



ADAPTIVE HUMAN MACHINE INTERACTION APPROACH FOR FEATURE SELECTION-EXTRACTION TASK IN MEDICAL DATA MINING

Iryna Perova, Yevgeniy Bodyanskiy

Kharkiv National University of Radio Electronics, Kharkiv, 61166, Ukraine,
rikywenok@gmail.com, yevgeniy.bodyanskiy@nure.ua

Paper history:

Received 09 March 2018
Received in revised form 19 June 2018
Accepted 22 June 2018
Available online 30 June 2018

Keywords:

Human Machine Interaction;
feature selection;
feature extraction;
Medical Data Mining;
Oja's neuron;
Neural Network Approach.

Abstract: Feature Selection task is one of the most complicated and actual in the areas of Data Mining and Human Machine Interaction. Many approaches to its solving are based on non-mathematical and presentative hypothesis. New approach to evaluation of medical features information quantity, based on optimized combination of feature selection and feature extraction methods is proposed. This approach allows us to produce optimal reduced number of features with linguistic interpreting of each of them. Hybrid system of feature selection/extraction based on Neural Network-Physician interaction is investigated. This system is numerically simple, can produce feature selection/extraction with any number of factors in online mode using neural network-physician interaction based on Oja's neurons for online principal component analysis and calculating distance between first principal component and all input features. A series of experiments confirms efficiency of proposed approaches in Medical Data Mining area and allows physicians to have the most informative features without losing their linguistic interpreting.

Copyright © Research Institute for Intelligent Computer Systems, 2018.
All rights reserved.

1. INTRODUCTION

The medical diagnostics is the one of the areas where human machine interaction problems [1-3] became really important due to different reasons. Physician is a person who decides what diagnosis should be made to patient and only physician is held fully liable for it. But most of measurements in human organism are made using technical systems. Some of these systems contain sub-systems of decision making and others don't have them. A number of measured parameters in patients is often too large. It is easy to see, that a physician obtains diagnostic information from different sub-systems and must make diagnostics when input information contains a large number of features. To make diagnostics in a good and honest manner a physician needs to obtain information about the most informative features before making any diagnostics. In the area of Medical Data Mining this problem is known as feature selection.

Actually, feature extraction and feature selection approaches in medical diagnostics are the most

discussable and complicated problems. Medical data sets often contain a large number of features, which coincide with a little number of patients. This fact makes ineffective the most known approaches for correct medical diagnostics without compressing input features space.

Preprocessing of input data is one of necessary stages of whole Data Mining task. During preprocessing the tasks of data reduction, feature selection and feature extraction have to be realized [4-9]. Feature selection and feature extraction approaches are the most important stages as possibility and quality of medical diagnostics process essentially depends on choosing specific features and their quantity.

At present time clear mathematical grounding was obtained due to feature extraction approaches such as principal component analysis (PCA) [10-12], discriminant analysis, principal manifolds analysis (PMA) [13]. Neural network approach can be used for solving feature extraction problem that is an actual task in deep learning systems (autoencoder Bottle Neck, Restricted Boltzmann Machine, e.a.)

[14-15]. In medical diagnostics process it is very important to have possibility to interpret obtained results using input features. That's why feature selection approaches became essentially useful in Medical Data Mining tasks, but usually they are based on non-formal intuitive hypotheses [16]. So, in this paper the authors propose to formalize and optimize this process using combined feature selection-extraction procedure in online mode.

From physician's point of view, feature selection approach is required for choosing reduced number n^R of features from n features (where $n^R < n$) with minimal loss of information value to provide mathematically grounded diagnostic process.

All these described earlier need to be presented from mathematical point of view to create mathematical system for feature selection, based on original dataset. So, feature selection approach is a process of choosing a map $x^R = f(x)$, where every pattern $x(k) = ((x_1(k), \dots, x_n(k))^T$, ($k = 1, 2, \dots, N$ – number of feature vector in dataset) belongs to n – dimensional feature space R^n , and transformed vector x^R – to n^R – dimensional space R^{n^R} , in conditions when $n^R < n$. Reduced feature space must contain the most informative features from input feature space R^n that permits to avoid the so-called curse of dimensionality. Thus, if n in our experiments is 34 (number of input features is 34 in dermatology dataset) than $n^R = 3$ (the dimensionality of input vectors for PCA-visualization). So, a symbol n^R describes the dimensionality of reduced space and symbol R^n denotes the n – dimensional real space of original input variables.

Therefore, the principal goal of such transformation is reducing of input feature space with saving of optimal data specifications to make medical diagnostic process clear and correct. Applying of hybrid systems of Computational Intelligence for different tasks in Medical Data Mining area is an actual problem, which can increase quality of medical diagnostics, clusterization, pattern recognition and so on. This situation occurs due to the ability of hybridization process to integrate advantages of different systems for achieving a goal. Applying of hybrid system for searching the most informative features in different medical applications is an actual task, especially in medical diagnostics to exclude any situations of information shortage (data incompleteness, data fuzziness, presence of gaps and outliers in data).

In this paper authors propose to integrate advantages of feature extraction and feature

selection systems and create a united hybrid feature extraction-selection system for estimation of features information content with extraction of the most informative features without losing its physical sense (possibility of linguistic interpretation).

2. HYBRID SYSTEM FOR FEATURE SELECTION-EXTRACTION BASED ON NEURAL NETWORK-PHYSICIAN INTERACTION

New approach based on using neural network with Oja's neurons integrated with physician is proposed. In the frame of this approach the most informative feature is the one, which has minimal distance with output signal of Oja's neuron in sense of Manhattan metrics. It's important to mark that proposed system can implement feature extraction-selection process in online mode.

At the first step of the proposed approach all measured features of a patient are fed to processing sequentially. If any of the features contains gaps, they will have to be filled by any procedure, described, for example, in [17].

Previously, input data have to be centered and normalized using expression

$$x_norm_i(k) = x_i(k) - \bar{x}_i \quad (1)$$

(where \bar{x}_i – arithmetic mean or median calculated for each feature), and coded to interval $[0;1]^n$ or $[-1;1]^n$ using expression (2) or (3)

$$\tilde{x}(k) = \frac{x_norm_i(k) - x_{i\min}}{x_{i\max} - x_{i\min}}, \quad (2)$$

$$0 \leq \tilde{x}(k) \leq 1,$$

$$\tilde{x}_i(k) = \frac{2 \cdot x_norm_i(k) - x_{i\max} - x_{i\min}}{x_{i\max} - x_{i\min}}, \quad (3)$$

$$-1 \leq \tilde{x}(k) \leq 1.$$

Encoded feature vector is fed to processing on group of Oja's neurons, where in output of the first Oja's neuron the first principal component $y_1(k)$ is obtained. Oja's neuron was proposed by Erkki Oja, professor from Aalto University and helps to implement extraction of the first principal component from features in dataset in online mode.

Self-learning algorithm for adaptive linear associator (ALA) based on normalized Hebb rule for computing of the first principal component, that

corresponds to eigen vector w_1 of Oja's neuron [18, 19] looks like

$$y_1(k) = w_1^T(k) \tilde{x}(k). \quad (4)$$

This algorithm minimizes local (one-step) learning criterion (energy function, Lyapunov function)

$$E(k) = \frac{1}{2} \|\tilde{x}(k) - \hat{x}(k)\|^2 \quad (5)$$

where $w_1(k)$ – current estimate of the synaptic weights of ALA, that corresponds to eigen vector w_1 ;

$y_1(k)$ – current estimate of the first principal component;

$\hat{x}(k) = w(k) \hat{y}_1(k)$ – recovered (“decompressed”) estimate of input normalized signal $\tilde{x}(k)$.

Considering the fact that gradient of criterion (5) looks like

$$\begin{aligned} \nabla_w E(k) &= -(\tilde{x}(k) - \hat{x}(k)) y_1(k) = \\ &= -(\tilde{x}(k) - w(k) y_1(k)) y_1(k), \end{aligned}$$

and Oja's rule for the first eigen vector has the following form:

$$\begin{aligned} w_1(k+1) &= w_1(k) + \\ &+ \eta(k) y_1(k) (\tilde{x}(k) - w_1(k) y_1(k)), \end{aligned} \quad (6)$$

we can write the learning procedure as follows:

$$\begin{aligned} w_1(k+1) &= w_1(k) + \\ &+ \eta(k) (\tilde{x}(k) \tilde{x}^T(k) w_1(k) - \\ &- w_1^T(k) \tilde{x}(k) \tilde{x}^T(k) w_1(k) w_1(k)) = \\ &= w_1(k) + \eta(k) (\tilde{x}(k) \tilde{x}^T(k) - \\ &- w_1^T(k) \tilde{x}(k) \tilde{x}^T(k) w_1(k) I_m) w_1(k), \end{aligned} \quad (7)$$

where $\eta(k)$ – learning rate parameter, that corresponds to Dvoretzky conditions.

Expressions (6-7) are the procedures of stochastic approximation type with low speed of converging

that cannot be used in non-stationary conditions. That is why, we can introduce the modified Oja's rule in the following form:

$$\begin{cases} w_1(k+1) = w_1(k) + r^{-1}(k) \cdot \\ \cdot y_1(k) (\tilde{x}(k) - w_1(k) y_1(k)). \\ r(k) = \alpha r(k-1) + y_1^2(k), \quad 0 \leq \alpha \leq 1, \end{cases} \quad (8)$$

It has the better operation speed and additional smoothing properties, which can be controlled using forgetting parameter α . In Fig.1 the modified Oja's neuron architecture is presented. This architecture solves a task of principal component analysis with sequentially obtaining of principal components in online mode as opposed to classical PCA that provides obtaining of principal components in batch mode. It is simple to show that on fix number of features classical PCA and modified Oja's neuron architecture provides an obtaining of equal eigen vectors. For example, after testing of proposed system on Iris Fisher dataset [20] equal eigen vectors were obtained using classical PCA and modified Oja's neuron architecture. Time for calculating eigen vector of Iris Fisher dataset by classical PCA using MacBook 12' with Intel Core m3 processor (1,1 GHz), 8Mb of RAM (1867 MHz) was 0,93 sec. Time for calculating eigen vector of Iris Fisher dataset by modified Oja's neuron architecture in the same condition was provided by 1105 iterations (2,57 sec). It took more processing time to modify Oja's neuron but only using of this system can provide paralleling work of Physician (Human) and Machine (Neural Network) to make Human-Machine (Neural Network-Physician) Interaction come true. All patients come to physician sequentially, physician processes features sequentially, so computer system needs to be working in online mode to provide online data processing.

At the next step, distances in the sense of Manhattan metrics between all features vectors and output signal of first Oja's neuron are calculated. A feature that has a minimal distance

$$d(\tilde{x}(z), y_1) = \sum_{i=1}^N |\tilde{x}_i(z) - y_1| \quad (9)$$

is chosen like the most informative one.

Then, this feature-winner is excluded from original data massive and system continues to process reduced dataset until all features will be turned over.

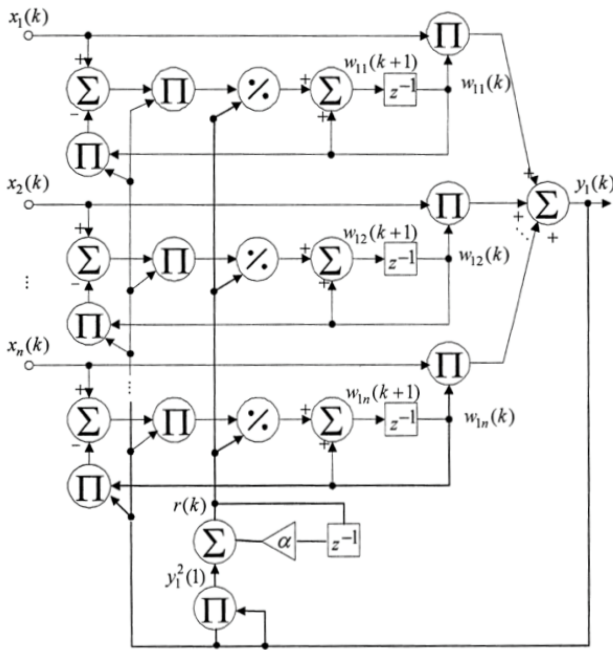


Figure 1 – Architecture of modified Oja’s neuron

Hybrid system of features selection-extraction based on Neural Network-Physician interaction is presented in Fig. 2. Input vector $x(k) = ((x_1(k), \dots, x_n(k))^T$ is fed to block of normalization and centering of original features.

At output of this block vector of transformed features $\tilde{x}(k) = ((\tilde{x}_1(k), \dots, \tilde{x}_n(k))^T$ appears. Then, the first principal component of input features $y_1(\tilde{x})$ is calculated in Oja’s neuron block using expressions (4), (8). In block of feature-winner detection, the process of defining a feature that has minimal distance in sense of Manhattan metrics with $y_1(\tilde{x})$ is implemented using (9). At the final stage in reduction block this feature-winner is excluded from original dataset and the system continues to look for the next feature with minimal distance (next most informative feature).

As a result of feature selection-extraction hybrid systems work, reduced dataset $x^R(k) = ((x_1^R(k), \dots, x_{n^R}^R(k))^T \in R^{n^R}$ is formed from original dataset $x(k) = ((x_1(k), \dots, x_n(k))^T \in R^n$, $n^R < n$. The reduced dataset contains the most informative features without losing their physical sense. This reduced dataset is fed to processing to physician and can help him make the final diagnosis.

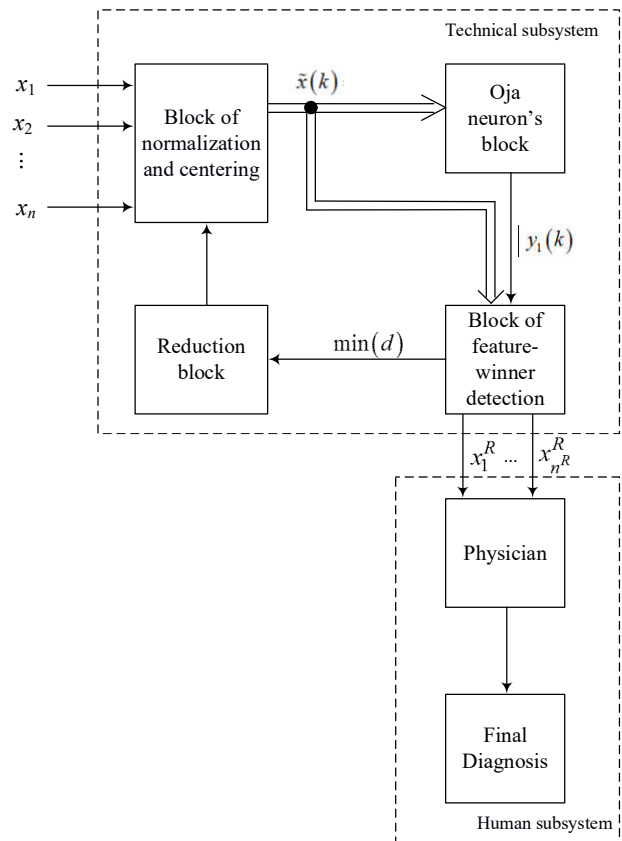


Figure 2 – Hybrid system of feature selection-extraction based on Neural Network-Physician interaction

3. EXPERIMENT

Hybrid system of feature selection-extraction based on Neural Network-Physician interaction was used for finding the most informative features in medical datasets from UCI Repository: dermatology.data (contains 34 features) [21], breast-cancer.data (contains 9 features) [22], pima-indian-diabetes.data (contains 8 features) [23], parkinsons.data (contains 21 features) [24]. All of these datasets contain classes-diagnoses that are overlapped in feature space (Fig.3-Fig.6) where classes-diagnoses were marked by different dots. In all figures dataset information was presented on the space of three first principal components. This approach is applied only to dataset visualization because all axes have no physical interpretation. Figures (see Fig. 3-Fig. 6) represent the most informative dividing of input information.

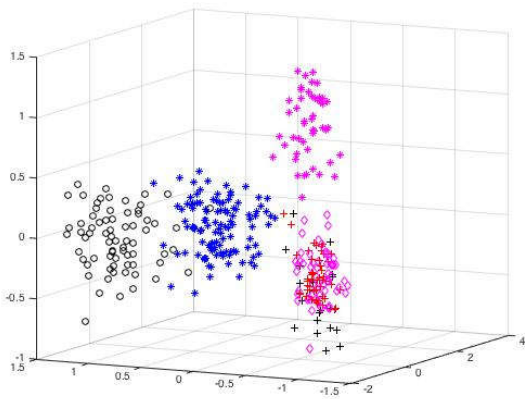


Figure 3 – Visualisation of Dermatology dataset

In each dataset procedure of feature selection-extraction was made. Final results are presented in Table 1.

In dermatology.data the most informative features are: elongation of the rete ridges, exocytosis, inflammatory mononuclear infiltrate, clubbing of the rete ridges, saw-tooth appearance of retes, scaling etc. In breast-cancer.data the most informative features are: Bare Nuclei, Uniformity Cell Size, Uniformity Cell Shape, Clump Thickness, etc.

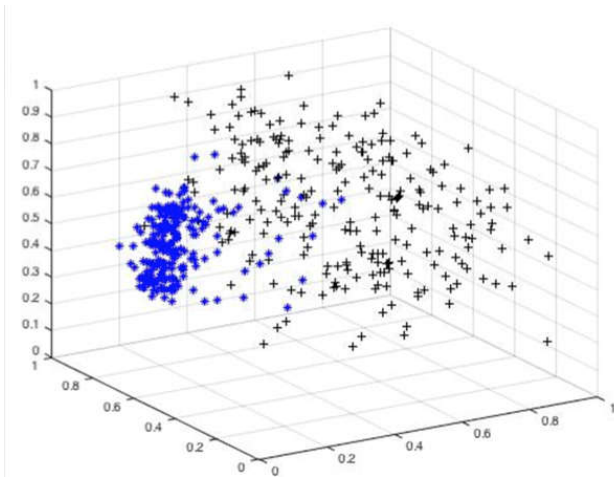


Figure 4 – Visualisation of Breast Cancer dataset

In pima-indians-diabetes.data the most informative features are: BMI, Diastolic, Glucose, NbPregnant, Age, etc. In parkinsons.data the most informative features are: Signal fractal scaling exponent, nonlinear dynamical complexity measure (RPDE), nonlinear measure of fundamental frequency variation (spread1), ratio of noise to tonal components in the voice (HNR), variation in amplitude (Shimmer_DDA).

Table 1. List of features medical datasets, starting of the most informative

№	Dataset	List of features, starting of most informative
1	dermatology.data	21, 16, 33, 20, 29, 2, 19, 4, 28, 27, 22, 6, 9, 12, 10, 34, 5, 32, 17, 25, 3, 14, 26, 1, 18, 11, 8, 15, 24, 7, 23, 31, 13, 30
2	breast-cancer.data	6, 2, 3, 1, 7, 5, 8, 4, 9
3	pima-indians-diabetes.data	6, 3, 2, 1, 8, 4, 7, 5
4	parkinsons.data	19, 17, 20, 16, 14, 21, 11, 18, 9, 12, 10, 5, 4, 1, 2, 7

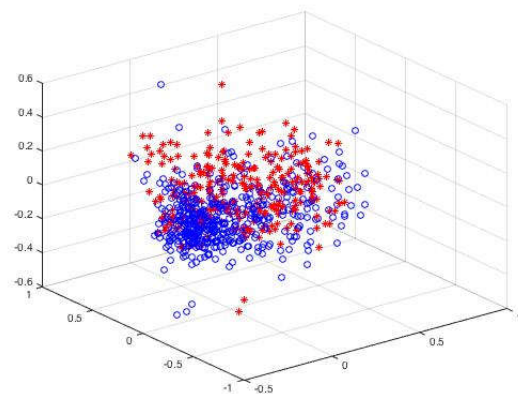


Figure 5 – Visualisation of Pima-Indian diabetes dataset

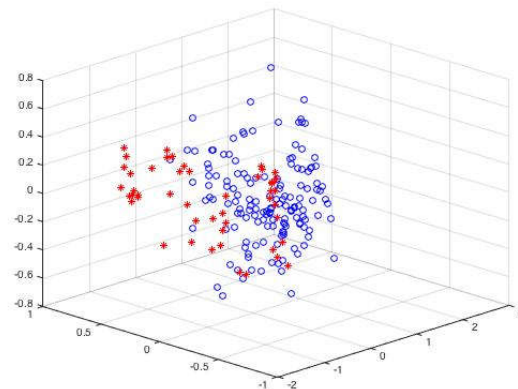


Figure 6 – Visualisation of Parkinsons dataset

Reduced feature space $x^R(k) = ((x_1^R(k), \dots, x_{n^R}^R(k))^T \in R^{n^R}$ can be processed by systems for medical fuzzy diagnostics, for example presented in [25-28].

4. CONCLUSION

In this paper a hybrid system of feature selection-extraction based on Neural Network-Physician interaction is proposed. This system enables to extract the most informative features without losing the physical sense of reduced feature space in online-mode, and it can simplify the Human Machine Interaction in the area of Medical Data Mining.

5. REFERENCES

- [1] A. Mosenia; S. Sur-Kolay; A. Raghunathan; N.K. Jha, "Wearable medical sensor-based system design: a survey," *IEEE Trans. on Multi-Scale Computing Systems*, vol. 3, issue 2, pp. 124-138, 2017.
- [2] S.P. Panda, "Automated speech recognition system in advancement of human-computer interaction", *2017 International Conference on Computing Methodologies and Communication (ICCMC)*, 2017, pp. 302-306, DOI: 10.1109/ICCMC.2017.8282696.
- [3] G. M. Phade, Prerna D. Uddharwar, P. A. Dhulekar, S. T. Gandhe, "Motion estimation for human-machine interaction," *2014 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, 2014, pp. 149-154, DOI: 10.1109/ISSPIT.2014.7300579.
- [4] J.C. Bezdek, P. Castelaz, "Prototype classification and feature selection with fuzzy sets," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 7, no. 2, pp. 87-92, 1977.
- [5] S.K. Pal, "Fuzzy set theoretic measures for automatic feature evaluation: II," *Information Sciences*, vol. 64, 1992, 165-179.
- [6] S.K. Pal, B. Chakraborty, "Fuzzy set theoretic measures for automatic feature evaluation," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 16, no. 5, pp. 754-760, 1986.
- [7] K.L. Priddy, S.K. Rogers, D.W. Ruck, G.L. Tarr, M. Kabrisky, "Bayesian selection of important features or feedforward neural networks," *Neurocomputing*, no. 5, pp. 91-103, 1993.
- [8] J.M. Steppe, K.W. Bauer Jr., "Improved feature screening in feedforward neural networks," *Neurocomputing*, vol. 13, pp. 47-58, 1996.
- [9] R.K. De, N.R. Pal, S.K. Pal, "Feature analysis: neural network and fuzzy set theoretic approaches," *Pattern Recognition*, vol. 30,
- [10] M. Pregenzer, G. Pfurtscheller, D. Flotzinger, "Automated feature selection with a distinctive sensitive learning vector quantizer," *Neurocomputing*, vol. 11, pp. 19-29, 1996.
- [11] C. R. Rao, "The use and interpretation of principal component analysis in applied research," *Sankhyā: The Indian Journal of Statistics, Series A*, vol. 26, no. 4, pp. 329-358, 1964.
- [12] M. Okamoto, "Optimality principal components multivariate analysis," *Proceedings of the 3rd Int. Symp.*, Dayton, 1967.
- [13] M. Okamoto, M. Kanazawa, "Minimization of eigenvalues of a matrix and optimality of principal components," *Ann. Math. Statist.*, vol. 39, no. 3, pp. 1-20, 1968.
- [14] C.M. Bishop, *Neural Networks for Pattern Recognition*, Oxford: Clarendon Press, 1995, 482 p.
- [15] A. Cichocki, R. Unbehauen, *Neural Networks for Optimization and Signal Processing*, Stuttgart: Teubner, 1993, 526 p.
- [16] J. Han, M. Kamber, *Data Mining: Concepts and Techniques*, Amsterdam: Morgan Kaufman Publ., 2006, 743 p.
- [17] P. Mulesa, I. Perova, "Fuzzy spacial extrapolation method using Manhattan metrics for tasks of medical data mining," *Proceedings of the International Conference on Computer Science and Information Technologies CSIT'2015*, Lviv, Ukraine, 2015, pp. 104-106. DOI: 10.1109/STC-CSIT.2015.7325443.
- [18] E. Oja, "A simplified neuron model as a principal component analyzer," *J. of Math. Biology*, vol. 15, pp. 267-273, 1982.
- [19] E. Oja, "Principal component, minor components, and liner neural networks," *Neural Networks*, vol. 5, pp. 927-935, 1992.
- [20] Iris Fisher dataset. [Online]. Available from: <http://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data>. Accessed: 1st of May, 2008.
- [21] Dermatology dataset. [Online]. Available from: <http://archive.ics.uci.edu/ml/machine-learning-databases/dermatology/dermatology.data>. Accessed: 1st of May, 2008.
- [22] Breast Cancer in Wisconsin dataset. [Online]. Available from: <http://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/breast-cancer-wisconsin.data>
- [23] Pima Indians Diabetes dataset. [Online]. Available from: <http://archive.ics.uci.edu/ml/machine-learning-databases/pima-indians-diabetes/pima-indians-diabetes.data>. Accessed: 1st of May, 2008.
- [24] Parkinson dataset. [Online]. Available from: <http://archive.ics.uci.edu/ml/machine-learning-databases/parkinsons/parkinsons.data>. Accessed: 1st of May, 2008.
- [25] I. Perova, Ye. Bodyanskiy, "Fast medical diagnostics using autoassociative neuro-fuzzy

memory,” *International Journal of Computing*, vol. 16, issue 1, pp. 34-40, 2017.

- [26] I. Perova, I. Pliss, “Deep hybrid system of computational intelligence with architecture adaptation for medical fuzzy diagnostics,” *I.J. Intelligent System and Applications*, vol. 7, pp. 12-21, 2017. DOI: 10.5815/ijisa.2017.07.02.
- [27] I. Pliss, I. Perova, “Diagnostic neuro-fuzzy system and its learning in medical data mining tasks in conditions of uncertainty about numbers of attributes and diagnoses,” *Automatic Control and Computer Sciences*, vol. 51, issue 6, pp. 391-398, 2017. DOI: 10.3103/S0146411617060062.
- [28] Ye. Bodyanskiy, I. Perova, O. Vynokurova, I. Izonin, “Adaptive wavelet diagnostic neuro-fuzzy system for biomedical tasks,” *Proceedings of the 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)*, Lviv-Slavske, Ukraine, February 20–24, 2018, pp. 299-303.



Iryna Perova graduated from Kharkiv National University of Radio Electronics in 2004. She got her PhD in 2008. She obtained an academic title of the Senior Researcher in 2015. She obtained an academic title of the Associate Professor in 2016.

Ph.D. Perova has been the associate professor in Biomedical engineering department at Kharkiv National University of Radio Electronics. She has more than 40 scientific publications including one invention. Her research interests are medical data mining, systems of computational intelligence,

neuro- and neo-fuzzy-systems for medical diagnostics tasks, on-line systems that have to do with control, identification, clustering, diagnostics and fault detection.



Yevgeniy Bodyanskiy in 1971 graduated with honour from Kharkiv National University of Radio Electronics. In 1980 he defended the Doctoral Thesis. In 1984 he was awarded the academic title of Senior Researcher. In 1990 he was awarded Dr. habil. sc. ing. degree. In 1994 he was awarded the academic title of Professor. His major fields of research are evolving hybrid systems of computational intelligence, data stream mining, data science, and big data.

Since 1974 he has been working at Kharkiv National University of Radio Electronics. In 1974–1976 he was a Researcher; in 1977–1983 he was a Senior Researcher; in 1986–1991 he was a Scientific Head of Control Systems Research Laboratory; in 1991–1992 he was a Research Fellow. Since 1992 he is a Professor of Artificial Intelligence Department at KhNURE, Scientific Head of Control Systems Research Laboratory at KhNURE. He has more than 660 scientific publications, including 42 inventions and 16 monographs. Research interests include hybrid systems of computational intelligence: adaptive, neuro-, wavelet-, neo-fuzzy-, real-time systems, including problems connected with control, identification, forecasting, clustering, diagnostics, fault detection in technical, economic, medical and ecological objects.

He is the IEEE senior member, member of 4 scientific and 7 editorial boards.