Mathematical Biosciences and Engineering

MBE, 18(3): 2599–2613.
DOI: 10.3934/mbe.2021132
Received: 28 February 2021
Accepted: 11 March 2021
Published: 17 March 2021

http://www.aimspress.com/journal/MBE

Research article

Predictive modeling based on small data in clinical medicine: RBF-based additive input-doubling method

Ivan Izonin[1,*], Roman Tkachenko[2], Ivanna Dronyuk[3], Pavlo Tkachenko[4], Michal Gregus[5] and Mariia Rashkevych[6]

[1] Department of Artificial Intelligence, Lviv Polytechnic National University, Kniazia Romana str., 5, Lviv 79905, Ukraine
[2] Department of Publishing Information Technologies, Lviv Polytechnic National University, S. Bandera str, 12, Lviv 79013, Ukraine
[3] Department of Automated Control Systems, Lviv Polytechnic National University, S. Bandera str, 12, Lviv 79013, Ukraine
[4] Department of Information Technologies, IT STEP University, Zamarstynivska str., 83a, Lviv 79090, Ukraine
[5] Faculty of Management, Comenius University in Bratislava, Odbojárov 10, Bratislava, Slovak Republic
[6] Department of Information Systems and Technologies, Lviv Polytechnic National University, Horbachevskoho str., 18, Lviv 79044, Ukraine

* **Correspondence:** Email: ivanizonin@gmail.com; Tel: +380988889687.

Abstract: The paper considers the problem of handling short sets of medical data. Effectively solving this problem will provide the ability to solve numerous classification and regression tasks in case of limited data in health decision support systems. Many similar tasks arise in various fields of medicine. The authors improved the regression method of data analysis based on artificial neural networks by introducing additional elements into the formula for calculating the output signal of the existing RBF-based input-doubling method. This improvement provides averaging of the result, which is typical for ensemble methods, and allows compensating for the errors of different signs of the predicted values. These two advantages make it possible to significantly increase the accuracy of the methods of this class. It should be noted that the duration of the training algorithm of the advanced method remains the same as for existing method. Experimental modeling was performed using a real short medical data. The regression task in rheumatology was solved based on only 77 observations. The optimal parameters of the method, which provide the highest prediction accuracy based on MAE and RMSE, were selected experimentally. A comparison of its efficiency with other methods of this class has been performed. The highest accuracy of the proposed RBF-based additive input-doubling method among the considered ones is established. The method can be modified by using other nonlinear artificial intelligence tools to implement its training and application algorithms and such methods can be applied in various fields of medicine.

Keywords: small data approach; neural networks; RBF; input-doubling method; predictive modeling; medicine

1. Introduction

Modern development of digital technologies stimulates the development of various spheres of human activity, including medicine. In particular, the advent of many micro devices in phones or other small gadgets, biochips, and other specialized information collection and transmission tools have created potential opportunities for the development of real-time tracking systems designed to monitor a patient's health. If we talk about monitoring one person, then there is a small data approach [1]. Over time, such data are combined with others to form Big data [2], based on which experts conduct intellectual analysis to develop medical decision-making systems for support on diagnosis or treatment.

There are many other sources of small data [3,4]. In particular, in countries with underdeveloped medical and IT (Information Technologies) areas, various patient tests are recorded in small electronic documents of the doctor or in personal paper cards of patients. Grouping and combining such data into large datasets for further intellectual analysis requires considerable material, time and human resources [5]. Furthermore, there is no absolute confidence in the effectiveness of the analysis of such data. There are a number of problems regarding the accuracy and reliability of such a dataset, the absence of gaps, noise and anomalies in it, and so on. Therefore, there is a very important task to develop a methodology of predictive analytics based on limited data processing in various fields of medicine.

The task in such formulation is not at all simple. Existing artificial intelligence (AI) tools do not provide sufficient prediction accuracy in processing short data sets [6-8]. There are a number of reasons for this. The first is overfitting, which is very typical of methods in this class [9]. Another problem is the very large impact of anomalies and omissions in the already limited amount of available information intended for processing [10]. Furthermore, the use of artificial intelligence methods to process short data sets does not provide a sufficient level of generalization, which affects the prediction accuracy. In turn, this raises concerns about the reliability of the result and the possibility of some generalization of the method, in particular on other data. Despite this, such task does not disappear and its number grows every year, but the available methodology for solving it is quite limited.

In [11], the authors investigated the development of a small data paradigm and its benefits in medicine. This theoretical work presents a number of very strong arguments in favor of the development of this area. The authors divided them into two main groups: scientific and applied. Each of these reasons is very well argued, and supported by real cases from medicine. This work was the impetus for developing of a new methodology for processing short data sets using artificial intelligence tools.

If we talk about the practical aspect of the problem, then one of the newest works in the direction of artificial intelligence methods for processing short data sets is a book [9]. Here are a number of machine learning methods for intelligent analysis in the case of limited data sets. Among the shortcomings of the work should be noted the description of only existing methods, as well as their ineffectiveness in the case of processing very short sets of medical data (starting from 10 vectors).

The review paper [12] is devoted to the problem of applying machine learning methods in order to improve medical services by reducing the necessary costs [13]. The authors studied the existing neural network tools to solve this task. They selected more than 3000 articles from various scientometric databases and conducted a detailed analysis of the application of various artificial neural networks in medical decision-making processes. The analysis showed the widespread use of this artificial intelligence tool in solving various medical problems, particularly in the United States

The "Biomedical Data Mining Using RBF Neural Networks" chapter of the book [14] describes the use of the Radial Basis Function Artificial Neural Network (RBF ANN) to improve the accuracy of cancer diagnosis. This task has a great practical importance for both patients and physicians. RBF ANN showed high accuracy and, consequently, the possibility of its practical application in various medical decision support systems.

In [15], the authors investigated the possibility of using RBF ANN to predict the onset of Parkinson's disease tremors in humans. The main practical value of this task is the reduction of battery power used for the Deep Brain Stimulation (DBS). The use of RBF ANN in this case provides accurate recognition of tremors and determines when it is necessary to carry out electrical activity in target brain areas DBS. The forecasting accuracy was more than 80%. In order to reduce the computational costs for the RBF ANN application, the authors proposed an improved forecasting method based on the additional use of particle swarm optimization. According to experimental studies presented in this paper, this approach has paid off.

Another successful example of the use of RBF ANN to solve medical tasks is a study [16]. The authors developed an adaptive back stepping controller based on this neural network type for medical devices to treat heart rhythm disorders. The main purpose of such development is to increase the productivity of the latter. Since this is a question of human life, such developments must be reliable, stable and effective. The authors provided all these conditions using the Lyapunov function. Experimental results in four patients showed a very high accuracy of the prognosis. Nevertheless, the authors' development requires more extensive experimental research.

In [17,18], new kernel functions were developed to increase the accuracy of RBF ANN. In [17], the cosine measure was proposed and a significant increase in the accuracy of neural networks of this type based on three different data sets was demonstrated. In [18], the authors used weighted cosine distance and also showed a significant increase in the accuracy of solving a medical problem, namely the prediction of leukemia. These developments can serve as a basis for improving all the methods analyzed above in the case of processing both large and small data sets. However, only a small number of studies have been devoted to the problem of processing short data sets based on RBF ANN.

The authors of [19, 20] presented a new method of processing short data sets. It is based on the use of the classic RBF ANN. As the training algorithm of RBF neural network involves random initial initialization of the weights, and possible random selection of *rbf*-centers, the method demonstrates different results after each run. The idea of the method from [19, 20] is to repeatedly run a neural network of this type with the optimal parameters of its operation to select the most accurate result "method of multiple runs". In the process of obtaining it, it is possible to fix the weights of this artificial intelligence tool, which provides the possibility of reproducing the results of its work.

In [21], a new RBF-based input-doubling method is introduced. The main idea of the method is to extend the short data set and apply the author's prediction procedures using RBF ANN. This approach demonstrates the improvement of the prediction accuracy when processing short data sets with a satisfactory time of the training procedure (given that we are talking about small data). This work is a continuation of the research begun in [21]. The aim of this paper is to increase the prediction accuracy with satisfactory indicators of the duration of the training procedure of the RBF ANN.

Our main contribution in this paper can be summarized as follows:

1) We have improved the existing RBF-based input-doubling method by added additional elements in the formula for calculating the output signal of the method, which improves its prediction accuracy;

2) We have shown that the adding of additional elements in the formula for calculating the predicted value of the developed method allows to compensate for the errors of various signs of the predicted values, which improves the prediction accuracy of methods of this class;

3) We have established the highest prediction accuracy of the proposed RBF-based additive input-doubling method with pre-selected values of the optimal parameters of its operation compare to other methods of this class.

The remainder of this paper is organized as follows: the second section provides basic information about RBF ANN as the basis of the improved method, and describes the general approach, training and application procedures of the improved method. The third section of this paper describes the procedures for selecting the optimal parameters of the proposed RBF-based additive input-doubling method, the results of its work, as well as comparing its effectiveness with other methods in this class. Conclusions and prospects for further research are presented in the conclusion section.

2. Materials and methods

2.1. Radial basic neural network

In general, the classical interpretation of the RBF neural network involves the sequential connection of three layers, the first of which is the input, the weights of which fix the coordinates of the centers of radial functions. Most often, when choosing the coordinates of the centers use the principle of randomization, or they are placed in the centers of clusters [22,23]. In the latter case, Kohonen's self-organizing maps are sometimes used as the input layer [24]. The radial elements of the second layer reproduce Gaussian functions (rarely other variants of functions) from the corresponding scalar products of the weights on the components of the input vectors. Instead of scalar products, Euclidean distances from the coordinates of the input vectors to the coordinates of the RBF centers are often used [25]. The third layer of elements forms a given response surface by a combination (often linear) of the corresponding Gaussian functions (Figure 1). The theoretical substantiation of efficiency of application of neural networks of this type follows from conclusions of the well-known Cover's theorem about the increase of probability of linear separability of images by a transformation of a difficult problem of classification in the space of higher dimension [26].

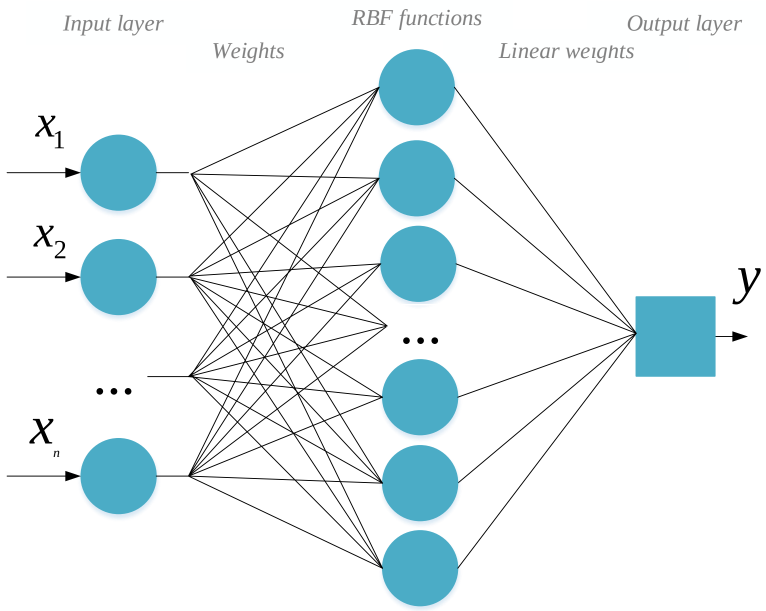


Figure 1. RBF topology.

Let consider the reflection of the hidden layer. Each hidden item is associated with a function φ. Each of these functions has a combined input and generates the value of the activity supplied to the output. The set of activity values of all hidden elements determines the vector on which the input vector is displayed: $\phi(x)=[\phi(x_1),]\phi(x_2),...,\phi(x_M)$, where M is a the number of hidden elements, and x is input vector.

The connections of the hidden layer element determine the center of the radial function for the hidden element. The input for each element is equal to the Euclidean norm: $net_j=\|X-W_j\|=\left[\sum_{i=1}^{n}(x_i-w_{ij})^2\right]$, where n is the number of input elements.

A variety of activation functions, such as the Gaussian function, is used for hidden elements $\phi(r)\approx\exp\left[-r^2\right]$ or function $\phi(r)\approx\sqrt{c^2+r^2}$.

Here are the main tasks that need to be solved in the process of setting up the RBF ANN [21]:

– to choose of the method of forming the coordinates of the *rbf*-centers and the type of basic functions;

– to set the number of *rbf*-centers and the parameter of basic functions, which specifies the magnitude of their scope (smooth factor);

– to train the output ANN layer.

In practice, two main methods are used to specify the coordinates of the centers: the choice as the centers of coordinates of the inputs of the training sample obtained randomly (from the general training sample), or the location of the centers of basic functions in the centers of clusters [27]. To implement the latter option as a shaper of the input signals of radial functions in this paper, the *kNN* algorithm is used. Details of the practical implementation of RBF ANN, which is used in this paper, algorithms for its training and application are given in [28].

2.2. RBF-based additive input-doubling method

Let the sample of independent variables be presented in the form of a matrix with dimension $n \times m$ in the form:

$$A = \left[a_{i,j} \right], \tag{1}$$

where $i = \overline{1,n}; j = \overline{1,m}$, n is the number of observations, m the number of independent variables of each observation.

Given the known output y_i of each of n observations, we can form a new matrix $\bar{A}$ whose dimension is $n \times (m+1)$ in the next form:

$$\bar{A} = \left[a_{i,j}; y_i \right]. \tag{2}$$

Algorithmic implementation of the training procedure of the RBF-based additive input-doubling method involves the following steps:

For each of the i^**-th* observations, where $i^* = \overline{1,n}$, we form a matrix B_{i^*} whose dimension is $n \times 2m$ in the next form:

$$B_{i^*} = \left[a_{i^*,j} \; a_{k,l} \right], \tag{3}$$

where i^* is a fixed value, $j = \overline{1,m}; k = \overline{1,n}; l = \overline{1,m}$.

The output for Eq (3) is formed as the difference of outputs between each new current point y_{i^*} and all other points of the training data set y_k in Eq (3):

$$z_{i^*} = (y_{i^*} - y_k). \tag{4}$$

Taking into account the values of outputs calculated based on Eq (4), we form a matrix $\bar{B}_{i^*}$ whose dimension is $n \times (2m+1)$:

$$\bar{B}_{i^*} = \left[a_{i^*,j} \; a_{k,l}; z_{i^*} \right]. \tag{5}$$

The final extended matrix of the training data whose dimension is $n^2 \times (2m+1)$ is formed by the following way:

$$\bar{B} = \begin{bmatrix} \bar{B}_1 \\ \bar{B}_2 \\ \dots \\ \bar{B}_n \end{bmatrix}. \tag{6}$$

Matrix Eq (6) is the basis of the training procedure of the RBF ANN.

Application mode of the improved method.

Let's consider the vector u_{p^0} the dimension of which is *m*, where p^0 is fixed, for which (vector) the value of the output signal is unknown. Based on it, in contrast to the existing method [21], we form two matrices C_1 i C_2 whose dimension is $n \times 2m$:

$$C_1 = \left[u_{p^0}\; a_i \right], \tag{7}$$

$$C_2 = \left[a_i\; u_{p^0} \right], \tag{8}$$

where a_i - all vectors from the initial training data set from Eq (1), $i = \overline{1, n}$

Based on matrices Eqs (7) and (8), taking into account the unknown value of the outputs $t_i^{(1)}, t_i^{(2)}$, we obtain extended matrices that will contain a set of independent attributes with the corresponding dependent attribute:

$$\bar{C}_1 = \left[u_{p^0}\; a_i ; t_i^{(1)} \right], \tag{9}$$

$$\bar{C}_2 = \left[a_i\; u_{p^0} ; t_i^{(2)} \right], \tag{10}$$

Now it is possible to build the response surfaces $t_i^{(1)}, t_i^{(2)}$ for $\bar{C}_1, \bar{C}_2$ by applying pre-trained RBF ANN.

The final required signal y of the improved method, in contrast to the basic method [21], is calculated according to the expression:

$$y^{pred} \cong \frac{\sum_{i=1}^{n} y_i}{n} + \frac{\sum_{i=1}^{n} t_i^{(1)}}{2n} - \frac{\sum_{i=1}^{n} t_i^{(2)}}{2n}. \tag{11}$$

It should be noted that in Eq (11), based on $\frac{\sum_{i=1}^{n} t_i^{(1)}}{2n} - \frac{\sum_{i=1}^{n} t_i^{(2)}}{2n}$ the principle of averaging, which is characteristic of the work of ensemble methods, is implemented. Furthermore, the use of both these terms is in contrast to the existing method, which uses only $\frac{\sum_{i=1}^{n} t_i^{(1)}}{2n}$ in Eq (11), provides partial mutual compensation of random errors of different signs in equality Eq (11). These two circumnutates significantly increase the accuracy of the output signal based on the improved method.

3. Results, comparison and discussion

The section provides information about the dataset used for modeling the improved method, a description of the procedures for the selection of the optimal parameters of its work, the results and a comparison with other methods of this class.

3.1.1. Dataset description

A short dataset in one of the branches of medicine was chosen to model the work of the improved method. The authors in [19] used two datasets to predict the compressive strength of trabecular bone

(CS in MPa). One of these datasets is real, and the other one is artificial. Each of them contains 77 observations. The actual dataset is compiled based on a review of patients with osteoarthritis. The practical value of this task is the need to predict the strength of the load on the bone when performing exercise in the elderly suffering from osteoarthritis. The purpose of accurate prognosis is to avoid physical injuries from overload, which occur quite often during this disease in the elderly.

The list of attributes and their main characteristics are given in Table 1.

Table 1. Dataset.

Characteristic	Maximum value	Average value	Minimum value
Patients' age	87,00	64,80	41,70
Patients' gender	1,00	-	0,00
Tissue porosity (BV/TV),	43,50	26,22	2,10
Trabecular thickness factor (tb.th)	419,00	259,69	154,00
Structure model index (SMI)	2,10	0,65	0,04
CS (in MPa)	28,80	16,40	1,90

During modeling, the real dataset was randomly divided into three data sets: training, validation and test. The ratio of this division: 70, 10 and 20%, respectively.

3.1.2. The optimal value of the number of clusters of the algorithm *k-nn*

One of the important steps in applying of RBF ANN is the choice of the number of clusters of the *kNN* algorithm. Effective selection of this parameter significantly affects the accuracy of the method. We used Mean Absolute Error Eq (12) and Root Mean Square Error Eq (13) as performance indicators.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_i^{pred})^2 . \qquad (12)$$

$$MAE = \sqrt{\sum_{i=1}^{n}\frac{(y_i - y_i^{pred})^2}{n}} . \qquad (13)$$

The experiment to select the optimal value of the investigated parameter based on Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) was as follows. Both existing and improved methods were run 100 times when changing the number of clusters of the *kNN* algorithm from 15 to 100. Other parameters of the method were the same (number of epochs RBF ANN-200 [21], smooth factor equal 0.4., Mean Square Error (MSE) loss function, 100 iterations for k-nn method). It should be noted that the RBF centers at each run were the same for both methods.

From the obtained results, the maximum value, the minimum value and the average value of MAE and RMSE errors were selected. The results of this experiment are summarized in Figure 2.

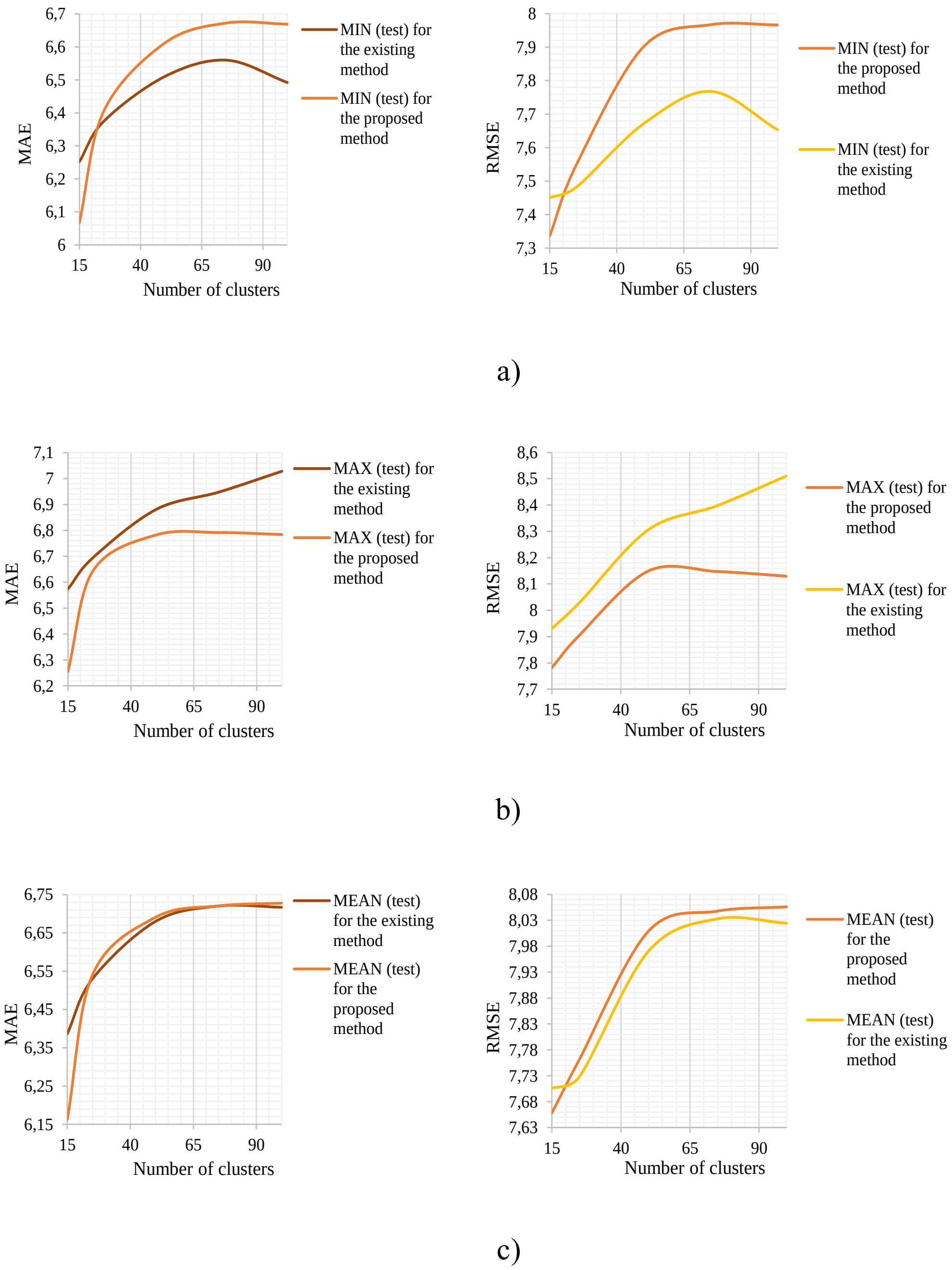


Figure 2. MAE and RMSE values for the test mode of existing and designed methods when changing the number of clusters [15, 100] of *k-nn* methods for 100 runs of both methods: a) minimal values; b) maximum values; c) average values.

From all graphs in Figure 2 it is clear that:

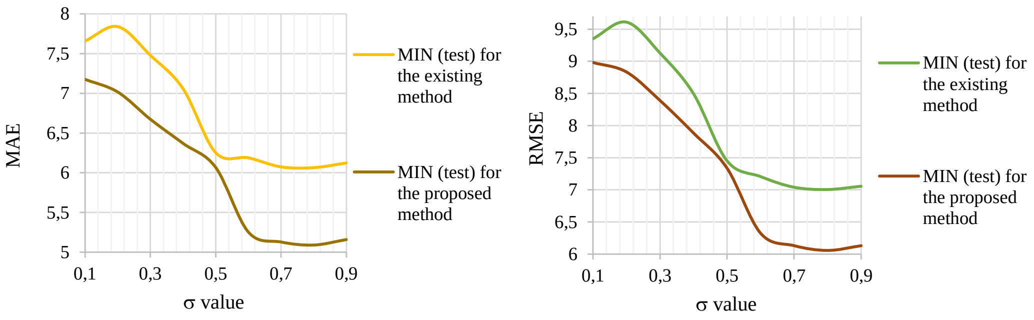
- as the number of clusters of the kNN algorithm increases, the errors of MAE and RMSE increase for both methods;
- the increasing of MAE and RMSE errors of the improved method with increasing the number of clusters k of the kNN algorithm is significantly higher in comparison with the existing method (except for small values of k);
- the lowest value of the error of the improved method was obtained for 15 clusters in all three

cases (minimum, maximum and average value);

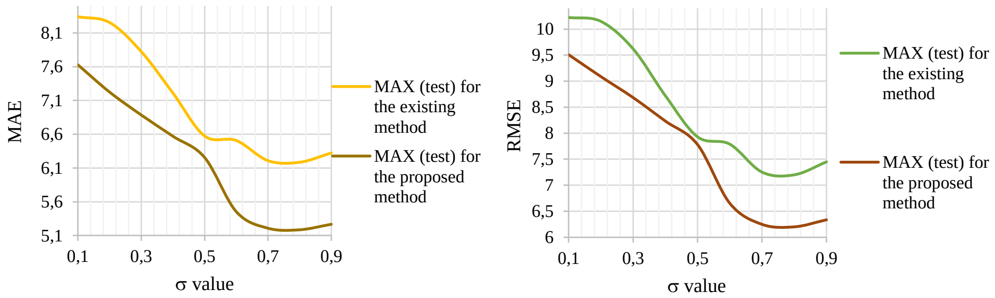
- the errors of the existing method at this value of the studied parameter (k = 15) are higher in comparison with the improved method both in the case of maximum and minimum values and in the case of the average value at 100 runs of both methods.

Taking all this into account, the value of the number of clusters of the kNN algorithm equal to 15.

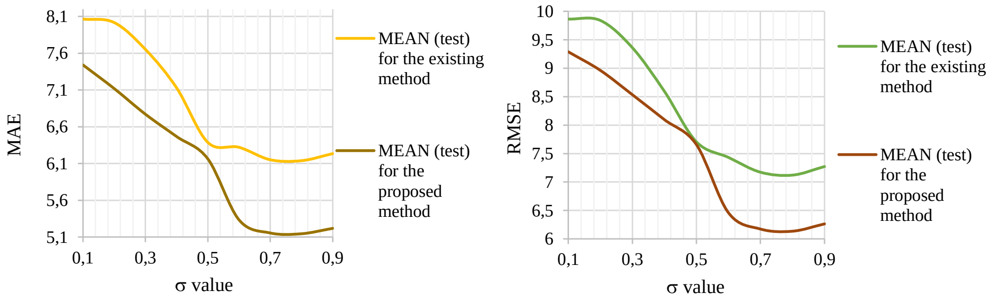
3.1.3. Selection of the optimal value of the smooth factor for the RBF ANN



a)



b)



c)

Figure 3. MAE and RMSE values for the test mode of existing and designed methods when changing smooth factor σ for 10 runs of both methods: a) minimal values; b) maximum values; c) average values.

One of the important parameters of the classical iterative RBF ANN, which is the basis of the existing and improved methods, is the smooth factor. The value of this parameter significantly affects the operation of the neural network of this type and, as a consequence - the operation of existing and improved methods.

In this work, a study was conducted on the selection of the optimal value of the smooth factor (σ). The experiment was as follows. Both existing and improved methods were run 100 times (one experiment) and for each experiment, we have changed the smooth factor values from 0.1 to 0.9 with a step of 0.1. Other parameters of the method were the same (number of epochs RBF ANN-200 [21], MSE loss function, 15 clusters, 100 iterations for *k-nn* method).

From the obtained results, the maximum value, the minimum value and the average value of MAE and RMSE errors for 100 runs of the method were selected. The results of this experiment are summarized in Figure 3.

All graphs from Figure 2 clearly show the following:

- the local maximum at small values of σ, ($\sigma < 0.2$) for both studied methods indicates the need to choose much larger values of this parameter for the practical implementation of both methods;

- as the value of the smooth factor increased, the errors of MAE and RMSE decreased for both methods;

- at a value of $\sigma = 0.5$, both methods show very similar results, although the improved method shows a smaller value of both errors;

- in the interval $0.7 < \sigma < 0.9$, both errors for both studied methods acquire saturation stages; this statement is true for the minimum, maximum and average values among 100 runs of both methods;

- in the interval $0.7 < \sigma < 0.9$, the improved method demonstrates a significant increase in the prediction accuracy compared to the existing one;

- the smallest value of errors for the improved method is received at the value of smooth factor $\sigma = 0.8$ both in case of the maximum and minimum values and in case of average value at its 100 runs in comparison with the existing method.

Therefore, the value of the smooth factor $\sigma = 0.8$ was used for further research.

3.1.4. Results

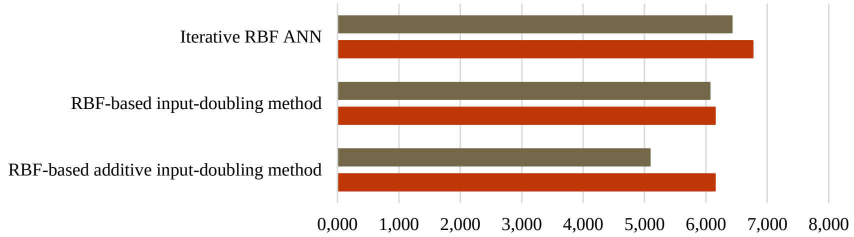
Based on the selected optimal parameters of the improved method the following results which are summarized in Table 2 are received.

Table 2. Results of the RBF-based additive input-doubling method.

MAE	RMSE	Training time (seconds)
Train mode		
6.150	6.237	13.906
Test mode		
5.089	6.057	-

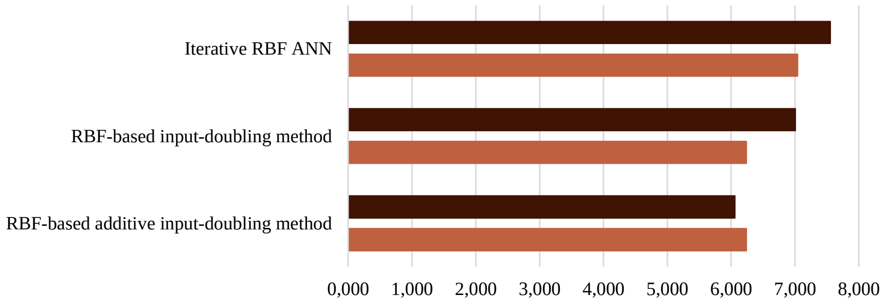
It should be noted that the minimum values obtained at 100 runs of the method are selected as the results of its work. This is justified by the possibility of fixing the synaptic weights, which provides all the necessary conditions for the full reproduction of the results of the method.

3.1.5. Comparison



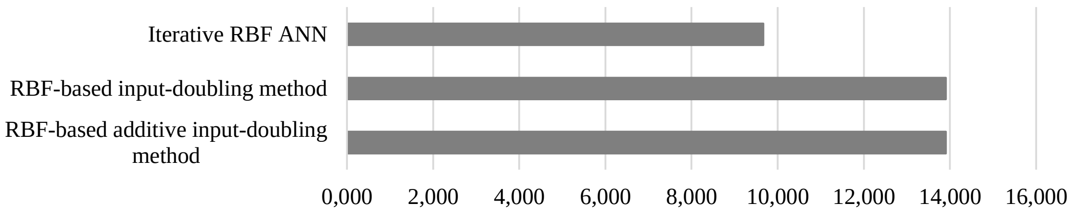
	RBF-based additive input-doubling method	RBF-based input-doubling method	Iterative RBF ANN
MAE (test)	5,089	6,064	6,425
MAE (train)	6,150	6,150	6,762

a)



	RBF-based additive input-doubling method	RBF-based input-doubling method	Iterative RBF ANN
RMSE (test)	6,057	7,005	7,552
RMSE (train)	6,237	6,237	7,039

b)



	RBF-based additive input-doubling method	RBF-based input-doubling method	Iterative RBF ANN
Training time, seconds	13,906	13,906	9,669

c)

Figure 4. Performance indicators for train and test modes of all investigated methods: a) MAE values; b) RMSE values; c) training time (in seconds).

The comparison of the efficiency of the improved method was performed with the existing method, as well as the classical iterative RBF ANN, which is the basis of the work of both studied methods. The comparison was based on both errors as well as the duration of the training procedure. The results of this study are summarized in Figure 4.

As can be seen from Figure 4, the highest operating errors in both training and application modes were obtained using the classical iterative RBF ANN. More accurate results were obtained using the existing RBF-based input-doubling method. However, the duration of the training procedure is increased by 5 seconds, which is about 50% of the duration of the classical RBF ANN, which is the basis of its work. This is due to two reasons: 1) the quadratic increase of the training sample according to (3)-(5)-(6); 2) doubling the number of input features according to (3) and as a consequence of the expansion of the RBF ANN's input layer, which is the basis of this method.

If we analyze the effectiveness of the proposed method, we can say the following:

- the improved method provides a significant increase of the prediction accuracy based on both errors compared to the existing method;
- the training errors for both existing and improved methods will be the same (or very close) due to the same procedure for the formation of the training sample according to (3)-(5)-(6);
- the duration of the training procedure will also be the same (or very close to the duration of the training procedure of the existing method), as explained in the previous case;
- the application time of the improved method will be longer than the application time of the existing method, due to the additional procedures and as a consequence of changing the search formula for the output signal of the method. However, since these are small and very small data sets, this shortcoming is not significant.

4. Conclusions

The development of modern medicine is accompanied by a large amount of diverse information about the object of observation. There are many applications in the field of medicine that require the use of artificial intelligence tools to obtain a faster or more accurate solution for quality diagnosis or treatment. In some cases, limited data sets significantly reduce the efficiency or simply make it impossible to apply artificial intelligence tools for the intellectual analysis of such information. This paper considers the current problem of regression analysis via artificial neural networks based on short sets of medical data. The authors proposed an improved RBF-based input-doubling method by introducing additional elements into the formula for calculating the output signal of the method. This improves prediction accuracy while the duration of the training procedure remains unchanged. The effectiveness of the advanced method was tested using a real short set of medical data, which contained only 77 observations. It is experimentally established that the additional elements in the formula for calculating the predicted value of the developed method allows to compensate for the errors of different signs of the temporary predicted values, which improves the accuracy of methods of this class as a whole. The selection of the optimal parameters of the developed method is carried out. The highest prediction accuracy of the proposed RBF-based additive input-doubling method is established by comparing the efficiency of its work with other methods of this class. The accuracy of the improved method is increased by 1% based on both MAE and RMSE compared to the basic method. The training time (13.9 seconds) of both methods is the same. Among the disadvantages of the studied methods should be noted significant time delays of their work. That is why future studies will consider non-iterative variants of RBF ANN [29] and will model the operation of the method based on them. This should provide a significant reduction of the training time of the method while maintaining, and even increasing the accuracy of its work.

Acknowledgments

The National Research Foundation of Ukraine funds this study from the state budget of Ukraine within the project "Decision support system for modelling the spread of viral infections" (№ 2020.01 / 0025).

Conflict of interest

The authors declare that they have no conflict of interests.

References

1. D. I. Christine, M. Thinyane, Small Data approaches provide nuance and context to health datasets, Available from: http://theconversation.com/small-data-approaches-provide-nuance-and-context-to-health-datasets-121972.
2. N. Melnykova, N. Shakhovska, M. G. ml, V. Melnykov, Using big data for formalization the patient's personalized data, *Proc. Comp. Scien.*, **155** (2019), 624–629.
3. E. K. Wang, F. Wang, R. P. Sun, X. Liu, A new privacy attack network for remote sensing images classification with small training samples, *Math. Biosci. Eng.* **16** (2019), 4456–4476.
4. T. Mao, L. Yu, Y. Zhang, L. Zhou, Modified mahalanobis-taguchi system based on proper orthogonal decomposition for high-dimensional-small-sample-size data classification, *Math. Biosci. Eng.* **18** (2021), 426–444.
5. L. Mochurad, M. Yatskiv, Simulation of a human operator's response to stressors under production conditions, *CEUR-WS*, **2753** (2020), 156–169.
6. V. Kotsovsky, F. Geche, A. Batyuk, On the computational complexity of learning bithreshold neural units and networks, *Adv. Intel. Syst. Comp.*, **1020** (2019), 189–202.
7. V. Kotsovsky, F. Geche, A. Batyuk, Finite generalization of the offline spectral learning, in *2018 IEEE 2nd Intern. Conf. on Data Stream Mining Processing (DSMP)*, 2018, 356–360.
8. S. Fedushko, M. G. ml, T. Ustyianovych, Medical card data imputation and patient psychological and behavioral profile construction, *Proc. Comp. Scien.*, **160** (2019), 354–361.
9. S. Huang, H. Deng, *Data Analytics A Small Data Approach*, 1*st* edition, Routledge & CRC Press, 2021.
10. T. Hovorushchenko, A Herts, Y. Hnatchuk, Concept of intelligent decision support system in the legal regulation of the surrogate motherhood, *CEUR-WS*, **2488** (2019), 57–68.
11. E.B. Hekler, P. Klasnja, G. Chevance G, N. M. Golaszewski, D. Lewis, I. Sim, Why we need a small data paradigm, *BMC Med.*, **17** (2019), 133.
12. N. Shahid, T. Rappon, W. Berta, Applications of artificial neural networks in health care organizational decision-making: a scoping review, *PLOS ONE*, **14** (2019), e0212356.
13. S. Kaczor, N. Kryvinska, It is all about services-fundamentals, drivers, and business models, *J. Serv. Sci. Res.*, **5** (2013), 125–154.

14. J. Wang, *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications*, IGI Global, 2008.
15. D. Wu, K. Warwick, Z. Ma, M. N. Gasson, J. G. Burgess, S. Pan, et. al., Prediction of parkinson's disease tremor onset using a radial basis function neural network based on particle swarm optimization, *Int. J. Neural. Syst.*, **20** (2010), 109–116.
16. M. E. Karar, Robust RBF neural network-based backstepping controller for implantable cardiac pacemakers, *I. J. Adap. Cont. Sign. Proc.*,**32** (2018), 1040–1051.
17. W. Aftab, M. Moinuddin, M. S. Shaikh, A novel kernel for RBF based neural networks, *Abs. Appl. Anal.*, **2** (2014), 176253
18. M. Moinuddin, I. Naseem, W. Aftab, S. A. Bencherif, A. Memic, A weighted cosine RBF neural networks, *J. Mol. Biol. Biotech*., **2** (2017), 2.
19. T. Shaikhina, N. A. Khovanova, Handling limited datasets with neural networks in medical applications: a small-data approach, *Artifl. Intel. Med.*, **75** (2017), 51–63.
20. T. Shaikhina, D. Lowe, S. Daga, D. Briggs, R. Higgins, N. Khovanova, Machine learning for predictive modelling based on small data in biomedical engineering, *IFAC-PapersOnLine*, **48** (2015), 469–474.
20. I. Izonin, R. Tkachenko, S. Fedushko, D. Koziy, K. Zub, O, Vovk, RBF-based input doubling method for small medical data processing, *Adv. Intell. Syst. Comput.*, 2021.
22. Y. Bodyanskiy, I. Pliss, A. Deineko, Multilayer radial-basis function network and its learning, *IEEE Int. Conf. Comp. Sci. Inf. Tech.*, (2020), 92–95.
23. S. Babichev, J. Škvor, Technique of gene expression profiles extraction based on the complex use of clustering and classification methods, *Diagnostics*, **10** (2020), 584.
24. T. Kohonen, Essentials of the self-organizing map, *Neur. Netw*., **37** (2013), 52–65.
25. S. Subbotin, Radial-basis function neural network synthesis on the basis of decision tree, *Opt. Mem. Neur. Networks*, **29** (2020), 7–18.
26. F. Samuelson, D. G. Brown, Application of cover's theorem to the evaluation of the performance of CI observers, *Int. Joint Conf. Neur. Netw.*, (2011), 1020–1026.
27. Ye. V. Bodyanskiy, A. O. Deineko, Ya. V. Kutsenko, On-line kernel clustering based on the general regression neural network and kohonen's self-organizing map, *Aut. Control Comp. Sci.*, **51** (2017), 55–62.
28. M. Deshp, Using neural networks for regression: radial basis function networks, Available from: https://pythonmachinelearning.pro/using-neural-networks-for-regression-radial-basis-function-networks/.
29. R. Tkachenko, H. Kutucu, I. Izonin, A. Doroshenko, Y. Tsymbal, Non-iterative neural-like predictor for solar energy in libya, *CEUR-WS*, **2105** (2018), 35–45.