Research article

Modified generalized neo-fuzzy system with combined online fast learning in medical diagnostic task for situations of information deficit

Yevgeniy Bodyanskiy[1], Olha Chala[2,*], Natalia Kasatkina[3] and Iryna Pliss[1]

[1] Control systems research laboratory, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

[2] Artificial intelligence department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

[3] Division of doctoral and post-graduate, National University of Food Technology, Kyiv, Ukraine

* **Correspondence:** Email: olha.chala@nure.ua; Tel: +380675460394.

Abstract: In the paper, we propose the modified generalized neo-fuzzy system. It is designed to solve the pattern-image recognition task by working with data that are fed to the system in the image form. The neo-fuzzy system can work with small training datasets, where classes can overlap in a features space. The core of the system under consideration is a modification of multidimensional generalized neuro-fuzzy neuron with an additional softmax activation function in the output layer instead of the defuzzification layer and quartic-kernel functions as membership ones. The learning procedure of the system combined cross-entropy criterion optimization using a matrix version of the optimal by speed Kaczmarz-Widrow-Hoff algorithm with the additional filtering (smoothing) properties. In comparison to the well-known systems, the modified neo-fuzzy one provides both numerical and computational implementation simplicity. The computational experiments have proved the effectiveness of the modified generalized neo-fuzzy-neuron, including the situation with shot training datasets.

Keywords: medical diagnostic; neo-fuzzy neuron; online learning; small training set; overlapping classes; membership functions

1. Introduction

Nowadays, approaches that are based on computational intelligence [1–3] have become widespread due to their effectiveness in solving various tasks in many fields. One of them is the healthcare field, in particular medical diagnostics tasks and data processing [4–7]. In some cases, information can appear in different notations—the input can have image form, leading to the image recognition classification problem, and classes (of diagnosis, for example) can overlap in a multidimensional feature space. Also, in real-world tasks, the data can be incomplete, or even the length of datasets can be small, forming a prior information deficit problem.

The approaches that proved their efficiency, including universal approximating properties and high-quality classification, in this field, medical diagnostic tasks are artificial neural networks, including both shallow neural networks [8] and deep ones [9,10]. Even though these systems are good classifications, they are crisp and can work only under conditions when classes can be separated in the features space.

In the fuzzy classification class of tasks, the approaches that proved their effectiveness are neuro-fuzzy systems [3,11–13]. They are designed as hybrid systems that combine the artificial neural network properties and fuzzy inference systems. Nevertheless, they require long training datasets to train the system correctly, and as was mentioned earlier, plenty of prior information is not always available during the diagnostic task solving.

Under conditions of a limited number of prior information, the probabilistic neural networks (PNNs) [14,15] can be used instead of multilayer neural networks and have shown good results in solving medical diagnostic tasks under described conditions. These networks tune according to the paradigm of lazy learning [16], based on the principle "Neurons at data points" [17] that means in the pattern layer of probabilistic neural networks, coordinates of centers of kernel activation functions coincide with coordinates of the observations from the training dataset. According to the classical architecture, these systems have to use all prior information from the training dataset, which is the main drawback of the system: when a new observation from the training dataset feeds into the system, a new neuron in the pattern layer is formed, restructuring the system. Of course, taking into account all useful information is a necessary step; however, this process is time-consuming and leads to the well-known problem Curse of dimensionality.

Considering situations of limited prior information in the training dataset and processing the data in the learning mode, the neo-fuzzy systems [18–21] have shown their effectiveness because they have good approximating properties, thus the learning procedure can be performed in real-time with a high learning speed [22].

Despite significant advantages, the recognition system, which is based on the neo-fuzzy neuron, cannot work under conditions of multiclass classification [23]. Also, the system contains too many activation functions and parameters, making it cumbersome for computational implementation.

To overcome this problem, we can use the idea of a generalized neo-fuzzy neuron [24]. In the general case, it is a multidimensional approximating system, however, it is not designed to solve classification problems, we propose to introduce the generalized neo-fuzzy neuron modification to solve pattern recognition-classification problem under conditions of a limited dataset and overlapping classes with high classification accuracy and speed.

2. Materials and methods

2.1. Architecture of modified neo-fuzzy-neuron for fuzzy classification

The architecture of the system for pattern recognition under the conditions of overlapping classes is shown in Figure 1, and contains four layers of information processing: the input (zero) layer receives a vector of features $x(k) = \left(x_1(k), ..., x_i(k), ..., x_n(k) \right)^T \in \mathrm{R}^n$ (here— $k = 1, 2, ..., N$ or the observation number in the training sample or the time of the discrete-time in the case when the information comes in a sequential online mode).
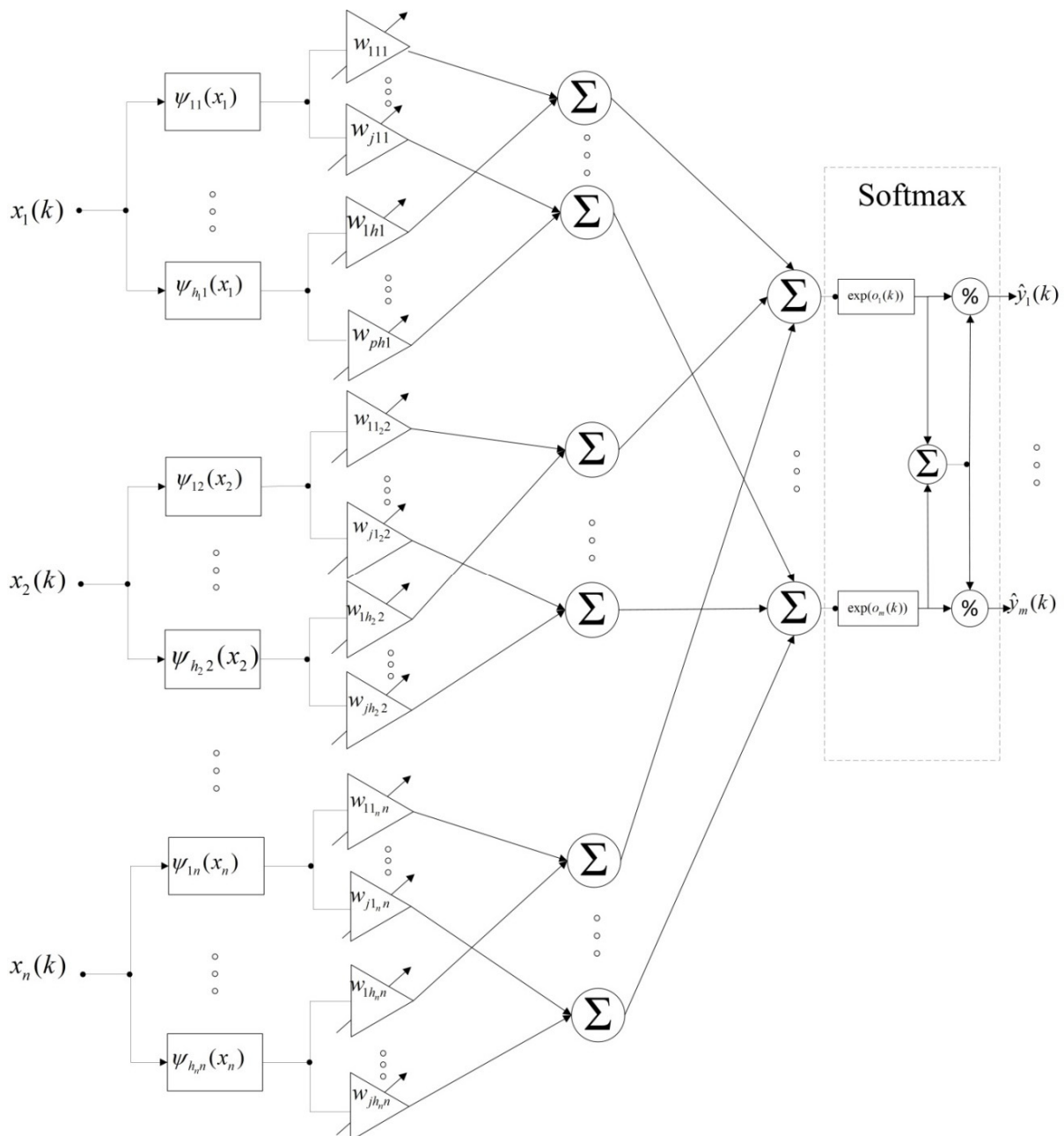
Figure 1. Architecture of modified generalized neo-fuzzy neuron for pattern recognition task.

This vector feeds into the first fuzzification layer, formed by $h = \sum_{i=1}^{n} h_i$ membership functions $\psi_{li}(x_i)$, as in traditional neuro-fuzzy systems, such as ANFIS or the Takagi-Sugeno-Kang system. Fuzzified signals are fed to the second layer of tuned synaptic weights w_{jli}, the total number of which is *mh* parameters. The system is trained by adjusting these weights in the optimization process (either batch or recurrent in online mode) of adopted criterion for the goal function learning. The third layer is formed with $m(n + 1)$ of elementary adders, on the output of which vector signal $o(k) = \left(o_1(k),...,o_j(k),...,o_m(k)\right)^T \in \mathrm{R}^m$ feeds to the input (fourth layer), formed by softmax activation functions [11]. Then on the outputs of the system appears the vector signal $\hat{y}(k) = \left(\hat{y}_1(k),...,\hat{y}_j(k),...,\hat{y}_m(k)\right)^T \in \mathrm{R}^m$. It sets the level of fuzzy membership of the observation vector to all classes $Cl_j, j = 1, 2, ... m$ —in the form:

$$\hat{y}_j(k) = \text{softmax}\, o_j(k) = \exp(o_j(k))\left(\sum_{p=1}^{m}\exp(o_p(k))\right)^{-1}, \; j = 1, 2, ... m. \tag{1}$$

Thus, the system under consideration implements a nonlinear transformation

$$\hat{y}_j(k) = SM(k) \times \left(w(k-1)\psi\left(x(k)\right)\right) \tag{2}$$

where, $SM(k) = \left(\text{softmax}\, o_1(k),...,\text{softmax}\, o_j(k),...,\text{softmax}\, o_m(k)\right)^T$, $\times$ is the symbol of direct product, $w(k-1) - \left(m \times h\right)$ is the matrix of adjusted synaptic weights that were obtained from the training dataset's observations, $\psi(x(k)) = (\psi_{11}(x_1(k)), \; \psi_{21}(x_1(k)), \; ..., \; \psi_{h_1 1}(x_1(k)), \; ..., \; \psi_{li}(x_i(k)),$ $..., \; \psi_{h_n n}(x_n(k)))^T$.

Triangular kernel functions are usually used as membership functions in neo-fuzzy neurons and satisfy the Ruspini unity partition conditions. This leads to the fact that in the features space, the separating hypersurfaces are approximated by piecewise linear functions. However, this type of functions does not always provide the required accuracy of the approximation. At the same time, on each learning step, only two neighboring functions are "fired" at each input, and only *2n* is adjusted instead of $h = \sum_{i=1}^{n} h_i$ synaptic weights, which naturally speeds up the learning process and is undoubtedly, is an advantage. Also, it is possible to increase the approximation accuracy by using B-splines, which provide the unity partition conditions as well. However, the number of adjusted weights at each step increases significantly.

The necessity for meeting the Ruspini partitioning conditions is because both the neo-fuzzy neuron and the generalized neo-fuzzy neuron do not have an output defuzzification layer, which is

always present in traditional neuro-fuzzy systems.

In the system under consideration, softmax-activation functions form the output layer that not only solves the defuzzification problem by ensuring that the condition is met but also provides a solution to the problem of pattern recognition:

$$\sum_{j=1}^{m} \hat{y}_j(\tau) = 1 \tag{3}$$

In other words, unity partition conditions do not impose on membership functions.

Indeed, it is possible to take traditional Gaussians or Cauchyans as membership functions that are used in neuro-fuzzy systems. However, at the same time, it would be necessary to tune absolutely all synaptic weights on each step that sophisticates the implementation.

Thus, we propose to use the fourth-degree polynomials—so-called quartic-kernel functions as the membership functions, which are shown in Figure 2.
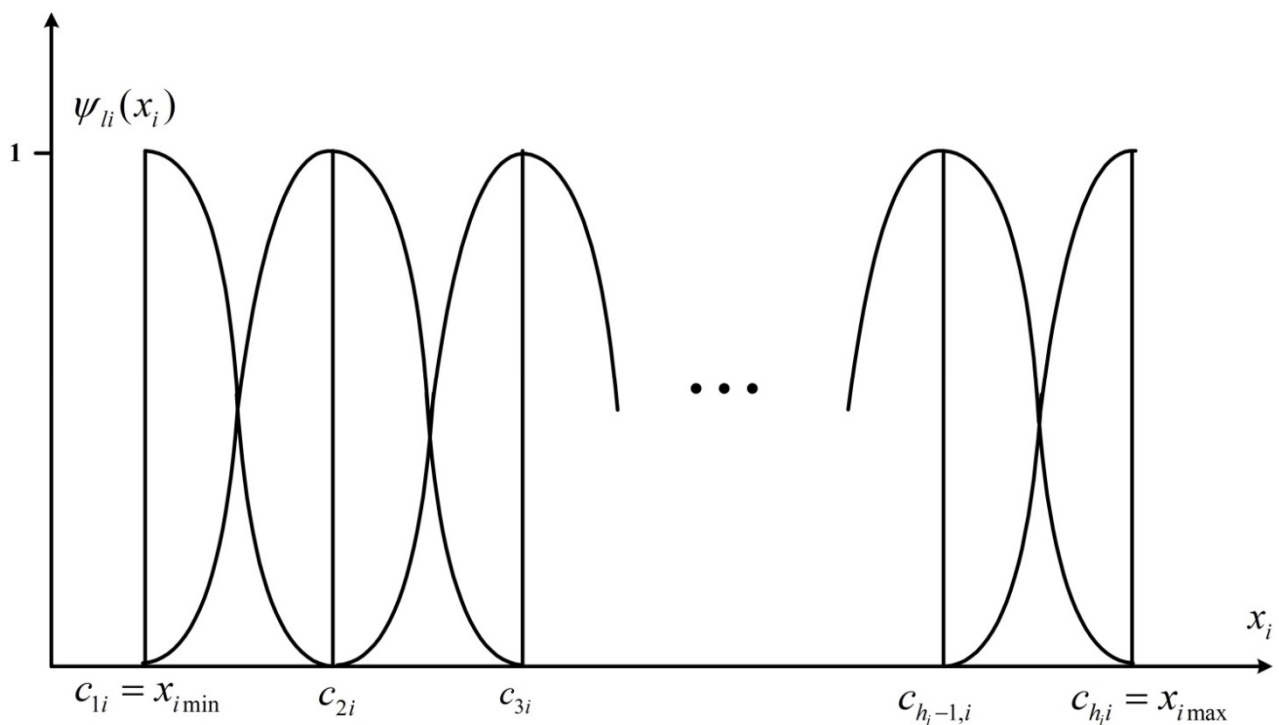
Figure 2. Quartic kernel functions as membership functions of modified neo-fuzzy-neuron.

Here, $\left[x_{i\,min}, x_{i\,max}\right]$ is the interval for changing the signal at the *i*-th input of the system,

$$r_i = \frac{x_{i\,min} - x_{i\,max}}{h_i - 1} \tag{4}$$

is the distance between two neighbouring centers of membership functions at *i*-th input,

$$y_{l_i}(x_i) = \left[\left(1 - \frac{(x_i - c_{li})^2}{r_i^2} \right)^2 \right]_+ \tag{5}$$

c_{li} is the parameters of the membership functions centers, $[\circ]_+ = \max\{0, \circ\}$ is the projector on positive ortant.

In the general case, they are arbitrarily placed on the absciss as shown in Figure 3.
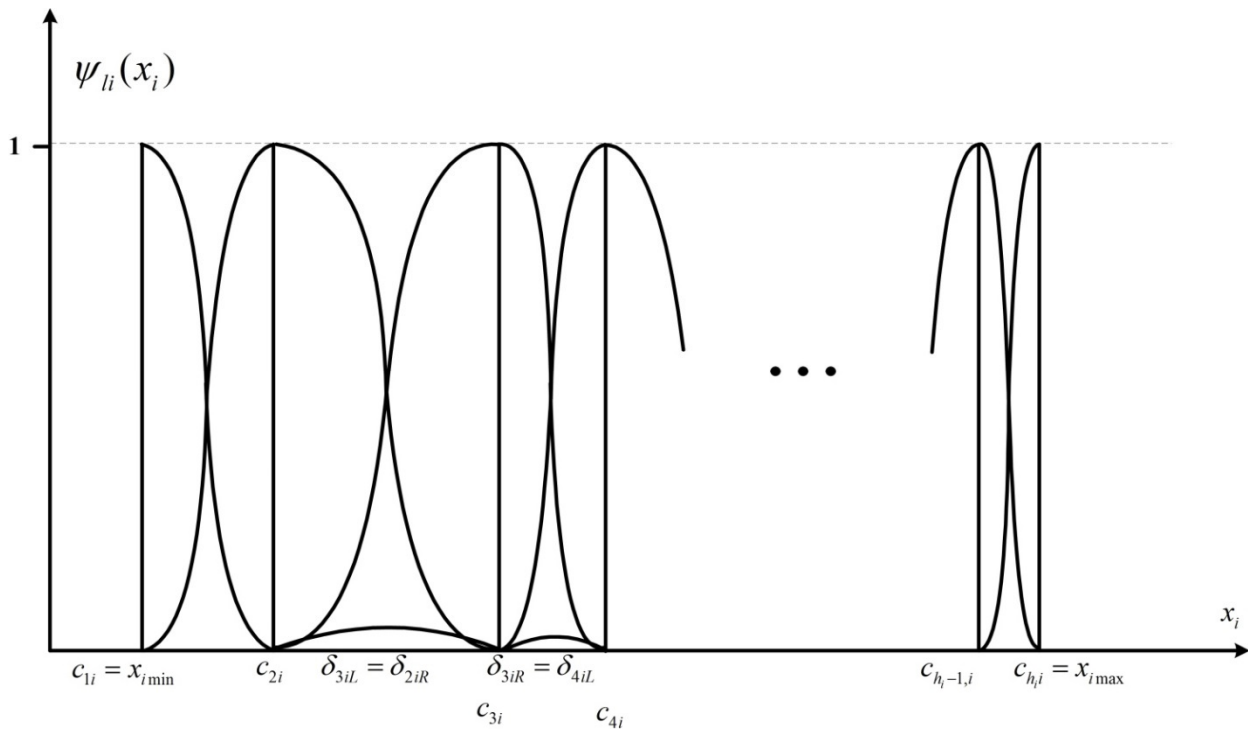
Figure 3. Unsymmetrical quartic kernel functions.

These functions in the analytical interpretation have a form

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

where, the indices L, R mean "left", "right".

It is easy to see that these functions are close to splines in terms of their precision properties. At the same time, on the inputs of the system, only two synaptic weights need to be tuned per learning step.

2.2. The leaning of modified generalized neo-fuzzy-neuron for fuzzy pattern recognition

The learning procedure of the system under consideration is implemented by optimizing the cross-entropy criterion. It nowadays is widely used in the problem of image-pattern recognition using deep convolutional neural networks [10]. This criterion has a form:

$$E(k) = \sum_{j=1}^{m} E_j(k) = -\sum_{j=1}^{m} y_j(k) \ln \hat{y}_j(k) = -\sum_{j=1}^{m} y_j(k) \ln \text{softmax } o_j(k) \tag{7}$$

here, $y_j(k)$ is where an external training signal that can take only two values

$$y_j(k) = \begin{cases} 1 & \text{if } x(k) \in Cl_j, \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

the so-called "one-hot coding".

Correspondingly, in the vector of these signals $y(k) = \left(y_1(k), ..., y_i(k), ..., y_m(k)\right)^T$, there is only 1 unit, and all other components are zero. The optimization process in the general case can be written in the form

$$w(k) = w(k-1) + \eta(k)(y(k) - w(k-1)\psi(x(k)))\psi^T(x(k)), \tag{9}$$

where, $\eta(k)$ is a learning rate.

The speed optimization of this procedure [25] leads to the algorithm:

$$
\begin{aligned}
w(k) &= w(k-1) + \frac{y(k) - w(k-1)\psi(x(k))}{\|\psi(x(k))\|^2} \psi^T(x(k)) = \\
&= w(k-1) + (y(k) - w(k-1)\psi(x(k)))\psi^+(x(k))
\end{aligned} \tag{10}
$$

where, $(\circ)^+$ is a symbol of pseudoinversion according to Moore-Penrose.

It is easy to see that the Kaczmarz-Widrow-Hoff algorithm [26,27] is a generalization of the matrix version in the artificial neural networks theory optimal algorithm for learning.

Filtering (smoothing) properties in the case of distorted perturbations of input observations can be added to the training procedure using a matrix modification of the so-called exponentially weighted stochastic approximation [28] in the form

$$
\begin{cases}
w(k) = w(k-1) + r^{-1}(k)(y(k) - w(k-1)\psi(x(k))\psi^T(x(k))), \\
\sigma(k) = \alpha\sigma(k-1) + \|\psi(x(k))\|^2, 0 \leq \alpha \leq 1
\end{cases} \tag{11}
$$

where, α is the forgetting factor. When $\alpha = 1$, the algorithm is converted into a matrix version of the

Goodwin-Ramage-Caines stochastic approximation algorithm [29] when $\alpha = 0$ the algorithm is converted into a matrix version of the Kaczmarz-Widrow-Hoff one.

When alpha is one, the algorithm transforms into a matrix version of the Goodwin-Ramage-Caines stochastic approximation algorithm [28]; when alpha is zero, the algorithm transforms into a matrix version of the Kaczmarz-Widrow-Hoff one.

Generally, not only the synaptic weights can be tuned, but also the location of the membership functions centers in the first layer. So, to improve the quality of learning, it is possible to use ideas such as lazy learning, which is used in Specht probabilistic neural networks, and Kohonen self-leaning [29] on the principle of the Winner takes all (WTA).

The tuning of the membership function can be described in the form of the following steps:
• Set the threshold of the inseparability of two neighboring centers.

$$\sigma_{i\,min} = \left| c_{li} - c_{l-1,i} \right|_{min}; \tag{12}$$

• feed the observation $x(1)$ at each input, form the first membership functions with centers:

$$c_{1i}(1) = x_i(1); \tag{13}$$

• feed the observation $x(2)$, and check the conditions:

$$\left| c_{1i}(1) - x_i(2) \right| \le \delta_{i\,min} \,\forall\, i = 1, 2, ..., n. \tag{14}$$

Form new centers $c_i(2)$ at these inputs where inequality is violated, and remains only where inequality satisfies;
• check the conditions:

$$c_{1i}(2) = c_{1i}\eta_1(2)\left(1 - \left(x_i(2) - c_{1i}(1)\right)\right) \tag{15}$$

and for those inputs where this inequality is true, tune the centers according to the WTA rule;
• check the conditions:

$$2\delta_{i\,min} \le \left| c_{1i}(1) - x_i(2) \right| \tag{16}$$

and for those inputs where it is done, form new centers according to the principle Neurons at data points:

$$c_{2i}(2) = x_i(2). \tag{17}$$

The process of forming new centers continues until it is formed

$$h_i = \frac{x_{i\,max} - x_{i\,min}}{\delta_{i\,min}} + 1 \tag{18}$$

membership functions at each input.

It is clear that as a result of such a lazy self-learning procedure, a different number of membership functions will be formed at each input. At the same time, their smallest number is two functions that are formed at the inputs that receive binary variables: "yes" or "no", "0" or "1", "there is a symptom" or "no symptom", and so on.

3. Results

3.1. Dataset overview

Pneumonia is a serious disease in which inflammation occurs in the lungs. Children are much more likely to develop severe complications than adults, and in general, this disease often proceeds with difficulty due to various reasons. Also, the pneumonia mortality rate is higher than HIV/AIDS, malaria, and measles combined [30]. According to the latest reports of the World Health Organization (WHO), it kills around 2 million children under five years old, making it the leading cause of the children mortality, most cases (approximately 95%) of which occur in Southeast Asia and Africa.

The leading causes of pneumonia are bacterial and viral. Both of them require different approaches to their management, but the key component in diagnostics of described cases is chest X-rays. However, obtaining radiographic data from a young patient is complicated compared to adults. In this case, it is hard to standardize the image to show the physiology of the patient on the X-ray. At the same time, the rapid interpretation of such results is rarely available in countries where such disease is progressing.

Data on medical image classification comes from the work [31] and contains pediatric X-ray images collected in a "Chest X-ray Images" dataset. It has two folders under the titles of two categories of chest X-ray images Normal and Pneumonia. In the second folder, two types of Pneumonia are specified by the file name: Bacteria and Viruses.

Table 1. The propositions of the "Chest X-ray Images" dataset.

|	Training set		Test set	
	Observation	Present	Observation	Present
Normal	1349	25.78	234	37.5
Bacteria	2538	48.51	242	38.78
Virus	1345	25.71	148	23.72
Total	5232	100	624	100

The total number of X-ray images in the training set is 5232, and the test set contains 624 observations and proportions of the train-test split considering the two types of pneumonia presented in Table 1.

Figure 4 shows observation from the dataset represents class Normal and shows clear lungs without any abnormal foggy areas.

Figure 5 shows observation from the dataset that represents classes "Bacterial pneumonia" (left) typically characterized by consolidation and air bronchogram sign, whereas "Viral pneumonia" (right) manifests with a Tree-in-bud sign and air tapping.
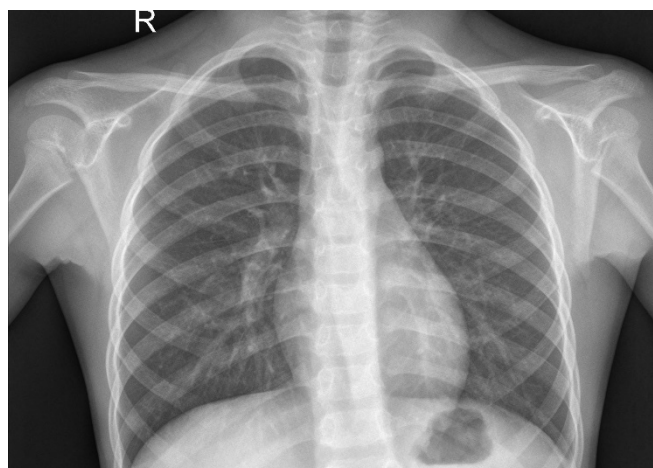
Figure 4. Example of chest X-ray from class Normal.
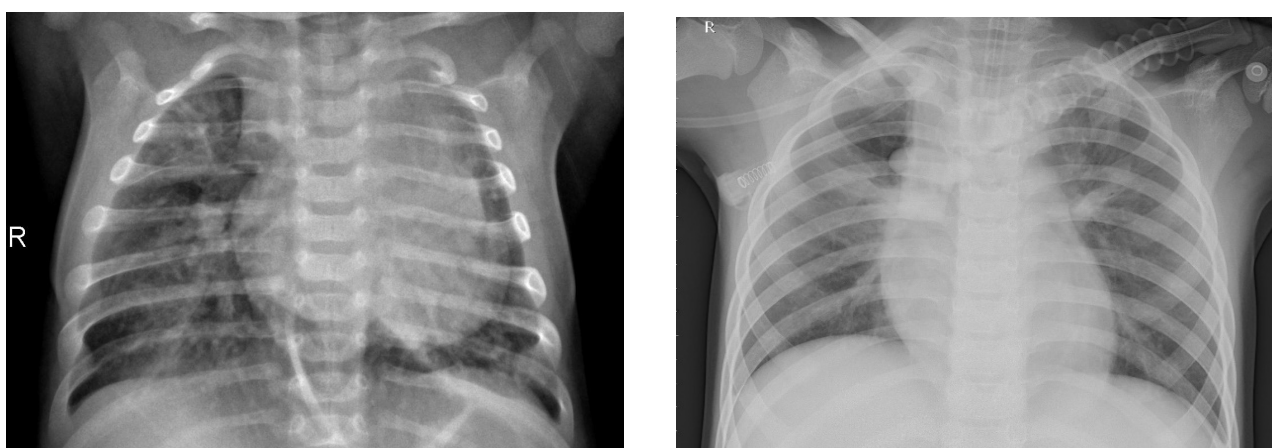
Figure 5. Example of chest X-ray from class "Bacterial pneumonia" (left), and class "Viral pneumonia" (right).

3.2. Models for comparison overview

Machine learning algorithms are widely used as classificators in the medical diagnostic field, for example, in reference [32], the support vector machines (SVM) using local binary patterns (LBPs) as features are classified the whole medical image, achieving state-of-art accuracy. In [33], the authors use not just linear but also RBF SVM for brain tumor detection, considering the situation where classes cannot be separated with a straight line, in other words, fuzzy case. Both SVMs showed great results, but the accuracy of RBF SVM is higher by around 4%. In [34], the polynomial SVM was used in the classification task of medical implant materials and showed high accuracy, however, the proposed method was time-consuming compared to the basic method.

The CNN-based deep neural system is popular in the medical classification task [35–37]. Among various architectures of CNN, ResNet characterizes by high accuracy and the ability to work with relatively small training sets. Also, according to the architecture, there is no need to fire all neurons in every epoch, improving accuracy and reducing learning time.

The distinctive feature of the Capsule Neural Networks [38–41] is equivariance, keeping the spatial relationship of objects in an image, and at the same time, the result does not impact the object's orientation and size. In [39], Capsule Neural Networks classified brain tumors on Magnetic Resonance Imaging (MRI) images and got 86.56% prediction accuracy. Moreover, the reference [38] presented a Capsule Neural Networks are based solution to classify four types of breast tissue biopsies from breast cancer histology images. They achieved 87% accuracy with the same high sensitivity.

So, as the baseline, for the traditional method Linear SVM, RBF SVM, polynomial SVM as the classifier, the ResNet and the Capsule Network with optimal configuration as CNN-based algorithms.

3.3. Environment setup

The hardware setup is the following: for SVM classification, an ordinary high-performance computer is enough, like 8G memory, i7 (2.70 GHz), and a 256G solid-state drive (SSD) disk. However, training both shallow and deep neural networks should use GPU to accelerate the process. In this paper, we used a Google Cloud GPU with standard configurations. Also, we used a virtual machine instance with one core of CPU, 16 G memory, and an Intel(R) Xeon(R) CPU@2.20GHz.

The software setup is the following: to test all three modifications of SVM classificators on the "Chest X-ray Images" dataset was developed with Python and ran on a laptop. The Capsule network and ResNet python implementations were run on a VM in Google GPU Cloud to test their performance.

3.4. Data pre-processing

This paper focuses on the image classification where observations are supposed to be typical, the same size, but, as can be seen from Figures 4 and 5, all images have different sizes. Also, the initial X-ray images have a different quality that typically depends on scattering radiations, receptor contrast, and contrasts themselves [42]. Thus, an image should be processed with Value Of Interest (VOI) transformation.

Some potential bias appears if we consider the variety of calibrations and range of dynamics. Hence, to make the image uniformly dynamic, histogram equalization was used. Then an image intensity was normalized in the range [0, 1] and standardized, so all observations in the dataset have the same size.

3.5. Experiment design and results

In the medical diagnostic field, it is hard to collect enough observations to form a long enough dataset for it to be used to train deep neural networks to obtain high classification accuracy. Hence, to analyze the ModGNFN performance on different size datasets, we split the original dataset "Chest X-ray" into subsets of the sizes 50, 100, 200, 400, 800 and 1600. The following experiment aims to process small datasets along with extremely small ones.

We observe all SVM models provide the lowest classification accuracy on the small number of samples. Methods this group poly SVM has an insignificant advantage. However, SVMs can train on a computer with a minimal configuration described in Section 3.3., a short training time, within 2 hours for the considered hardware configuration. It took more than 22 hours for the CapsNet to train,

which is the longest time. This network also requires the most memory for training. At the same time, the advantage of the proposed networks is at each moment k, only two neighboring functions are fired, so only two synaptic weights are tuned at each input. This property speeds up the learning procedure leading, so the time consumption of the system is less than 4 hours on the small datasets.

Table 2. The comparative analysis of the obtained results.

Methods	Chest X-ray50	Chest X-ray100	Chest X-ray200	Chest X-ray400	Chest X-ray800	Chest X-ray1600
RBF SVM	0.379	0.436	0.501	0.565	0.581	0.679
Linear SVM	0.345	0.408	0.449	0.513	0.623	0.698
Poly SVM	0.418	0.475	0.536	0.587	0.670	0.712
ResNet-18	0.485	0.617	0.701	0.842	0.816	0.853
ResNet-20	0.527	0.641	0.735	0.867	0.895	0.922
CapsNet	0.510	0.597	0.698	0.750	0.792	0.869
ModGNFN	0.625	0.702	0.776	0.824	0.840	0.856

Table 3. The comparative analysis of the obtained results with augmentation.

Methods	Chest X-ray50	Chest X-ray100	Chest X-ray200	Chest X-ray400	Chest X-ray800	Chest X-ray1600
RBF SVM	0.413	0.481	0.547	0.622	0.648	0.712
Linear SVM	0.372	0.454	0.536	0.570	0.654	0.703
Poly SVM	0.523	0.552	0.585	0.657	0.714	0.781
ResNet-18	0.592	0.664	0.752	0.846	0.875	0.908
ResNet-20	0.657	0.708	0.763	0.873	0.903	0.915
CapsNet	0.564	0.631	0.716	0.784	0.835	0.892
ModGNFN	0.710	0.748	0.825	0.839	0.846	0.885

However, on extremely small datasets, in terms of accuracy, the network yields both the proposed generalized neo-fuzzy-neuron system (ModGNFN) network and the ResNet network. Both networks show comparable recognition accuracy, but the proposed network has an advantage on very small (up to 400 elements) sets.

A comparison of Tables 2 and 3 shows that augmentation improves the accuracy of classification based on the increase in the amount of data online due to the transformation of the original image.

4. Discussion

Medical image is a unique field formed with tomography (ultrasound), radiographic projection (X-ray), and photography (dermatology) that contain elementary structures and show characteristics, distinct features, relative location of organs, and, in general, diagnostically relevant information [43]. Medical images consist of various geometrical forms shaping together an organ as well as its relative location and showing characteristics and distinct features. The main trait that helps recognize these patterns uses texture measurements, edge orientation, histograms of a hue, intensity and saturation.

Therefore, medical images contain overlapping patterns, representing classes that can overlap in features space.

The proposed diagnostic system is based on the modified generalized neo-fuzzy-neuron and focuses on highlighting such overlapping patterns. As inputs, the system uses a vector of features that is pre-processed image input, for example, with VGG16. In the first hidden layer, two quartic-kernel membership functions are fired, which allow filtering of the most vital features of patterns in the image. In this system, quartic-kernel functions allow narrowing down training time, speeding up processing procedure. The system includes two summation layers: the first one group together features of initial images with similar membership levels, which help to unite complex pattern elements that describe the shape and location of the organ. The obtained simple patterns can have different intensity levels on the initial image, and the aggregation of a group of such patterns allows to find a part in the image showing organ damage. Figure 5 shows an example of such a pattern. The second hidden layer is formed with adders where obtained outputs from the previous layer are grouped in a united pattern of features that provide further diagnostics. Then, in the output layer, obtained results from the second layer are deffuzified, computing the membership levels of an image to each class. This architecture combines quartic-kernel membership functions, and summation of obtained features allows getting base patterns and classifying them under conditions of the small and extremely small dataset with an accuracy of 0.776. The ensemble approach in ResNet showed its advantages with a further dataset increase from 400 elements. The system under consideration, on long datasets, obtains classification accuracy from 0.895 and more. Thus, the proposed modified generalized neo-fuzzy-neuron is designed for fast classification of medical images on small and extremely small datasets that allows using it in online mode.

The further development of the considered approach can be implemented by combining the classification and forecasting, particularly for the epidemic disease (such as COVID pandemic) in its early stages as well as its duration using the GROOMS [44] methodology.

5. Conclusions

In the paper, we considered the medical diagnostic problem under the conditions of overlapping classes and a small training dataset. We introduced the modified neo-fuzzy neuron. Its distinctive feature is an additional softmax layer, which allows simultaneously solving classification problems and implementing defuzzification of output signals. The usage of quartic-kernel functions as a membership allows the function to reduce the number of adjusted parameters at each learning stage. To optimize the algorithm by speed is to narrow down the number of training samples, which is very important under conditions of prior information that happens quite often in the medical field. The proposed system characterizes by the high accuracy and speed of simple numerical implementation, which the experiment results with real medical data confirms.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. C. L. Mumford, L. C. Jain, *Computational Intelligence*, Springer, Berlin Heidelberg, 2009.
2. R. Kruse, C. Borgelt, C. Braune, S. Mostaghim, M. Steinbrecher, *Computational Intelligence, A Methodological Introduction*, Springer-Verlag, Berlin, 2016.
3. J. Kacprzyk, W. Pedrycz, *Springer Handbook of Computational Intelligence*, Springer Verlag, Berlin Heidelberg, 2015.
4. P. Berka, J. Rauch, D. A. Zighed, *Data Mining and Medical Knowledge Management: Cases and Applications*, IGI Global, 2009.
5. R. Kountchev, B. Iantovics, *Advances in Intelligent Analysis of Medical Data and Decision Support Systems*, Springer International Publishing, Heidelberg, 2013.
6. Y. Bodyanskiy, I. Perova, O. Vynokurova, I. Izonin, Adaptive wavelet diagnostic neuro-fuzzy network for biomedical tasks, *14th International Conference on Advanced Trends in Radioelecrtronics, Telecommunications and Computer Engineering (TCSET)*, (2018), 711–715. https://doi.org/10.1109/TCSET.2018.8336299
7. Y. Syerov, N. Shakhovska, S. Fedushko, Method of the data adequacy determination of personal medical profiles, in (Eds.), *Advances in Artificial Systems for Medicine and Education II* (eds. Z. Hu, S. V. Petoukhov, M. He), Springer International Publishing, Cham, (2018), 333–343.
8. K. L. Du, M. N. S. Swamy, *Neural Networks and Statistical Learning*, Springer, London, 2013.
9. J. Schmidhuber, Deep learning in neural networks: An overview, *Neural Networks,* **61** (2015), 85–117. https://doi.org/10.1016/j.neunet.2014.09.003
10. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, United States, 2016.
11. P. V. C. Souza, Fuzzy neural networks and neuro-fuzzy networks: A review the main techniques and applications used in the literature, *Appl. Soft Comput.,* **92** (2020), 106275.
12. D. Graupe, *Deep Learning Neural Networks: Design and Case Studies*, World Scientific, New Jersey, 2016.
13. R. Tkachenko, I. Izonin, P. Tkachenko, Neuro-Fuzzy Diagnostics Systems Based on SGTM Neural-Like Structure and T-Controller, in *Lecture Notes in Computational Intelligence and Decision Making* (eds. S. Babichev, V. Lytvynenko), Springer International Publishing, Cham, (2021), 685–695. https://doi.org/10.1007/978-3-030-82014-5_47
14. D. F. Specht, Probabilistic neural networks, *Neural Network*, **3** (1990), 109–118. https://doi.org/10.1016/0893-6080(90)90049-Q
15. D. F. Specht, Probabilistic neural networks and polynomial ADALINE as complementary techniques to classification, *IEEE Trans. Neural Networks*, **1** (1990), 111–121. https://doi.org/10.1109/72.80210
16. O. Nelles, *Nonlinear Systems Identification*, Springer, Berlin, 2002.
17. D. R. Zahirniak, R. Chapman, S. K. Rogers, B. W. Suter, Pattern recognition using radial basis function network, *Aerospace Appl. Artif. Intell.*, **1990** (1990), 249–260.
18. T. Yamakawa, 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 and Neural Networks*, (1992), 477–483.
19. E. Uchino, T. Yamakawa, Neo-fuzzy-neuron based new approach to system modeling, with application to actual system, in *Proceedings Sixth International Conference on Tools with Artificial Intelligence,* (1994), 564–570. https://doi.org/10.1109/TAI.1994.346442

20. T. Miki, T. Yamakawa, Analog implementation of neo-fuzzy neuron and its on-board learning, *Comput. Intell. Appl.*, **1999** (1999), 144–149.

21. D. Zurita, M. Delgado, J. A. Carino, J. A. Ortega, G. Clerc, Industrial time series modelling by means of the neo-fuzzy neuron, *IEEE Access*, **4** (2016), 6151–6160. https://doi.org/10.1109/ACCESS.2016.2611649

22. 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*, (2003), 375–379.

23. Y. Bodyanskiy, N. Kulishova, O. Chala, The extended multidimensional neo-fuzzy system and its fast learning in pattern recognition tasks, *Data*, **3** (2018), 63. https://doi.org/10.3390/data3040063

24. R. P. Landim, B. Rodrigues, S. R. Silva, 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)*, (1998), 67–72. https://doi.org/10.1109/SBRN.1998.730996

25. E. Parzen, On estimation of a probability density function and mode, *Ann. Math. Statist.*, **33** (1962), 1065–1076.

26. S. Kaczmarz, Approximate solution of systems of linear equations, *Int. J. Control*, 57 (1993), 1269–1271. https://doi.org/10.1080/00207179308934446

27. B. Widrow, M. E. Hoff, *Adaptive switching circuits*, Stanford University, Stanford Electronics Labs, California, 1960.

28. Y. Bodyanskiy, V. Kolodyazhniy, A. Stephan, An adaptive learning algorithm for a neuro-fuzzy network, in *Computational Intelligence Theory and Applications Fuzzy Days*, (2001), 68–75. https://doi.org/10.1007/3-540-45493-4_11

29. G. C. Goodwin, P. J. Ramadge, P. E. Caines, Discrete time stochastic adaptive control, *SIAM J. Control Optim.*, **19** (1981), 829–853. https://doi.org/10.1137/0319052

30. R. A. Adegbola, Childhood pneumonia as a global health priority and the strategic interest of the bill and melinda gates foundation, *Clin. Infect. Dis.*, **54** (2012), S89–S92. https://doi.org/10.1093/cid/cir1051

31. D. Kermany, K. Zhang, M. Goldbaum, Labeled optical coherence tomography (OCT) and chest X-ray images for classification, *Mendeley Data*, **2** (2018).

32. Z. Camlica, H. R. Tizhoosh, F. Khalvati, Medical image classification via SVM using LBP features from saliency-based folded data, in *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, (2015), 128–132.

33. M. A. Khan, N. A. Syed, Image processing techniques for automatic detection of tumor in human brain using SVM, *Int. J. Adv. Res. Comput. Commun. Eng.*, **4** (2015), 541–544.

34. I. Izonin, A. Trostianchyn, Z. Duriagina, R. Tkachenko, T. Tepla, N. Lotoshynska, The combined use of the wiener polynomial and SVM for material classification task in medical implants production, *Int. J. Intell. Syst. Appl.*, **10** (2018), 40–47. http://doi.org/10.5815/ijisa.2018.09.05

35. Q. Li, W. Cai, X. Wang, Y. Zhou, D. Feng, M. Chen, Medical image classification with convolutional neural network, in *2014 13th International Conference on Control Automation Robotics and Vision (ICARCV)*, (2014), 844–848.

36. X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R. M. Summers, ChestX-Ray8: Hospital-scale chest X-Ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017), 2097–2106. https://doi.org/10.1109/CVPR.2017.369

37. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, et al., ImageNet large scale visual recognition challenge, *Int. J. Comput. Vis.*, 115 (2015), 211–252. https://doi.org/10.1007/s11263-015-0816-y

38. T. Iesmantas, R. Alzbutas, Convolutional capsule network for classification of breast cancer histology images, in *Image Analysis and Recognition*, (2018), 853–860. https://doi.org/10.1007/978-3-319-93000-8_97

39. P. Afshar, A. Mohammadi, K. N. Plataniotis, Brain tumor type classification via capsule networks, in *2018 25th IEEE International Conference on Image Processing (ICIP)*, (2018), 3129–3133. https://doi.org/10.1109/ICIP.2018.8451379

40. A. Jiménez-Sánchez, S. Albarqouni, D. Mateus, Capsule networks against medical imaging data challenges, in *Intravascular Imaging and Computer Assisted Stenting and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis,* (2018), 150–160. https://doi.org/10.1007/978-3-030-01364-6_17

41. E. Xi, S. Bing, Y. Jin, Capsule network performance on complex data, preprint, arXiv:171203480.

42. M. Mahesh, The essential physics of medical imaging, *Med. Phys.*, **40** (2013), 077301. https://doi.org/10.1118/1.4811156

43. H. D. Tagare, C. C. Jaffe, J. Duncan, Medical image databases: A content-based retrieval approach, *J. Am. Med. Inf. Assoc.*, **4** (1997), 184–198. https://doi.org/10.1136/jamia.1997.0040184

44. S. J. Fong, G. Li, N. Dey, R. G. Crespo, E. Herrera-Viedma, Finding an accurate early forecasting model from small dataset: A case of 2019-nCoV Novel Coronavirus outbreak, preprint, arXiv:2003.10776.