NEURAL NETWORKS: BRIEF DESCRIPTION**

A CNN model is a state-of-the-art method that has been largely utilized for image processing. A CNN model has the ability to extract global features in a hierarchical manner that ensures local connectivity as well as the weight-sharing property. It consists of the following layers [9, 10].

**Convolutional Layer:** The Convolutional layer is considered as the main working ingredient in a CNN model and plays a vital determining part of this model. A kernel (filter), which is basically an  $n \times n$  matrix successively goes through all the pixels and extracts the information from them.

**Stride and Padding:** The number of pixels a kernel moves in a step is determined by the stride size; conventionally, the size of the stride is set to 1. Let we have an input data matrix of size  $5 \times 5$ , which is scanned with a  $3 \times 3$  kernel. When

we use a  $3 \times 3$  kernel, and stride size 1, then the convolved output is a  $3 \times 3$  matrix; however, when we use stride size 2, the convolved output is  $2 \times 2$ . Interestingly, if we use a  $5 \times 5$  kernel on the above input matrix with stride 1, the output will be a  $1 \times 1$  matrix. Thus, the size of the output image changes with both the size of the stride and the size of the kernel. To overcome this drawback, we can utilize extra rows and columns at the end of the matrices that contain 0 s. This adding of rows and columns that contain only zero values is known as *zero padding*.

**Nonlinear Performance:** Each layer of the NN produces linear output, and by definition adding two linear functions will also produce another linear output. Due to the linear nature of the output, adding more NN layers will show the same behavior as a single NN layer. To overcome this issue, a rectifier functions such as Rectified Linear Unit (ReLU), Leaky ReLU, Tanh, Sigmoid, etc., are introduced to make the output nonlinear.

**Pooling Operation:** A CNN model produces a large amount of feature information. To reduce the feature dimensionality, a down-sampling method named a pooling operation has been performed. A few pooling operation methods are well known such as [9,10]: Max Pooling, Average Pooling.

For our analysis, we have utilized the Max Pooling operation that selects the maximum values within a particular patch.

**Drop-Out:** Due to the overtraining of the model, it shows very poor performance on the test dataset, which is known as over-fitting. These over-fitting issues have been controlled by removing some of the neurons from the network, which is known as Drop-Out.

**Decision Layer:** For the classification decision, at the end of a CNN model, a decision layer (usually MLP) is introduced. Normally, a Softmax layer or SVM layer is introduced for this purpose. This layer contains a normalized exponential function and calculates the loss function for the data classification.

## CNN MODEL FOR IMAGECLASSIFICATION

In the next fig. 6 the architecture of VGG-16 is presented which was used in our work as detector of informative features. It was trained by different algorithms: stochastic gradient descent (SGD), differential evolution [14, 15] and basin hopping [11].

As classifier of obtained features in our research it was suggested to use FNN NEFClass. FNN NEFClass was firstly suggested by D. Nauck and W. Kruse in [12]. It was modified and developed in [13, 14] (so-called FNN NEFClass M) The learning algorithms for FNN NEFClass: stochastic gradient SG, conjugate gradient descent (CGS) and genetic algorithm were developed and investigated in [14] for the problem of optical images pattern recognition.

FNN NEFCLASS was successfully applied for analysis of medical images of cervix tissue obtained with use of colposcope and diagnostics [16]. The main advantages of FNN NEFClass as classifier are: possibility to work with incomplete and fuzzy input data; performing fuzzy classification of input patterns

(images) using so-called membership functions; speed and high accuracy [13, 14, 15].

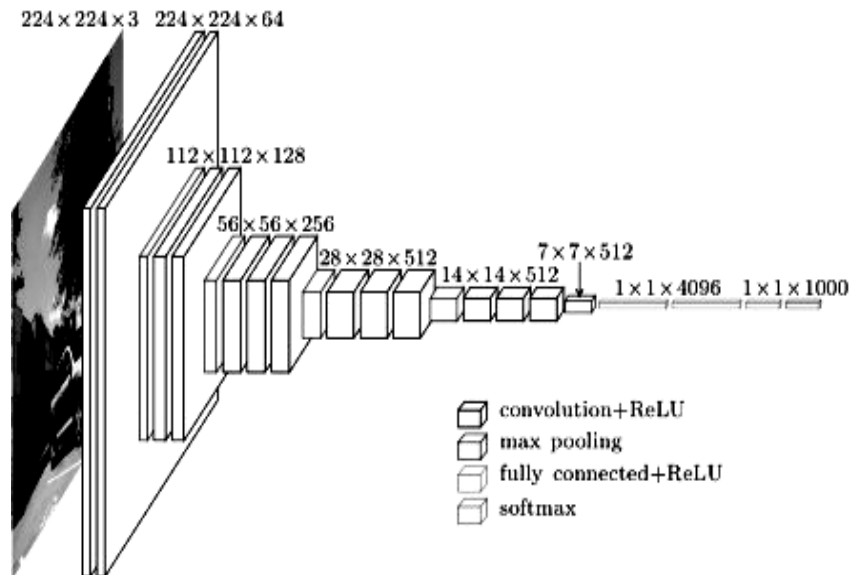


Fig. 6. Convolutional neural network VGG-16

## EXPERIMENTAL INVESTIGATIONS AND ANALYSIS

As it was already mentioned in our investigation pre-trained CNN VGG-16 was used. Method of training transfer was applied for this purpose. There are two main training scenarios:

**Features extraction.** In this case the last full-connected layer is deleted and the rest part of CNN is used as extractor for new data sets.

**Fine tuning.** In this case new data set is used for fine training of previously pre-trained neural network. In our research CNN VGG-16 was used for features extraction in medical images of breast tumors. After that the found features were fed as input data to FNN NEFClass. As algorithms of training FNN three algorithms were used: basin hopping [11], stochastic gradient descent and differential evolution [15].

## EXPERIMENTS DESCRIPTION

The series of experiments were carried out and the results were compared with works of predecessors. In the following tables 2, 3 the results of classification with different parameters are presented. All sample was divided into training and testing subsamples with ratio 80% / 20%.

In the first experiment we varied the number of linguistic variables (terms) and rules that to determine the best parameters values (table 2).

**Table 2.** Classification results of FNN NEFClass

Initial number of fuzzy sets (linguistic terms) / number of rules	40×, %	100×, %	200×, %	400×, %
2/2	73	74	74,2	73,5
4/2	75,3	74,8	75,7	75,4
6/2	78,2	79	78,4	78
8/2	76	75,4	76,5	75,8
2/4	75	74	73,8	73
4/4	78,3	76,3	75,7	75,4
6/4	82	83	82,4	83,2
8/4	82,2	81,5	81,5	83,8
2/6	75,4	73,8	74,4	73,2
<b>4/6</b>	<b>90</b>	<b>91</b>	<b>90,5</b>	<b>90</b>
6/6	89	89,7	90,2	89,5
8/6	90,3	90,5	92	91,2
4/8	89,3	89,8	89,7	89,3
6/8	89,2	88	89,4	88,4
8/8	88	87,2	87,2	87

From this table one can readily see that beginning from 6 fuzzy sets per variable and 6 rules the accuracy doesn't increase but complexity of training raises.

As it follows from the table for two classes the best values of parameters for FNN NEFClass are 4 fuzzy sets per variable and 6 rules. For comparison let's present the results of the previous work obtained with different classifiers for the same problem [6] (see table 3).

**Table 3.** Comparison of different classifiers accuracy

Classifier/magnification coefficient	40×, %	100×, %	200×, %	400×, %
Linear SVM	89	89	88	88
Polynomial SVM	88	90	89	85
Random forest	89,18	88	87,74	80
NEFClass	90	91	90,5	90

As we can see from the table 3 FNN NEFClass shows better results than previous classifiers: SVM machine and Random forest suggested in [6].

In our work for training of FNN NEFClass three algorithms were applied, namely, basin hopping, stochastic gradient descent and differential evolution. Using algorithms basin hopping and stochastic gradient descent we obtained approximately equal results that may mean to be close to optimal results while the training results of differential evolution appeared to be much worse.

It's worth to note that in this problem the number of features extracted by CNN VGG16 was very large — 4096 features. Therefore it was decided to cut the



number of features and reduce dimensionality of classification problem. For this aim principal components method (PCM) [17] was applied. In the table 4 the results of such reduction are presented.

**Table 4.** The dependence of total variance on number of components and approximate training time

Number of principal components	Variation	Approximate training time (in hours)
100	0,84058	~2
200	0,89736	~3
250	0,91232	~4
500	0,95486	~9

From the table 4 it follows that the results of reduction with 250 principal components are most acceptable as the complexity of training increases approximately proportional to dimension of input data. Due to lack of time the next experiments were performed using data with magnification factor 100× (2081 images). In the next table 5 the accuracy of classification is presented with various NEFClass parameters.

**Table 5.** Classification accuracy with 250 features

Number of fuzzy sets / number of rules	100×, %
4/4	80,64
4/6	87,24
4/8	88,18

In the table 6 the dependence of classification accuracy versus number of features is presented. One can see from this table that accuracy decreased only by some percent due such features reduction. But by this reduction we substantially have cut the training time.

**Table 6.** Classification accuracy with different number of features

Number of linguistic terms, number of rules/number of features	100	250	4096
4/4	75,23%	80,64%	86,3%
4/6	83,34%	87,24%	91%
4/8	84,21%	88,18%	89,8%

From this table one can easily see that the accuracy drops with decrease of features number but insignificant by 3–5% if compare results with 100 and 250 features. For comparison the classification with the full set of features 4096 was performed and we detected that with decrease features number in 20 times the accuracy falls only by 3–5%, in average. This conclusion confirms the efficiency of PCM method application for reduction of dimensionality of medical images classification problems.

## CONCLUSION

- The problem of analysis of breast tissue medical images and classification of detected tumor in two classes: benign and malignant is considered and discussed.
- For pattern recognition of breast tumors new hybrid CNN- FNN network is suggested in which CNN VGG 16 is used for informative features extraction while FNN NEFClass is used for classification of detected tumors.
- For training FNN NEFClass algorithms basin hopping, stochastic gradient descent and differential evolution were suggested and their efficiency investigated.
- The experimental investigations of suggested hybrid CNN-FNN network in the problem of classification real images of breast tumors using dataset BreakHis were carried out.
- The comparison of classification accuracy of the suggested hybrid CNN-FNN network with known work based on use of classification algorithms SVM and Random forest was performed which confirmed the efficiency of the suggested approach.
- The problem of reducing number of features in medical images classification problem using PCM method was investigated and its efficiency explored.

## REFERENCES

1. *Boyle P.* World Cancer Report 2012 / P. Boyle, B. Levin, Eds. — Lyon: IARC, 2012. — Available at: [http://www.iarc.fr/en/publications/pdfsonline/wcr/2008/wcr\\_2012.pdf](http://www.iarc.fr/en/publications/pdfsonline/wcr/2008/wcr_2012.pdf)
2. *Lakhani S.R.* WHO classification of tumours of the breast / S.R. Lakhani, S. Schnitt et al. — 4th ed. — Lyon: WHO Press, 2012.
3. *Zhang Y.* Breast cancer diagnosis from biopsy images with highly reliable random subspace classifier ensembles / Y. Zhang, B. Zhang, F. Coenen et al. // *Machine Vision and Applications*. — 2013. — Vol. 24, N. 7. — P. 1405–1420.
4. *Zhang Y.* One-class kernel subspace ensemble for medical image classification / Y. Zhang, B. Zhang, F. Coenen et al. // *EURASIP Journal on Advances in Signal Processing*. — 2014. — Vol. 2014, N 17. — P. 1–13.
5. *Doyle S.* Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features / S. Doyle, S. Agner, A. Madabhushi et al. // in *Proceedings of the 5th IEEE International Symposium on Biomedical Imaging (ISBI): From Nano to Macro*. — Vol. 61. — IEEE, May 2008. — P. 496–499.
6. *Singh Aditi.* Classifying Biological Images Using Pre-trained CNNs / Aditi Singh, Hadi Mansourifar, Hasnain Bilgrami et al. — Available at: <https://docs.google.com/document/d/1H7xVK7nwXcv11CYh7h15F6pM0m218FQloAXQODP-Hsg/edit?usp=sharing>
7. *Spanhol F.* A dataset for breast cancer histopathological image classification / F. Spanhol, L.S. Oliveira, C. Petitjean et al. // *IEEE Transactions of Biomedical Engineering*, 2016.
8. *Bengio Y.* Representation learning: A review and new perspectives / Y. Bengio, A. Courville, P. Vincent // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. — 2013. — Vol. 35. — P. 1798–1828.

9. *LeCun Y.* Deep learning / Y. LeCun, Y. Bengio, G. Hinton // *Nature*. — 2015. — Vol. 521. — P. 436–444.
10. *Krizhevsky A.* Imagenet classification with deep convolutional neural networks / A. Krizhevsky, I. Sutskever, G.E. Hinton // *Advances in Neural Information Processing Systems*. — 2012. — Vol. 25. — P.1097–1105.
11. *Olson B.* Basin Hopping as a General and Versatile Optimization Framework for the Characterization of Biological Macromolecules / B. Olson, I. Hashmi, K. Molloy et al. // *Advances in Artificial Intelligence*. —2012. — Article ID 674832.
12. *Nauck Detlef.* New learning strategies for NEFCLASS / Detlef Nauck, Rudolf Kruse // *In Proc. Seventh International Fuzzy Systems Association World Congress IFSA'97*. — Prague: Academia Prague, 1997. — Vol. IV. — P. 50–55.
13. *Zaychenko Yu.P.* Fuzzy neural networks for economic data classification / Yu.P. Zaychenko, Fatma Sevae, A.V. Matsak // *Vestnik of National Technical University of Ukraine "KPI", section "Informatic, control and computer engineering"*. — 2004. — Vol. 42. — P. 121–133.
14. *Zaychenko Yu.P.* The investigations of fuzzy neural networks in the problems of electro-optical images recognition / Yu.P. Zaychenko, I.M. Petrosyuk, M.S. Jaroshenko // *System research and information technologies*. — 2009. — N 4. — P. 61–76.
15. *Zgurovsky M.* The Fundamentals of Computational Intelligence: System Approach / M. Zgurovsky, Yu. Zaychenko // Switzerland: Springer International Publishing AG. — 2016. — 308 p.
16. *Zaychenko Yu.* Recognition of objects on Optical Images in Medical Diagnostics Using Fuzzy Neural Network NEFClass / Yu, Zaychenko, V. Huskova // *Intern. Journal Information Models and Analysis*. — 2015. — Vol. 4, N 1. — P. 13–22.
17. *Jindal N.* Enhanced Face Recognition Algorithm using PCA with Artificial Neural Networks / N. Jindal, V. Kumar // *International Journal of Advanced Research in Computer Science and Software Engineering*. — 2013. — Vol. 3. — P. 864–872.

*Received 27.08.2018*

---

From the Editorial Board: the article corresponds completely to submitted manuscript.