

# Generalized neo-fuzzy-neuron with membership functions of special type in medical diagnostics

Yevgeniy Bodyanskiy<sup>a</sup> and Olha Chala<sup>b</sup>

<sup>a</sup> Control systems research laboratory, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

<sup>b</sup> Artificial intelligence department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

## Abstract

The neuro-fuzzy system for medical diagnostics under conditions of short training datasets and overlapping classes is proposed. The proposed system is based on the multidimensional generalized neuro-fuzzy neuron which was modified by introducing an additional softmax output layer and membership functions of the special form. As for the learning, the optimal by speed adaptive algorithm provided additional soothing properties. In contrast to well-known neuro-fuzzy systems, the proposed one is simple in numerical implementation and has fewer tuning parameters. In this case, the possibility of processing information that is specified in different scales is provided. Experiments have confirmed the effectiveness of the proposed system, especially in situations where information is fed for processing in online mode.

## Keywords 1

Neo-fuzzy neuron, online learning, medical diagnostics, membership functions, overlapping classes

## 1. Introduction

Nowadays, artificial intelligence methods, in particular computational intelligence [1-3] gained its spread to solve many tasks which are related to medical information processing, and specifically in medical diagnostics task [4-12]. Generally this task can be considered as pattern classification-recognition problem under conditions of deficit of priori information and overlapping classes with its arbitrary form in multidimensional space of features that can be defined in numerical, nominal and ordinal scale. Artificial neural networks, both traditional shallow [13] and more advanced deep [14, 15] ones, which provide high quality classification due to their universal approximating properties, proved to be the best here. As for situations with overlapping classes, the most effective are neuro-fuzzy systems [3, 16, 17], which are hybrid systems of computational intelligence that combine the properties of both artificial neural networks and fuzzy reasoning systems. These systems require significant amounts of a priori information from training samples for their training settings, which are not always available in medical diagnostic tasks.

In situations of limited training dataset, an alternative to multilayer neural networks can be used probabilistic ones [18, 19], which have shown themselves quite well in medical applications [20]. Probabilistic neural networks implement their training on the basis of the principle "Neurons in data points" [21], i.e. the so-called lazy learning [22], when the centers of kernel activation functions of the image layer coincide with the coordinates of observations from the training dataset. The disadvantage of probabilistic neural networks is that it is a necessity of the a priori task of the whole training dataset. This means that if the training dataset receives one classified observation, then in order to take this into account it is necessary to restructure the network architecture as a whole.

---

Proceedings Name, Month XX-XX, YYYY, City, Country

EMAIL: yevgeniy.bodyanskiy@nure.ua (A. 1); olha.chala@nure.ua (A. 2)

ORCID: 0000-0001-5418-2143 (A. 1); 0000-0002-7603-1247 (A. 2)



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

In situations of limited training dataset and sequential online data entry and processing in the learning mode, the so-called neo-fuzzy systems [23-26] can be most effective, characterized by good approximating properties, and can be tuned online with the maximum possible learning speed [27].

At the same time, the recognition system is based on neo-fuzzy neurons, not designed to work with many classes [28], redundant, contains too many membership functions and tuned parameters, which makes it cumbersome and difficult to implement.

This problem can be overcome by using the idea of a generalized neo-fuzzy neuron [29], which in the general case is an approximating multidimensional system and is not meant to solve classification problems. Therefore, it is advisable to introduce its modification, which is designed specifically to solve the problems of image recognition with high accuracy and speed in terms of overlapping classes.

Hence, the subject of the research in this paper analysis process of medical data in online mode under conditions of overlapping classes of diagnosis. The generalized modified neo-fuzzy neuron and process of its online learning under conditions of priori information deficit.

## 2. Architecture of the modified generalized neo-fuzzy-neuron for classification-diagnostic task solving

The generalized neo-fuzzy neuron in the general case is a multidimensional nonlinear system that is able to adjust its parameters-synaptic weights online, while implementing the mapping.

$$o(\tau) = F(x(\tau)), o \in R^m, x \in R^n. \quad (1)$$

Through the approximation of priori unknown multidimensional function according to observation- vectors  $o(\tau) = (o_1(\tau), \dots, o_j(\tau), \dots, o_m(\tau))^T$ ,  $x(\tau) = (x_1(\tau), \dots, x_i(\tau), \dots, x_n(\tau))^T$ ,  $\tau = 1, 2, \dots, T$  – the number of the observation vectors in the training dataset, or the index of the current discrete time in the case when the information for processing is fed in real time. In addition to the actual problem of approximation, GNFN can be used in solving problems of extrapolation-prediction of multidimensional sequences, nonlinear identification, and adaptive control of nonlinear objects [29].

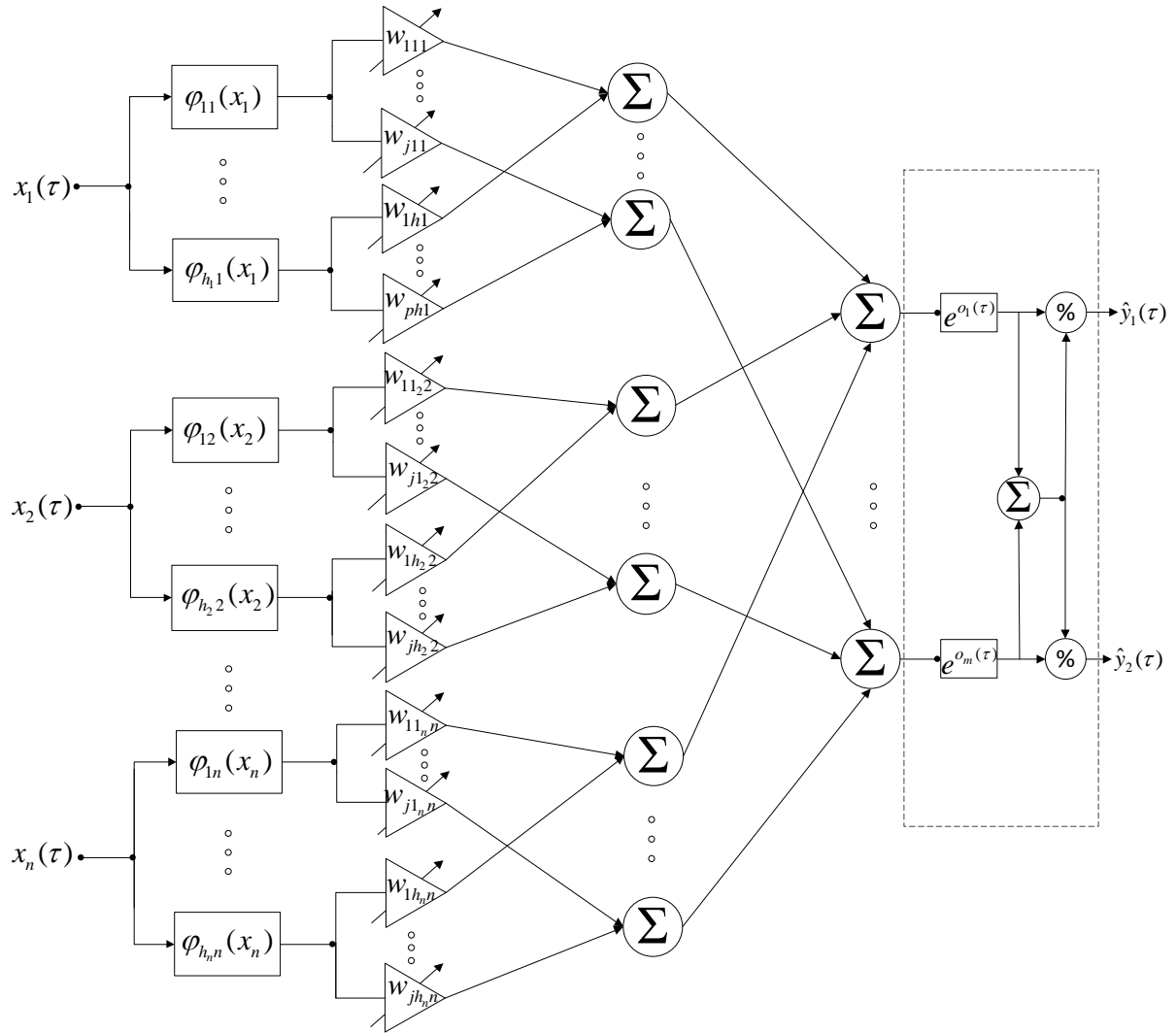
In order to enable the solution of image classification-recognition problems, an additional output layer formed by softmax activation functions [14] should be added to GNFN, by means of which the procedure of defuzzification of output signals is first implemented  $o_j(\tau)$ , and secondly, it is possible to use crossentropy as a learning criterion, which is the most common in the problems of pattern recognition.

In Figure1 is shown the architecture of a generalized neo-fuzzy neuron with an additional layer of defuzzification to solve pattern recognition problem of medical diagnostics, the distinctive feature of which is that the classes of diagnoses arbitrarily overlap, can have a rather complex shape, and input signals can be set in any scale, not just in the traditional numerical. It is easy to see that this scheme is a kind of hybrid of traditional GNFN and Takagi-Sugeno-Kang neuro-fuzzy-system, which is a multidimensional universal approximator, the training of which may encounter some computational difficulties [16]. It should also be added that the system under consideration can in principle be implemented by connecting ordinary neo-fuzzy neurons in parallel with the softmax output function. However, such a scheme will contain  $n$  times more membership functions in the first layer, which complicates its implementation.

The nonlinear transformation is implemented by the system and shown in Figure 1, in the general case can be written as

$$\begin{cases} o_j(\tau) = \sum_{i=1}^n \sum_{l=1}^{h_i} \varphi_l(x_i(\tau)) w_{jli}(\tau-1), \\ \hat{y}_j(\tau) = \text{softmax } o_j(\tau) = \frac{e^{o_j(\tau)}}{\sum_{p=1}^m e^{o_p(\tau)}}, \forall j = 1, 2, \dots, m \end{cases} \quad (2)$$

where  $w_{ji}(\tau-1)$  – tuned synaptic weights, which were obtained on the basis of  $\tau-1$  previous observations,  $y_i(x_i)$  – one-dimensional membership functions, the total number of which is equal to  $\sum_{i=1}^n h_i$  (here  $h_i$  – the number of these functions at the  $i$ -th input of the system). Using of these functions the fuzzification of the input variables space is realized, i.e. this dimension is increased, which allows solving linearly inseparable recognition problems.



**Figure 1:** Generalized neo-fuzzy neuron with membership functions of special type in medical diagnostics tasks

In traditional neo-fuzzy neurons as such functions, triangular ones are usually used, which satisfy the Ruspini partition condition. The advantages of these functions are, firstly, the system does not require an output layer of defuzzification, and, secondly, for processing each observation  $x_i(\tau)$  the only two neighbouring functions on each input are fired. This in its turn leads to the fact that at each moment of the current time are tuned only  $2n$  synaptic weights  $w_{ji}(\tau)$  that simplify the implementation of the learning procedure. In principle, in neo-fuzzy neurons, any function that

satisfies Ruspini partitioning conditions can be used, such as B-splines, but in this case, each step must be tuned for all  $\sum_{i=1}^n h_i$  synaptic weights.

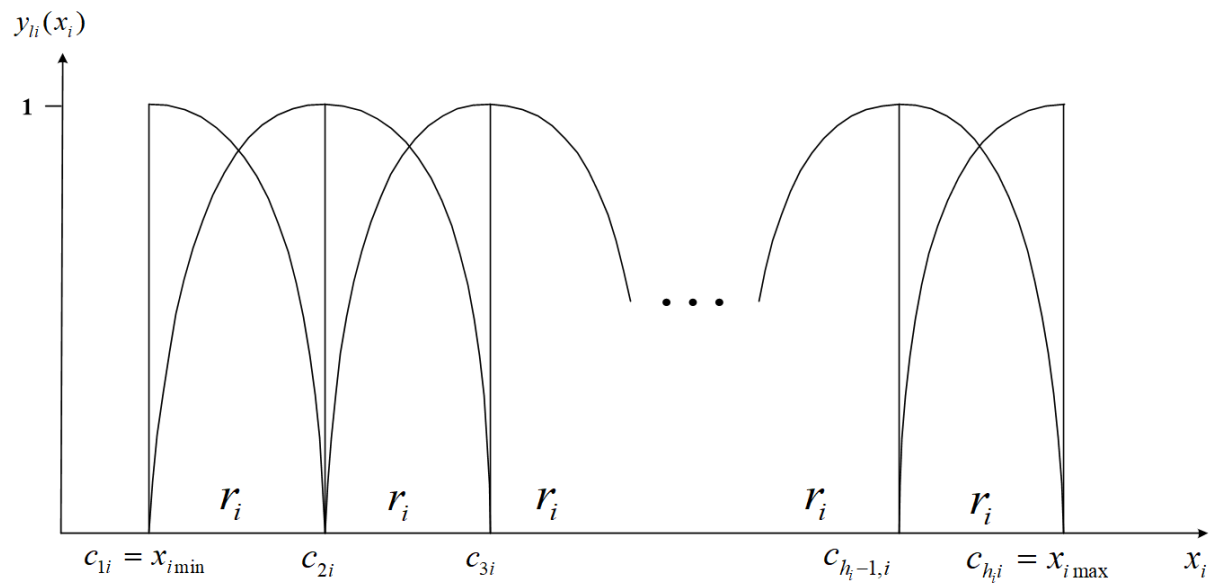
Since in the system under consideration the softmax layer performs the defuzzification procedure at the same time, there is no need to fulfill the conditions of unity partitioning. Therefore, any functions used in traditional neuro-fuzzy systems can be used as membership ones. These are usually kernel functions (Gaussians, Cauchyans e.a.), which are widely used in Parzen "windows" [30], or Nadaraya-Watson estimates [31]. Since these functions have an infinite interval of determination on each step, it is necessary to tune all the synaptic weights of the system.

Therefore, we propose to use Epanechnikov's kernel functions [32], which are shown in Figure2, as membership functions and can be written in analytical form as it is presented in formula 3

$$y_{li}(x_i) = \left( 1 - \frac{(x_i - c_{li})^2}{r_i^2} \right) \delta_{li} \quad (3)$$

where  $c_{li}$  – centers (extremum points) of the corresponding functions,  $r_i$  – the distance between two tuned centers (in the case of their uniform distribution on the interval of determination  $[x_{i \min}, x_{i \max}]$ ),

$$\delta_{li} = \begin{cases} 1 & \text{if } |x_i - c_{li}| < r_i, \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$



**Figure 2:** Epanechnikov membership functions

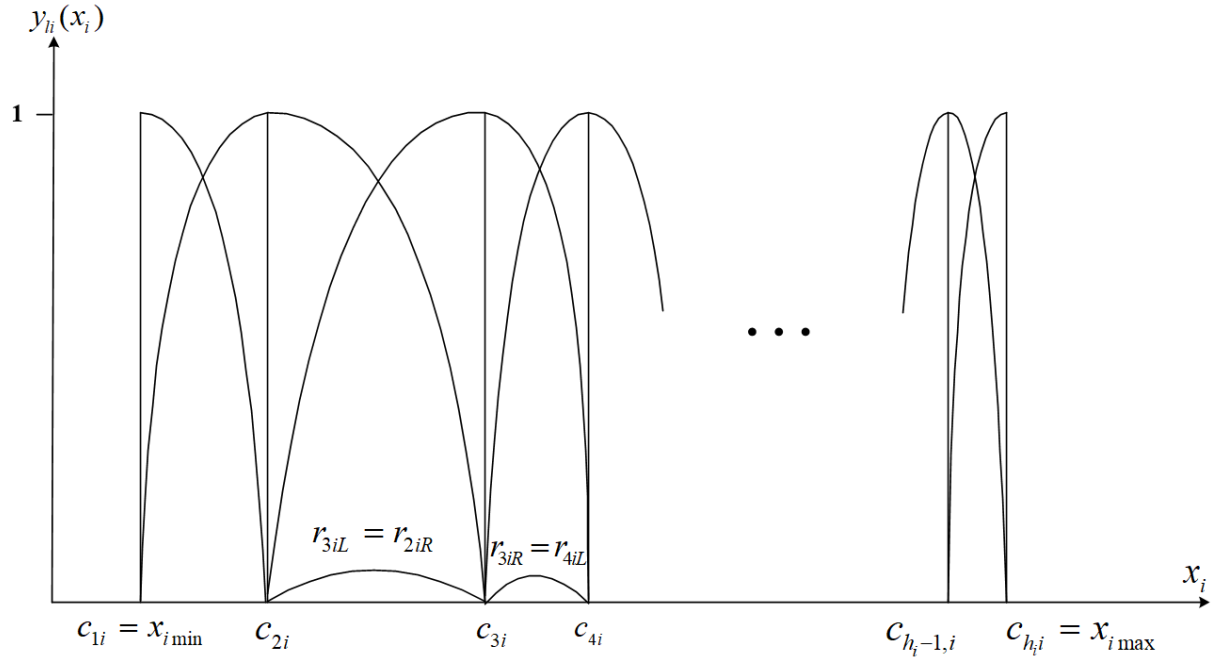
If the  $i$ -th input is evenly spaced  $h_i$  functions, the distance between the centers is described by the expression

$$r_i = \frac{x_{i \min} - x_{i \max}}{h_i - 1}. \quad (5)$$

It is important to notice that in the case of binary inputs which are often used in medical diagnostic tasks requires only two membership functions with centers  $c_{1i} = 0, c_{2i} = 1$ , and  $r_i = 1$ .

It should also be noted that in comparison with triangular membership functions that implement piecewise-linear approximation, approximation using Epanechnikov's linear functions provides a higher accuracy separating hypersurfaces' reconstruction in the feature space.

In a more general case, the centers of membership functions can be located unevenly, as shown in Figure 3.



**Figure 3:** Unsymmetrical Epanechnikov membership functions

In this case, the membership functions are not symmetrical and can be written as

$$\begin{aligned}\varphi_{liL}(x_i) &= \left[ 1 - \frac{(x_i - c_{li})^2}{(c_{l-1,i} - c_{li})^2} \right]_+, \\ \varphi_{liR}(x_i) &= \left[ 1 - \frac{(x_i - c_{li})^2}{(c_{l+1,i} - c_{li})^2} \right]_+\end{aligned}\quad (6)$$

where  $[\cdot]_+ = \max\{0, \cdot\}$ .

Thus, output signals are generated at the GNFN outputs

$$o_j(\tau) = \sum_{i=1}^n \sum_{l=1}^{h_i} \varphi_{li}(x_i(\tau)) w_{jli}(\tau - 1) \quad (7)$$

(to the tact of refining synaptic weights based on observation  $x_i(\tau)$ ), which pass through the output layer of softmax activation functions, forming levels of fuzzy membership  $\hat{y}_j(\tau)$  of each observation  $x(\tau)$  to each of possible classes-diagnoses  $c_{ji}$ ,  $j = 1, 2, \dots, m$ .

### 3. Learning of the generalized neo-fuzzy-neuron with softmax output layer in classification-diagnostic task solving

As a learning criterion, the crossentropy criterion can be used, which is most often currently implemented in convolutional deep neural networks [14].

$$J(\tau) = \sum_{j=1}^m J_j(\tau) = - \sum_{j=1}^m y_j(\tau) \ln \hat{y}_j(\tau) \quad (8)$$

where  $y_j(\tau)$  – an external reference signal that determines whether or not the observation  $x(\tau)$  is related to  $j$ -th class and can only take two values 0 or 1 (so-called “one-hot coding”).

Let's introduce further into consideration  $\sum_{i=1}^n h_i \times 1$  – vector of the mutual membership functions  $\varphi(x) = (\varphi_1(x_1), \varphi_2(x_2), \dots, \varphi_{h_1}(x_1), \dots, \varphi_{l_i}(x_i), \dots, \varphi_{h_n}(x_n))^T$  and the vector of synaptic weights associated with the  $j$ -th output of the system  $w(x) = (w_{j11}, \dots, w_{jh_1}, \dots, w_{jl1}, \dots, w_{jh_n})^T$  then let's write down initial signals and local criterion of training in the form

$$o_j(\tau) = w_j^T(\tau-1)\varphi(x(\tau)), \quad (9)$$

$$\hat{y}_j(\tau) = \frac{e^{o_j(\tau)}}{\sum_{p=1}^m e^{o_p(\tau)}} = \frac{e^{w_j^T(\tau-1)\varphi(x(\tau))}}{\sum_{p=1}^m e^{w_p^T(\tau-1)\varphi(x(\tau))}}, \quad (10)$$

$$J_j(\tau) = -y_j(\tau) \ln \hat{y}_j(\tau) = -y_j(\tau) \ln \frac{e^{w_j^T(\tau-1)\varphi(x(\tau))}}{\sum_{p=1}^m e^{w_p^T(\tau-1)\varphi(x(\tau))}}. \quad (11)$$

Using as a learning algorithm a gradient procedure for optimizing the criterion  $J_j(\tau)$  easy to write

$$w_j(\tau) = w_j(\tau-1) - \eta(\tau) \nabla_{w_j} J_j(\tau) = w_j(\tau-1) + (y_j(\tau) - w_j^T(\tau-1)\varphi(x(\tau)))\varphi(x(\tau)) \quad (12)$$

where  $\eta(\tau)$  – learning rate parameter. Optimizing this algorithm for the rate of convergency, coming to the following procedure

$$w_j(\tau) = w_j(\tau-1) + \frac{y_j(\tau) - w_j^T(\tau-1)\varphi(x(\tau))}{\varphi^T(x(\tau))\varphi(x(\tau))} \varphi(x(\tau)) \quad (13)$$

which is essentially the optimal Kaczmarz-Widrow-Hoff algorithm [33, 34], which provides the maximum possible rate of convergency in the class of gradient procedures.

The regularized version of this algorithm can also be used to protect against the possible effect of an "exploding" gradient.

$$w_j(\tau) = w_j(\tau-1) + \frac{(y_j(\tau) - \hat{y}_j(\tau))\varphi(x(\tau))}{\gamma + \|\varphi(x(\tau))\|^2} \quad (14)$$

(here  $\gamma > 0$  – momentum term) or exponentially weighted stochastic approximation [35]

$$\begin{cases} w_j(\tau) = w_j(\tau-1) + r^{-1}(\tau)(y_j(\tau) - w_j^T(\tau-1)\varphi(x(\tau)))\varphi(x(\tau)), \\ r(\tau) = \alpha r(\tau-1) + \|\varphi(x(\tau))\|^2 \end{cases} \quad (15)$$

where  $0 \leq \alpha \leq 1$  – forgetting factor that provides additional filtering properties of the learning process and coincides with the algorithm of Kaczmarz -Widrow-Hoff [36] when  $\alpha = 0$ . Having  $\alpha = 1$  this procedure is transformed into a Goodwin-Ramage-Caines stochastic approximation algorithm [37].

The quality of the learning process of the system under consideration can be improved by tuning not only the synaptic weights, but also the corresponding memberships of the first layer of the system [38]. It is quite simple to do this, using the ideas of lazy learning [21], which is the basis of probabilistic neural networks.

Introducing some threshold of proximity of two neighbouring centers on each input

$$r_{i\min} = |c_{l-1,i} - c_{li}|_{\min} = |c_{l+1,i} - c_{li}|_{\min}, \quad (16)$$

the process of setting centers can be organized as follows: the first center coincides with the first observation from the training dataset, i.e.  $c_1 = x(1)$ . Upon receipt of the second observation  $x(2)$  check this distance at all coordinates from  $c_1$ . If this distance is less than  $r_{i\min} \forall i$ , then a new center is not formed. If this distance is greater than  $r_{i\min} \forall i$ , then the center is set  $c(2)$ . This process takes place until the entire system of centers, the number of which is determined by the value (formula 17), is established:

$$h_i = \frac{x_{i\max} - x_{i\min}}{r_{i\min}} + 1. \quad (17)$$

It is clear that this procedure can be implemented only for inputs that receive information in a numerical scale.

## 4. Experiment results

The experiment was held with the usage of “Urinary biomarkers for pancreatic cancer” [39] dataset from the Kaggle repository, since this type of cancer is extremely aggressive and lethal, because on early stages it is extremely hard to detect since there are no symptoms, thus, the survival rate within five years after diagnosis less than ten per cent. However, if is caught early, the odds of surviving are quite promising. Hence, the detection of pancreatic cancer is a burning issue.

The dataset itself contains fourteen features, however, the key ones among them are creatinine, LYVE1, REG1B, and TFF1. Where, creatinine is a protein for an indication of kidney function, LYVE1( lymphatic vessel endothelial hyaluronan receptor 1) – a protein that can take part in tumour metastasis, REG1B is a protein that can be related to pancreas regeneration, and TFF1(trefoil factor 1) can be associated to regeneration and repair of the urinary tract. Besides that, the dataset contains age and sex that can also be related to the fact of having pancreatic cancer, also there are a few other biomarkers as well, however, not all patients were measured.

The goal function of this dataset can take three possible values 3 (pancreatic cancer), 2 (non-cancerous pancreas condition) and 1 (healthy), also the dataset length is 590 instances. The available data from the dataset are presented in the Table 1.

**Table 1**  
Sample of the dataset

patient_cohort	sample_origin	age	sex	diagnosis	stage	benign_sample_diagnosis	plasma_CA19_9	creatinine	LYVE1	REG1B	TFF1	REG1A
Cohort1	BPTB	33	F	1	NaN	NaN	11.7	1.83222	0.893219	52.94884	654.282174	1262.000
Cohort1	BPTB	81	F	1	NaN	NaN	NaN	0.97266	2.037585	94.46703	209.488250	228.407
Cohort2	BPTB	51	M	1	NaN	NaN	7.0	0.78039	0.145589	102.36600	461.141000	NaN
Cohort2	BPTB	61	M	1	NaN	NaN	8.0	0.70122	0.002805	60.57900	142.950000	NaN
Cohort2	BPTB	62	M	1	NaN	NaN	9.0	0.21489	0.000860	65.54000	41.088000	NaN

It can be seen that raw data has different scales such as nominal and numerical. As for preprocessing the one-hot coding was used for the first scale.

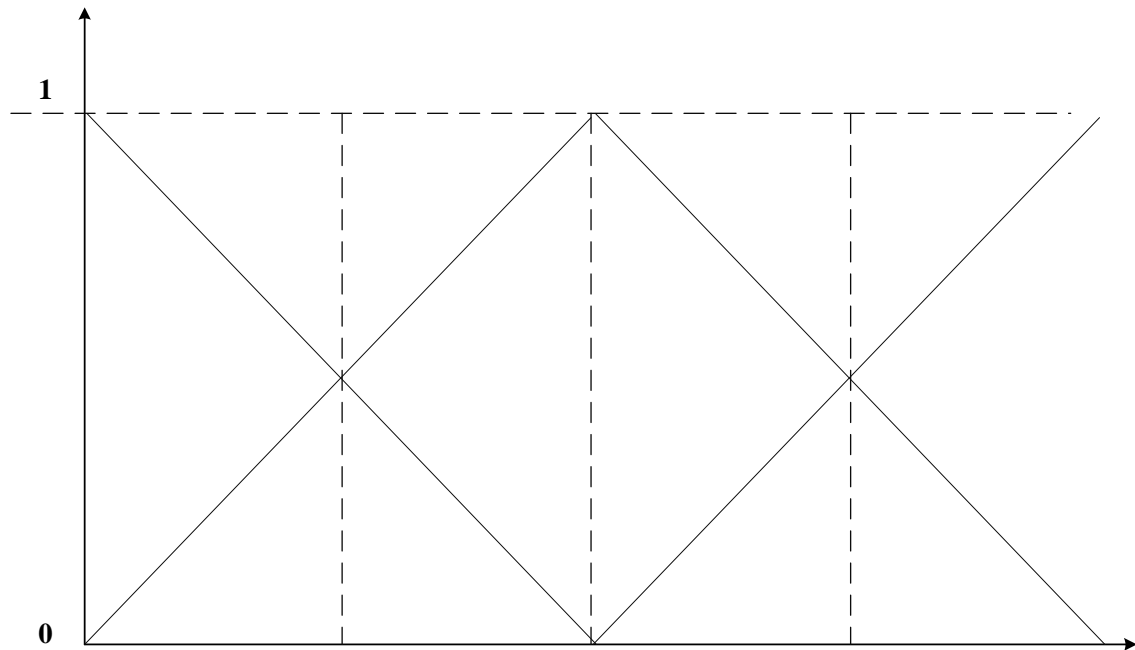
The idea that underlie in the following experiment is the test of the efficiency of the chosen activation function. Thus, the proposed network was taken as a basis. However, the activation function was changed from the Epanechnikov to the mentioned first order B-splines instance of which can be seen in the Figure 5 and the traditional Gaussian activation function which, by its nature can provide membership levels – it can work with overlapping classes.

**Table 2**  
Comparison analysis of the membership functions

Membership functions	Accuracy, %	Time, sec
Epanechnikov	94,78	49
First order B-splines	81,301	51
Gaussian	72,37	133

From the Table 2 it can be seen that the Epanechnikov’s membership function does not yield to the first-order B-splines in accuracy, but still having comparable time as it was expected from the piece-wise approximation. Having Gaussian as a membership activation function, the system requires additional normalisation step so the Ruspini condition partition will be satisfied and it is important to notice, that the function itself has cumbersome computations. This slows down the data processing

and leading to the unnecessary steps of the whole process, additionally, normalisation leading to the loss of important information thus the accuracy is low.



**Figure 5:** Instance of the first order B-splines (triangle membership functions)

The next step was the comparison of the proposed approach with the existing ones. In this case, the advanced systems such as deep neural networks that could have provided with high accuracy can be used, however, considering the volume of dataset and the importance of the speed, and the hardware resource usage, the usage of DNNs is ineffective. Thus, the simple machine learning approach “Support vector machine” (SVM) and the “Multilayer perceptron” (MLP) we are chosen for the further comparison.

**Table 3**

Quantitative indicators of diagnostics process

Membership functions	Accuracy, %	Time, sec
GNFN	94,78	49
SVM	77,83	98
MLP	79,98	106

As it can be seen the both accuracy and time values that were obtained from SVM and MLP algorithms are yield to the proposed approach since machine learning algorithms are crisp and can work only under conditions when classes are not overlapped. Besides that the volume of the dataset may have influenced on the speed of both approaches leading to the more significant time-consumption in comparison to the proposed approach.

## 5. Conclusion

In order to solve the problems of medical diagnostics from the theory point of view of the image recognition-fuzzy classification task, a hybrid system of computational intelligence is proposed. This system is essentially a combination of a generalized neo-fuzzy neuron, neuro-fuzzy system and a convolutional deep neural network. The system under consideration is designed to operate under conditions of overlapping classes of arbitrary shape and limited training dataset quite often found in real situations associated with a deficit of a priori information. The optimal speed algorithm is used to



train the system which can be provided with additional filtering properties. The results of the experiment prove that the proposed system is simple in numerical implementation and provides high quality diagnosis, in comparison to traditional artificial neural networks, under conditions of ambiguity and lack of a priori information. High speed of the proposed system allows processing information in online mode.

## 6. References

- [1] C. Mumford, L. Jain, Computational Intelligence, Collaboration, Fuzzy and Emergence, Berlin: Springer, Verlag, 2009. doi: 10.1007/978-3-642-01799-5.
- [2] R. Kruse, C. Borgelt, F. Klawonn, C. Moewes, M. Steinbrecher, P. Held, Computational Intelligence. A Methodological Introduction. Berlin: Springer-Verlag, 2013. doi: 10.1007/978-1-4471-5013-8.
- [3] J. Kacprzyk, W. Pedrycz, Springer Handbook of Computational Intelligence, Berlin Heidelberg: Springer, Verlag, 2015. doi: 10.1007/978-3-662-43505-2.
- [4] M. Schmitt, H.-N. Teodorescu, A. Jain, A. Jain, S. Jain, Computational Intelligence Processing in Medical Diagnosis. Springer-Verlag Berlin Heidelberg, 2002. doi: 10.1007/978-3-7908-1788-1.
- [5] Eu. Giannopoulou, Data Mining in Medical and Biological Research. N.Y.: ITAC, 2008. doi: 10.5772/95.
- [6] P. Berka, S. Rauch, D. Zighed, Data Mining and Medical Knowledge Management: Cases and Applications. N.Y.: Herskey, 2009. doi: 10.4018/978-1-60566-218-3.
- [7] R. Kantchev, R.: Advances in Intelligent Analysis of Medical Data and Decision Support Systems. Springer, 2013. doi: 10.1007/978-3-319-00029-9.
- [8] Ye. Bodyanskiy, I. Perova, O. Vynokurova, I. Izonin, Adaptive wavelet diagnostic neuro-fuzzy network for biomedical tasks. Conference: 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET), 2018, pp. 711 – 715. doi: 10.1109/TCSET.2018.8336299.
- [9] Yu. Syerov, N. Shakhovska, S. Fedushko, Method of the Data Adequacy Determination of Personal Medical Profiles. Advances in Artificial Systems for Medicine and Education II. Springer Nature Switzerland AG, 2018, pp. 333-343. doi: 10.1007/978-3-030-12082-5\_31.
- [10] R. Tkachenko, I. Izonin, and P. Tkachenko, “Neuro-Fuzzy Diagnostics Systems Based on SGTM Neural-Like Structure and T-Controller,” in Lecture Notes in Computational Intelligence and Decision Making, vol. 77, S. Babichev and V. Lytvynenko, Eds. Cham: Springer International Publishing, 2022, pp. 685–695. doi: 10.1007/978-3-030-82014-5\_47.
- [11] S. Nataliya, V. Oleh, and H. Mariia, “Generalized formal model of big data,” arXiv:1905.03061 [cs], May 2019, Accessed: Nov. 02, 2021. [Online]. Available: <http://arxiv.org/abs/1905.03061>
- [12] I. Izonin, R. Tkachenko, M. Gregus ml., K. Zub, and P. Tkachenko, “A GRNN-based Approach towards Prediction from Small Datasets in Medical Application,” Procedia Computer Science, vol. 184, pp. 242–249, 2021, doi: 10.1016/j.procs.2021.03.033.
- [13] R. Tkachenko, I. Izonin, P. Vitynskyi, N. Lotoshynska, and O. Pavlyuk, “Development of the Non-Iterative Supervised Learning Predictor Based on the Ito Decomposition and SGTM Neural-Like Structure for Managing Medical Insurance Costs,” Data, vol. 3, no. 4, p. 46, Oct. 2018, doi: 10.3390/data3040046.
- [14] I. Izonin, R. Tkachenko, S. Fedushko, D. Koziy, K. Zub, and O. Vovk, “RBF-Based Input Doubling Method for Small Medical Data Processing,” in Advances in Artificial Systems for Logistics Engineering, vol. 82, Z. Hu, Q. Zhang, S. Petoukhov, and M. He, Eds. Cham: Springer International Publishing, 2021, pp. 23–31. doi: 10.1007/978-3-030-80475-6\_3.
- [15] K.-L. Du and M. N. S. Swamy, Neural Networks and Statistical Learning. London: Springer London, 2014. doi: 10.1007/978-1-4471-5571-3.
- [16] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning. MIT Press, 2016.
- [17] D. Graupe, Deep Learning Neural Networks: Design And Case Studies. N.Y.: World Scientific, 2016.

- [18] S. Osowski, *Sieci neuronowe do przetwarzania informacji*, Warszawa: Oficyna Wydawnicza Politechniki Warszawskiej, 2006.
- [19] Souza, P.V.C.: Fuzzy neural networks and neuro-fuzzy networks: A review the main techniques and applications used in the literature. *Applied Soft Computing*, vol. 92 (2020).
- [20] 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.
- [21] D. F. Specht, "Probabilistic neural networks," *Neural Network*, vol. 3, pp. 109-118, 1990.
- [22] Y. Bodyanskiy, A. Deineko, I. Pliss, O. Chala, "Probabilistic Neuro-Fuzzy System in Medical Diagnostic Task and its Lazy Learning-Selflearning," *Informatics & Data-Driven Medicine (IDDM) 2020*, pp. 29-35.
- [23] D.R. Zahiriak, R. Chapman, Rogers, Pattern recognition using radial basis function networks, in: *Sixth Annual Aerospace Applications of AI Conf*, Dayton, 1990, pp. 249–260.
- [24] O. Nelles, *Nonlinear Systems Identification*. Berlin: Springer, 2001. doi:10.1007/978-3-662-04323-3.
- [25] T. Yamakawa, E. Uchino, J. Miki, H. Kusanagi, A neo-fuzzy neuron and its application to system identification and prediction of the system behavior, in: *Proceedings of the 2nd International Conference on Fuzzy Logic & Neural Networks*, Iizuka, Japan, 1992.
- [26] E. Uchino, T. Yamakawa, *Soft Computing Based Signal Prediction, Restoration, and Filtering*, in: D. Ruan (Ed.), *Intelligent Hybrid Systems*, Springer US, Boston, MA, 1997, pp. 331–351. [https://doi.org/10.1007/978-1-4615-6191-0\\_14](https://doi.org/10.1007/978-1-4615-6191-0_14).
- [27] Miki T, Yamakawa T, "Analog implementation of neo-fuzzy neuron and its on-board learning," *In Computational Intelligence and Applications*, Piraeus: WSES Press, 1999, pp. 144-149.
- [28] D. Zurita, M. Delgado, J.A. Carino, J.A. Ortega, G. Clerc, Industrial Time Series Modelling by Means of the Neo-Fuzzy Neuron, volume 4 of *IEEE Access*, 2016 6151–6160. <https://doi.org/10.1109/ACCESS.2016.2611649>.
- [29] Y. Bodyanskiy, I. Kokshenev, V. Kolodyazhniy, An adaptive learning algorithm for a neo-fuzzy neuron, in: *Proceedings of the 3rd Conference of the European Society for Fuzzy Logic and Technology*, Zittau, Germany, 2003.
- [30] Y. Bodyanskiy, N. Kulishova, O. Chala, The Extended Multidimensional Neo-Fuzzy System and Its Fast Learning in Pattern Recognition Tasks, 3 volume of *Data*, 2018. <https://doi.org/10.3390/data3040063>.
- [31] R. P. Landim, B. Rodrigues, S. R. Silva, and W. M. Caminhas, "A neo-fuzzy-neuron with real time training applied to flux observer for an induction motor," in *Proceedings 5th Brazilian Symposium on Neural Networks (Cat. No.98EX209)*, Belo Horizonte, Brazil, 1998, pp. 67–72. doi: 10.1109/SBRN.1998.730996.
- [32] E. Parzen, "On Estimation of a Probability Density Function and Mode," *Ann. Math. Statist.*, vol. 33, no. 3, pp. 1065–1076, Sep. 1962, doi: 10.1214/aoms/1177704472.
- [33] G. S. Watson, "Smooth Regression Analysis," *Sankhya: The Indian Journal of Statistics, Series A*, Vol. 26, No. 4, 1964, pp. 359-372.
- [34] V. A. Epanechnikov, "Non-Parametric Estimation of a Multivariate Probability Density," *Theory Probab. Appl.*, vol. 14, no. 1, pp. 153–158, Jan. 1969, doi: 10.1137/1114019.
- [35] S. Kaczmarz, "Approximate solution of systems of linear equations," *International Journal of Control*, vol. 57, no. 6, pp. 1269–1271, Jun. 1993, doi: 10.1080/00207179308934446.
- [36] B. Widrow and M. E. Hoff, *Adaptive Switching Circuits*. IRE WESCON Convention Record, 1960.
- [37] G. C. Goodwin, P. J. Ramadge, and P. E. Caines, "A globally convergent adaptive predictor," *Automatica*, vol. 17, no. 1, pp. 135–140, Jan. 1981, doi: 10.1016/0005-1098(81)90089-3.
- [38] Y. Bodyanskiy, V. Kolodyazhniy, and A. Stephan, "An Adaptive Learning Algorithm for a Neuro-fuzzy Network," *Computational Intelligence. Theory and Applications. Fuzzy Days*, vol. 2206, pp. 68–75, 2001.
- [39] S. Debernardi et al., "A combination of urinary biomarker panel and PancRISK score for earlier detection of pancreatic cancer: A case-control study," *PLoS Med*, vol. 17, no. 12, p. e1003489, Dec. 2020, doi: 10.1371/journal.pmed.1003489.