

Fast Probabilistic Neuro-Fuzzy System for Pattern Classification Task

Yevgeniy Bodyanskiy¹, Anastasiia Deineko², Irina Pliss³, Olha Chala⁴,

^{1,2,4} *Artificial Intelligence Department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine*

³ *Control Systems Research Laboratory, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine*

Abstract – The probabilistic neuro-fuzzy system to solve the image classification-recognition task is proposed. The considered system is a “hybrid” of Specht’s probabilistic neural network and the neuro-fuzzy system of Takagi-Sugeno-Kang. It is designed to solve tasks in case of overlapping classes. Also, it is supposed that the initial data that are fed on the input of the system can be represented in numerical, rank, and nominal (binary) scales. The tuning of the network is implemented with the modified procedure of lazy learning based on the concept “neurons at data points”. Such a learning approach allows substantially reducing the consumption of time and does not require large amounts of training dataset. The proposed system is easy in computational implementation and characterised by a high classification speed, as well as allows processing information both in batch and online mode.

Keywords – Lazy learning; membership function; neural network; neuro-fuzzy system; probabilistic pattern recognition.

I. INTRODUCTION

Today artificial neural networks have become widespread for an image recognition task due to their universal approximating properties. These characteristics allow restoring separating hypersurfaces of any complex form. Here, it should be noted that deep neural networks (DNNs) have recently become the most common apparatus for the solving of the classification-recognition task. In the meantime, DNNs are characterised by a low learning speed and require significant training dataset sizes. These reasons limit their usage in tasks when the data processing speed under conditions of a priori information deficit comes to the fore. In this case, probabilistic neural networks (PNNs) proposed by D. Specht can be the most effective ones [1], [2]. They are based on ideas of Bayesian inference and Parzen estimates of the data distribution density in classes of observations. Even though the network accuracy is somewhat inferior to deep ones, its speed and implementation simplicity, in some cases, becomes an advantage. Moreover, some modifications were introduced in [3]–[11] that provided the ability to process non-stationary data flows under overlapping class conditions based on the ideas of fuzzy classification [12]. Meanwhile, there is some bulkiness of these networks because a standard PNN can be a bit tricky because its first hidden pattern layer in the general case contains a number of R-neurons, which is equal to the number of observations in the training dataset.

To overcome this inconvenience, we can use the approach based on the hybrid systems of computational intelligence [13]–[15], and primarily neuro-fuzzy systems [16], [17]. The last ones have many advantages over classic neural networks, keeping multipurpose approximating properties combined with online mode learning abilities.

II. ARCHITECTURE OF PROBABILISTIC NEURO-FUZZY SYSTEM

The architecture of the probabilistic neuro-fuzzy system that has four layers of data processing is presented in Fig. 1. On the zero (receptor) layer of the system, vectors of observations are sequentially fed. They are forming a training dataset in the form $X = \{x(1), x(2), \dots, x(k), \dots, x(N)\}$,

$$x(k) = (x_1(k), x_2(k), \dots, x_i(k), \dots, x_n(k))^T \in R^n$$

(total N observations, meanwhile this number can be changed over time), also it is assumed that N_1 observations belong to the class Cl_1 , N_2 to Cl_2 , N_j to Cl_j , and finally N_m to the m -th class Cl_m . Herewith input data can be organised so that:

$$\left\{ \begin{array}{l} \text{if } x(k) \in Cl_1 \text{ then } k = \tau_1 = 1, 2, \dots, N_1, \\ \text{if } x(k) \in Cl_2 \text{ then } k = \tau_2 = N_1 + 1, \dots, N_1 + N_2, \\ \vdots \\ \text{if } x(k) \in Cl_j \text{ then } k = \tau_j = N_1 + N_2 + \dots + N_{j-1} + 1, \dots, \\ N_1 + N_2 + \dots + N_j, \\ \vdots \\ \text{if } x(k) \in Cl_m \text{ then } k = \tau_m = N_1 + N_2 + \dots + N_{m-1} + 1, \dots, \\ N_1 + N_2 + \dots + N_m = N. \end{array} \right. \quad (1)$$

These observations are fed to the first hidden fuzzification layer formed with fuzzy membership functions $\varphi_i(x_i, c_i)$, $i = 1, 2, \dots, h_i$, $i = 1, 2, \dots, n$ where the usual one-dimensional Gaussian functions are used:

$$\varphi_i(x_i, c_i) = \left(\frac{(x_i - c_i)^2}{2\sigma^2} \right), \quad (2)$$

where c_i – centre of corresponding membership function;
 σ – widespread parameter.

The number of these membership functions at each input x_i can be different: from two in the case of a binary input variable

that takes only two values “yes” or “no” to N as in a standard probabilistic neural network (PNN), i.e.,

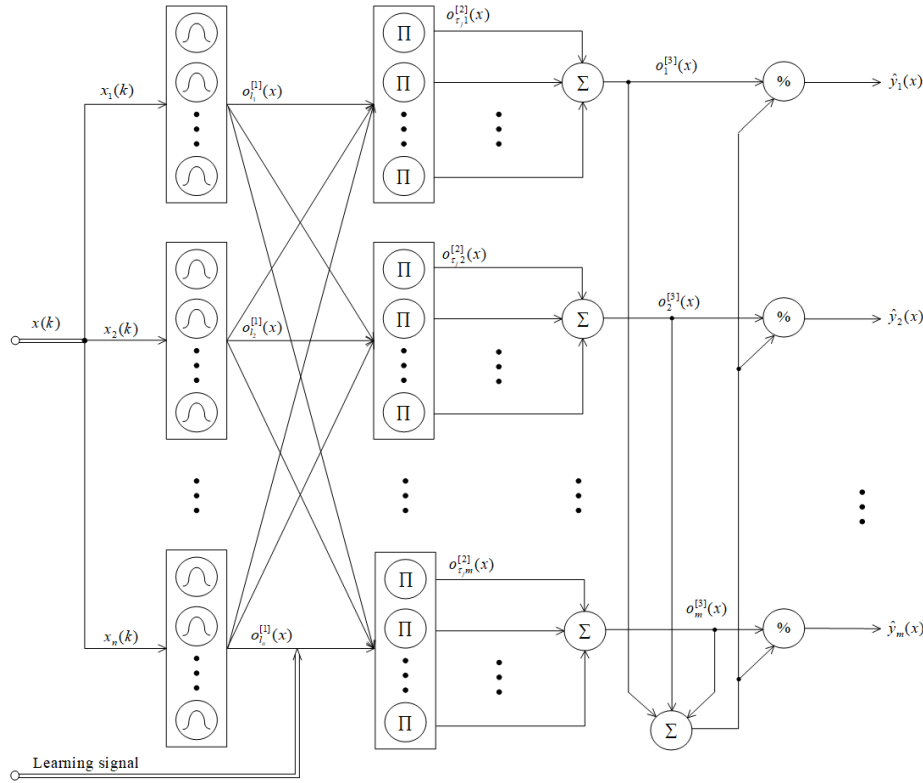


Fig. 1. Architecture of probabilistic neuro-fuzzy system.

$$2 \leq h_i \leq N \quad (3)$$

for each input or for the system in general

$$2n \leq h = \sum_{i=1}^n h_i \leq N_n. \quad (4)$$

The output signals of the first layer

$$o_i^{[1]} = \varphi_i(x_i, c_i) \quad (5)$$

are fed to inputs on the second hidden aggregation layer forming N standard multiplication blocks that form bell shaped activation functions in multidirectional Gaussian form

$$\varphi_{\tau_j}^j(x_i, c_{\tau_j}) = \prod_{i=1}^n \exp\left(-\frac{(x_i - c_i)^2}{2\sigma^2}\right) = \exp\left(-\frac{\|x - c_{\tau_j}\|^2}{2\sigma^2}\right), \quad (6)$$

where vector-centre $c_{\tau_j} = (c_{i_1}, \dots, c_{i_n})^T$ is formed based on one-dimensional centres for each class j . Herewith the number of multiplication blocks for each class is N_j . Thus, in the second layer, the N multidimensional kernel activation functions are formed, and on its output, the signals appear

$$o_{\tau_j}^{[2]} = \varphi_{\tau_j}^j(x(\tau_j), c_{\tau_j}). \quad (7)$$

It is easy to see that the first two layers of the proposed system coincide with input layers of the popular neuro-fuzzy system of Takagi-Sugeno-Kang or ANFIS [16], [17] and implement the same function as the PNN pattern layer. The advantages of the system popped up in case when some components of input vectors–images of the training dataset coincide. This situation appears when the input signal individual components are either binary or rank, or nominal variables, which is common in real tasks. Here, the number of membership functions in the corresponding input h_i is smaller than the amount of training data N .

The third hidden layer is formed by $m + 1$ adders, herewith from the first m Parzen estimates of the data distribution density in each class are formed

$$o_j^{[3]} = \sum_{\tau_j=N_1+N_2+\dots+N_{j-1}}^{N_1+N_2+\dots+N_j} o_{\tau_j}^{[2]} = p_j(x) \quad (8)$$

and $(m + 1)$ -th adds all output signals of this layer

$$o^{[3]} = \sum_{j=1}^m o_j^{[3]} = \sum_{j=1}^m \sum_{\tau_j=N_1+N_2+\dots+N_{j-1}}^{N_1+N_2+\dots+N_j} o_{\tau_j}^{[2]}. \quad (9)$$

Finally, the output layer of the system computes the probability of membership of each observation x that is not from the training dataset to each class

$$\hat{y}_j(x) = \frac{o_j^{[3]}(x)}{o^{[3]}(x)} = \frac{o_j^{[3]}(x)}{\sum_{j=1}^m o_j^{[3]}(x)}. \quad (10)$$

It is easy to see that the output layer performs the operation similar to the defuzzification in popular neuro-fuzzy systems. Still, in our case, it has purely probabilistic meaning.

III. FAST LEARNING OF PROBABILISTIC NEURO-FUZZY SYSTEM

It is proposed to use a modified lazy learning procedure based on the concept “neurons at data points” [18], [19] for the tuning of the PNFS. Exactly based on this conception, the classic PNNs learn; at the same time when the image from the training dataset $x(\tau_j)$ is fed to the input of the system, the multidimensional activation function $\varphi_{\tau_j}^j(x, c_{\tau_j})$ is formed, and the current observation determines its centre $c_{\tau_j} = x(\tau_j)$. This means that learning is almost instantaneous, but the number of activation functions is determined by the number of the training dataset N .

Using a neuro-fuzzy approach allows narrowing down the number of membership functions, based on which multidimensional activation function is formed in the second hidden layer. In the simplest case, the components of the observation $x(\tau_j)$ form N membership functions c_{τ_j} where individual coordinates determine centres $x(\tau_j)$, and we do not get any gain in the number of functions. Although with significant N value, these one-dimensional centres c_{τ_j} can be placed very close, and almost coincide on the axis x_i . Note also that within a neuro-fuzzy approach, all input signal coordinates are pre-processed and coded

$$0 \leq x_i(\tau_j) \leq 1. \quad (11)$$

The gain in the number of functions appear when within two different training observations $x(\tau_j)$ and $x_i(\delta_j)(\tau_j \neq \delta_j)$ some coordinates coincide $x_i(\tau_j) = x_i(\delta_j)$.

Such a situation is common when, for instance, one of the coordinates is a binary value of 0 or 1. Then it is necessary to form only two membership functions with centres in 0 and 1. The profit in the number of functions appears with nominal and rank variables as well. In this case, it is needed to create membership functions where their number is equal to the number of names or ranks.

To reduce the number of membership functions, by specifying the maximum possible number h_i for each coordinate x_i the distance between two neighbouring functions is defined by the value

$$\Delta_i = \frac{1}{h_i - 1}. \quad (12)$$

Clearly that for the binary variables $\Delta_i = 1$, in this case, centres are placed in points 0 and 1. Centres are placed in points $0, \Delta, 2\Delta, \dots, 1$ for the continuous numeric variables. If any value from the training dataset $x_i(\tau_j)$ appears to be in the interval between two centres $l_i\Delta, (l_i + 1)\Delta$, it is assigned to the closest one. Further, based on these one-dimensional functions, multidimensional activation functions are formed in the second hidden layer. The number of multiplication blocks in this layer is described by the N value, and it is assumed that there are no identical training observations.

Hence, PNFS learning is reduced to the setting of one-dimensional membership function centres and reconfiguration of multiplication blocks in the aggregation layer. It is clear that such system tuning happens almost instantly.

IV. THE RESULTS OF A COMPUTATIONAL EXPERIMENT

The two datasets were chosen to evaluate the proposed system. The first dataset is “ML hand-written digits” from the UCI repository where tags are rank values, and the second one is “Fashion MNIST”, which has nominal tag values. The first dataset has 1797 images that are matrices 8×8 . The fact that they are low resolution gives the ability to show how well the proposed network works in a fuzzy case. The second dataset has 60 000 instances in the training and 6000 in the test dataset. Each sample of these datasets is a matrix 24×24 . As we can see, the resolution is higher than that of the first dataset, but due to the size of dataset, we can see how the proposed network performs in online mode. Both of these datasets contain overlapping classes. The fuzzy case that appears in the “ML hand-written digits” dataset is represented in Fig. 2 that shows only part of the full visualization. The samples from the both datasets are presented in Fig. 3.

For the comparative analysis, the popular machine learning method K-nearest neighbour (KNN) and evolving fuzzy-probabilistic neural network (EFPNN) proposed in [19] and the probabilistic neuro-fuzzy system (PNFS) were taken. During the first part of the experiment, “ML hand-written digits” were taken, and the spent time, along with the accuracy, were tracked. The obtained results are shown in Table I.

TABLE I
ACCURACY AND TIME CONSUMPTION OF ALGORITHMS WITH “ML HAND-WRITTEN DIGITS” DATASET

Algorithms for comparison	Classification accuracy	Time, sec
KNN	81.98	0.18
EFPNN	96.39	7.02
PNFS	93.07	5.52

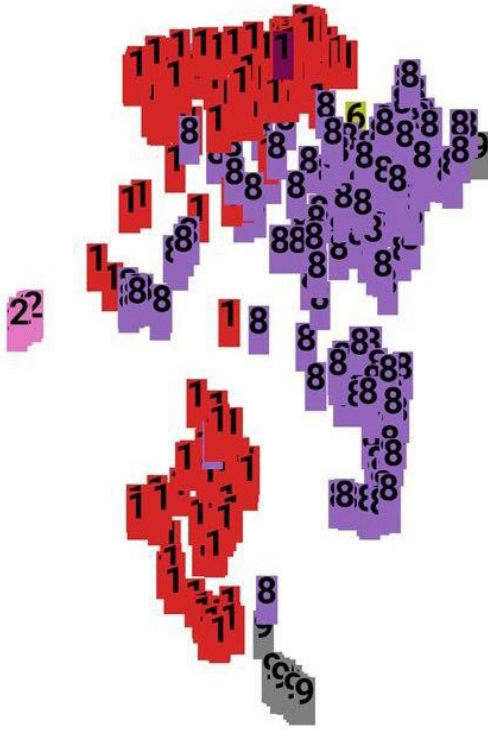


Fig. 2. Overlapping classes in the “ML hand-written digits” dataset.

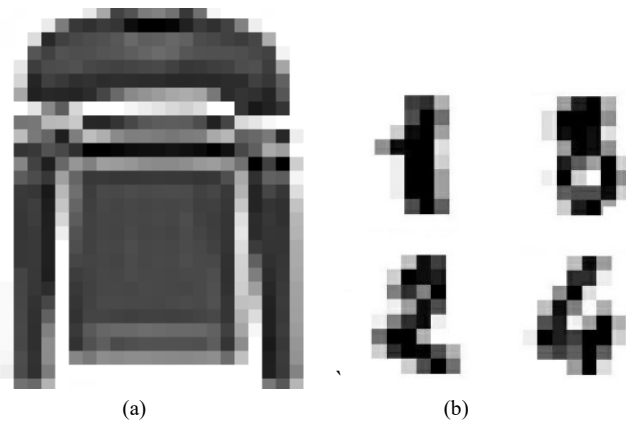


Fig. 3. Example of the observation from: (a) dataset “Fashion MNIST”, (b) dataset “ML hand-written digits”.

As we can see, the probabilistic neuro-fuzzy system works faster than the evolving fuzzy-probabilistic neural network; however, it yields EFPNN in classification accuracy. It has to be taken into account that the K-nearest neighbour algorithm was performed on GPU and the probabilistic neuro-fuzzy system on CPU, which led to the gap in the time spent. Nevertheless, if the probabilistic neuro-fuzzy system had performed on the CPU, the consumed time would be comparable because the speed of the modern CPU would be higher than the recent GPU from five to ten times.

For the second part of the experiment, the “Fashion MNIST” dataset was taken. During this part of the experiment, the initial dataset was taken as nine datasets with different volume, and the classification accuracy was tracked. Each of these datasets was formed randomly. With the initial “Fashion MNIST”

dataset, the biggest dataset that was formed had 15 000 samples. The wide-spread parameter was chosen empirically for the initial dataset to obtain the best results. For the long dataset, the wide-spread parameter was 0.74. The results are presented in Fig. 4.

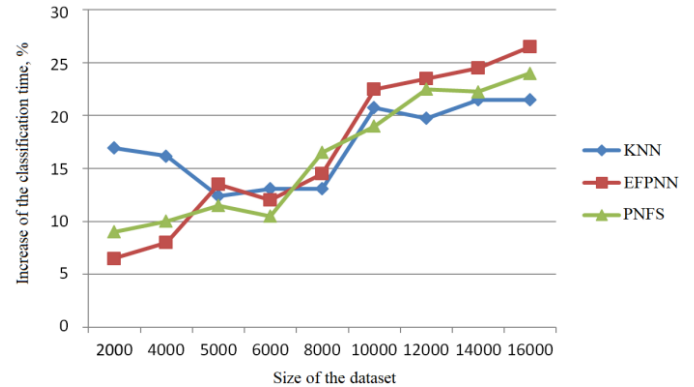


Fig. 4. Dependency of dataset size and the classification time.

As seen in Fig. 4, the KNN algorithm provides the smallest increase in computation time, depending on the dataset size. However, this algorithm has lower accuracy compared to algorithms based on probabilistic neural networks. The proposed PNFS algorithm requires lower computational burden compared to the EFPNN algorithm with similar classification accuracy. On datasets with a larger size, more than 8000 elements, this algorithm shows a nearly linear increase in computational costs with an increase in the sample size.

The increase in computational costs when going from short datasets to the m -th class ones is probably related to features of software implementation. The small dataset fills evenly in RAM, and therefore as it grew, the additional processing costs increased insignificantly. The long dataset requires more resources for swapping pages by the memory operator, which significantly increases RAM costs.

V. CONCLUSION

The proposed probabilistic neuro-fuzzy system for the image recognition-classification task solving is a generalization of Specht’s probabilistic neural network for the case of overlapping classes. It is designed to work with both long and short training datasets. The system is quite similar to the Takagi-Sugeno-Kang’s neuro-fuzzy system of zero rank (Wang–Mendel system) by the architecture. Still, PNFS significantly exceeds it in terms of learning speed, based on the principle of “neurons at data points”. The tuning process is reduced to the setting of one-dimensional membership function centres in the fuzzification-reconfiguration layer, multiplication blocks in the aggregation layer, and happens almost immediately.

The proposed system is simple in numerical implementation, and the experiment confirms its effectiveness.

REFERENCES

- [1] D. F. Specht, "Probabilistic neural networks," *Neural Networks*, vol. 3, no. 1, pp. 109–118, 1990. [https://doi.org/10.1016/0893-6080\(90\)90049-Q](https://doi.org/10.1016/0893-6080(90)90049-Q)
- [2] D. F. Specht, "Probabilistic neural networks and polynomial ADALINE as complementary techniques to classification", *IEEE Trans. on Neural Networks*, vol. 1, pp. 111–121, 1990. <https://doi.org/10.1109/72.80210>
- [3] Ye. Bodyanskiy, Ye. Gorshkov, V. Kolodyazhnyi and J. Wernstedt, "A learning of probabilistic neural network with fuzzy inference," in Pearson D.W., Steele N.C., Albrecht R.F. (eds), *Artificial Neural Nets and Genetic Algorithms*, Springer, Vienna, 2003, pp. 3–17. https://doi.org/10.1007/978-3-7091-0646-4_3
- [4] Ye. Bodyanskiy, Ye. Gorshkov and V. Kolodyazhnyi, "Resource-Allocating Probabilistic Neuro-Fuzzy Network," *Proc. 2nd Conf. of European Union Society for Fuzzy Logic and Technology (EUSFLAT 2003)*, Zittau, Germany, 10–12 September 2003, pp. 392–395.
- [5] Ye. Bodyanskiy, Ye. Gorshkov, V. Kolodyazhnyi, J. Wernstedt, "Probabilistic neuro-fuzzy network with non-conventional activation functions," in: Palade V., Howlett R.J., Jain L. (eds), *Knowledge-Based Intelligent Information and Engineering Systems, KES 2003. Lecture Notes in Artificial Intelligence*, vol. 2773, Berlin, Heidelberg, New York, Springer, 2003, pp. 973–979. https://doi.org/10.1007/978-3-540-45226-3_133
- [6] L. Rutkowski, "Adaptive probabilistic neural networks for pattern classification in time-varying environment," *IEEE Trans. on Neural Networks*, vol. 5, no. 4, pp. 811–827, 2004. <https://doi.org/10.1109/TNN.2004.828757>
- [7] Ye. Bodyanskiy and O. Shubkina, "Semantic annotation of text documents using modified probabilistic neural network," *6th IEEE Int. Conf. on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, Prague, Czech Republic, 2011, pp. 328–331. <https://doi.org/10.1109/IDAACS.2011.6072767>
- [8] Ye. Bodyanskiy, O. Shubkina, "Semantic annotation of text documents using evolving neural network based on principle "Neurons at Data Points" 4th Int. Workshop on Inductive Modeling "IWIM 2011", Kyiv, Ukraine, 2011, pp. 31–37. <https://doi.org/10.1109/IDAACS.2011.6072767>
- [9] Ye. Bodyanskiy, I. Pliss and V. Volkova, "Modified probabilistic neuro-fuzzy network for text document processing," *Int. J. Computing*, vol. 11, no. 4, 2012, pp. 391–396.
- [10] J.-H. Yi, J. Wang, G.-G. Wang, "Improved probabilistic neural networks with self-adaptive strategies for transformer fault diagnosis problem," *Advances in Mechanical Engineering*, vol. 8, no. 1, 2016, pp. 1–13. <https://doi.org/10.1177/1687814015624832>
- [11] P. Zhernova, I. Pliss, O. Chala "Modified fuzzy probabilistic neural network," *Intellectual Systems for Decision Making and Problems of Computational Intelligence ISDMCI'2018*, Kherson, PP Vyshemirsky V. S., pp. 228–230, 2018.
- [12] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. N.Y.: Plenum Press, 1987.
- [13] C. Mumford, L. Jain (Eds), *Computational Intelligence, Collaboration, Fuzzy and Emergence*. Berlin, Springer - Verlag, 2009. <https://doi.org/10.1007/978-3-642-01799-5>
- [14] R. Kruse, C. Borgelt, F. Klawonn, C. Moewes, M. Steinbrecher, P. Held, *Computational Intelligence. A Methodological Introduction*. Berlin: Springer-Verlag, 2013. <https://doi.org/10.1007/978-1-4471-5013-8>
- [15] J. Kacprzyk, W. Pedrycz, *Springer Handbook of Computational Intelligence*. Berlin, Heidelberg, Springer-Verlag, 2015. <https://doi.org/10.1007/978-3-662-43505-2>
- [16] P.V.C. Souza, "Fuzzy neural networks and neuro-fuzzy networks: A review the main techniques and applications used in the literature," *Applied Soft Computing*, vol. 92, July 2020. <https://doi.org/10.1016/j.asoc.2020.106275>
- [17] D.R. Zahimiak, R. Chapman, S.K. Rogers, B.W. Suter, M. Kabriski, V. Pyatti, "Pattern recognition using radial basis function network," *Proceedings Aerospace Application of Artificial Intelligence*. Dayton, Ohio, 1990, pp. 249–260.
- [18] O. Nelles, *Nonlinear Systems Identification*. Berlin, Springer, 2001. <https://doi.org/10.1007/978-3-662-04323-3>
- [19] Ye. Bodyanskiy, A. Deineko, I. Pliss, O. Chala "Evolving fuzzy-probabilistic neural network and its online learning" *2020 10th International Conference on Advanced Computer Information Technologies*. Deggendorf, Germany, 16–18 September, 2020, in Press. <https://doi.org/10.1109/ACIT49673.2020.9208904>
- Yevgeniy Bodyanskiy.** In 1971, he graduated with honours 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. 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 the period of 1974–1976, he was a Researcher; from 1977 to 1983, he was a Senior Researcher; from 1986 to 1991, he was a Scientific Head of Control Systems Research Laboratory; from 1991 to 1992, he was a Research Fellow. Since 1992, he has been 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 related to 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.
Contact data: Office 511, Nauky Ave., 14, Kharkiv, 61166, Ukraine.
E-mail: yevgeniy.bodyanskiy@nure.ua
ORCID iD: <https://orcid.org/0000-0001-5418-2143>
- Anastasiia Deineko.** In 2011, she graduated with honours from Kharkiv National University of Radio Electronics. In 2015, she received the degree of Candidate of Technical Sciences. Her major fields of research are evolving hybrid systems of computational intelligence, data stream mining. Since 2015, she has been working at Kharkiv National University of Radio Electronics. In the period of 2015–2017, she was an Assistant of Artificial Intelligence Department; from 2017 to 2018, she was a Senior Lecturer of Artificial Intelligence Department. Since 2019, she has been an Associated Professor of Artificial Intelligence Department at KhNURE. She has more than 35 publications, including publications in scientometric databases, such as Scopus and Web of Science. Research interests include hybrid systems of computational intelligence: adaptive, neuro-, fuzzy-, kernel-, real-time systems, including problems related to forecasting, clustering, pattern recognition.
Contact data: Office 255 AI Dept., Nauky Ave., 14, Kharkiv, 61166, Ukraine.
E-mail: anastasiia.deineko@nure.ua
ORCID iD: <https://orcid.org/0000-0002-3279-3135>
- Irina Pliss.** In 1970, she was qualified as Electrical Engineer in the specialty of Automatics and Telemechanics at Kharkiv National University of Radio Electronics (KhNURE). In the same year, she was an Engineer of Automatics Department at KhNURE. In the period of 1970–1973, she was a Junior Researcher of Automatics Department, KhNURE. From 1973 to 1976, she was a PhD student of Automatics Department, KhNURE. From 1978 to 1992, she was a Senior Researcher of Control Systems Research Laboratory, KhNURE. In 1978, she received the degree of Candidate of Technical Sciences (PhD) at KhNURE in the scientific specialty 05.13.06 – Technical Cybernetics and Information Theory. In 1984, she was a Senior Researcher, the scientific specialty 05.13.01 – Technical Cybernetics and Information Theory and since 1992 – Leading Researcher of Control Systems Research Laboratory, KhNURE. She has currently published over 280 scientific papers, including 8 monographs and book chapters published by Springer, Nova Science Publishers Inc., John Wiley (New York, USA). Her scientific interests are hybrid systems of computational intelligence, data stream mining, big data, deep learning, and evolving systems.
Contact data: Office 511, Nauky Ave., 14, Kharkiv, 61166, Ukraine.
E-mail: iryna.pliss@nure.ua
ORCID iD: <https://orcid.org/0000-0001-7918-7362>
- Olha Chala.** In 2019, she graduated with an honour and obtained Bachelor degree in Computer Science at Kharkiv National University of Radio Electronics (KhNURE). In 2019, she initiated studies at the master study programme "Systems of Artificial Intelligence" at KhNURE. Since 2018, she has been working at Control Systems Research Laboratory, KhNURE. She has 10 scientific publications. Her current scientific interests are neuro-fuzzy systems, deep learning, and pattern recognition task.
Contact data: Office 511, Nauky Ave., 14, Kharkiv, 61166, Ukraine.
E-mail: chala.olha@nure.ua
ORCID iD: <https://orcid.org/0000-0002-7603-1247>