

21. Belyaeva, I. N., Bogachev, V. E., Chekanov, N. A. (2012). Programma postroeniya funktsii Grina dlya obyknovennogo differentsial'nogo uravneniya tret'ego poryadka. Svidetel'stvo o gosudarstvennoy registratsii programmy dlya EVM No. 2012661078.
22. Kamke, E. (1965). Spravochnik po obyknovennym differentsial'nym uravneniyam. Moscow: Nauka, 704.
23. Mihlin, S. G. (1947). Prilozheniya integral'nyh uravneniy k nekotorym problemam mehaniki, matematicheskoy fiziki i tehniki. Moscow-Leningrad: OGIZ izdatel'stvo tekhniko-teoreticheskoy literatury, 304.

Запропоновано генетичний метод для прогнозування показників здоров'я населення на основі нейромережесих моделей. Принципова відмінність запропонованого генетичного методу від існуючих аналогів полягає у використанні диплоїдного набору хромосом в особин популяції, яка еволюціонує. Така модифікація робить залежність фенотипу особини від генотипу менш детермінованою і, вкотре, сприяє збереженню різноманітності генофонду популяції і варіабельності ознак фенотипу впродовж виконання алгоритму. Крім цього, запропоновано модифікацію генетичного оператора мутацій. На відміну від класичного методу, особини, які піддаються дії оператору мутації, обираються не випадковим чином, а у відповідності до їх мутаційної стійкості, що відповідає значенню функції пристосованості особини. Таким чином, мутьуть особини, що характеризуються гіршими значеннями цільової функції, а геном сильних особин залишається незмінним. У цьому випадку зменшується вірогідність втрати досягнутого впродовж еволюції екстремуму функції внаслідок дії оператора мутацій, а перехід до нового екстремуму здійснюється у випадку накопичення достатньої питомих ваги кращих ознак в популяції.

Порівняльний аналіз роботи моделей, синтезованих за допомогою розробленого генетичного методу, показав, що найкращі результати досягнуті у моделі на основі нейронної мережі довгої короткочасної пам'яті. Під час створення і навчання моделі на основі мережі довгої короткочасної пам'яті було досліджено можливість використання методу рою часток для оптимізації параметрів мережі. Результати експериментальних досліджень показали, що розроблена модель дає найменшу помилку передбачення кількості нових випадків туберкульозу – середня абсолютна помилка складає 6,139, що менше у порівнянні з моделями, побудованими за допомогою інших методів.

Практичне використання розроблених методів дасть можливість своєчасно коригувати плановані лікувально-діагностичні, профілактичні заходи, завчасно визначати необхідні ресурси для локалізації та ліквідації захворювань з метою збереження здоров'я населення

Ключові слова: нейронні мережі, генетичний алгоритм, фенотип, модифікований генетичний оператор мутації, прогнозування показників здоров'я населення

UDC 004.93

DOI: 10.15587/1729-4061.2020.197319

CONSTRUCTION OF A GENETIC METHOD TO FORECAST THE POPULATION HEALTH INDICATORS BASED ON NEURAL NETWORK MODELS

I. Fedorchenko

Senior Lecturer*

E-mail: evg.fedorchenko@gmail.com

A. Oliinyk

PhD, Associate Professor*

A. Stepanenko

PhD, Associate Professor*

T. Zaiko

PhD, Associate Professor*

S. Korniienko

PhD, Associate Professor

Software Developer

A. Kharchenko

Software Developer

*Department of Software Tools

Zaporizhzhia Polytechnic National University

Zhukovskoho str., 64,

Zaporizhzhia, Ukraine, 69063

Received date 21.10.2019

Accepted date 17.02.2020

Published date 29.02.2020

Copyright © 2020, I. Fedorchenko, A. Oliinyk,

A. Stepanenko, T. Zaiko, S. Korniienko, A. Kharchenko

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0>)

1. Introduction

Quality of life of population is determined by different indicators, in particular health indicators, whose condition is predetermined by environmental factors. According to medical research conducted in recent years [1], there is a close relationship between the anthropogenic air pollution in certain areas and the increased population morbidity. As estimated by the World Health Organization (WHO), air pollution is the biggest factor of environmental health risks at present [2]. Based on this assessment, about 3.7 million of additional deaths are related to ambient air pollution, 4.3 million – to

air pollution indoors. Since many people are exposed to both indoor and outdoor polluted air, causes and deaths from various diseases caused by different sources cannot be determined through the usual generalization of data. The biggest health problems caused by direct influence of air pollution are related to diseases of blood circulation, respiratory diseases, cancer, neuro-mental disorders, as well as some others [3, 4].

Consequently, the health condition and population morbidity in a region can be considered as derivatives from the environment.

The use of known statistics methods for forecasting the dependence of health indicators, as well as mathematical

models suggested in known literary sources [1–14], is associated with certain limitations and requirements for target functions. When applying such methods, it is impossible to increase the accuracy of the forecast when parameters change, for example, in forecasting the dependence of indicators of public health on pollutant emissions in the air. These restrictions, when searching for optimal solutions, do not make it possible to increase the accuracy of the forecast to the desired value, which necessitates constructing such models that could provide higher performance accuracy of the forecast. Such models can be models based on artificial neural networks that are able to process multi-dimensional data of different types, as well as characterized by high approximating and generalizing properties.

Therefore, it is a relevant task to develop methods and models for forecasting population health indicators based on neural network technologies.

2. Literature review and problem statement

The results of research on establishing a mathematical dependence of population health indicators based on neural network models on pollutant emissions volumes are reported in paper [5]. In the proposed model an independent variable is the volume of pollutant emissions, and the dependent one is a morbidity indicator (1)

$$K_{morb} = f(x_{emiss}), \quad (1)$$

where K_{morb} is the morbidity indicator, x_{emiss} are indicators that characterize the impact of emission volumes.

Based on the above data and analysis of statistical data [2–5], one can conclude that the desired mathematical model would be stochastic rather than deterministic.

Paper [6] reports the study results on that, in addition to the volume of pollutant emissions, the morbidity rate is affected by a set of other factors whose exact number is hardly determined. If one denotes these factors x_1, x_2, \dots, x_n , then the generalized model of dependence (1) can be represented in form (2):

$$K_{morb} = f(x_{emiss}, x_1, x_2, \dots, x_n). \quad (2)$$

While analyzing [7], it was found that the main factor influencing human health emissions is the presence of toxic substances in their composition. A separate group of diseases was identified in the study of atmospheric air pollution effect on public health [5–7]. This group includes chronic obstructive lung diseases, bronchus, bronchial asthma, as well as lung cancer, diseases of the cardiovascular and nervous system.

According to research by the Central Geophysical Observatory [2], in 2015, 4.5 million tons of harmful substances were emitted into the atmospheric air in Ukraine, of which 62 % were from stationary sources and 38 % – from mobile sources. The main air pollutants are energy and metallurgy enterprises (55 % and 22 % of all pollution from stationary sources). The main air polluters are the enterprises of energy and metallurgy (55 % and 22 % of all dirt from stationary sources). They are responsible for the increased content of specific harmful substances: formaldehyde, phenol, hydrogen fluoride, ammonia, with especially large quantities of nitrogen dioxide and carbon monoxide. Therefore, the effect of these

toxic substances emitted from stationary sources is taken into consideration in the research; it was decided to build a model. A model should be developed to predict the morbidity rate using the population of Ukraine as an example; the people who fell ill given the unfavorable environmental situation in cities, depending on the types and concentrations of pollutants. The practical value of the developed model implies that it can be used to predict the dynamics of health indicators in the future for other cities. Using the developed model would enable timely adjustments to the planned medical and diagnostic, preventive measures, to determine the necessary resources in advance to localize and eliminate the diseases in order to preserve public health. It is worth noting that the methods proposed in the paper for the synthesis of models of public health indicators can also be used to process data from other sources and other countries.

Study [7] reports the results according to which the character and degree of exposure to toxic substances, their ability to provoke pathological conditions in the human body, vary depending on the combination of meteorological and climatic factors. Precipitation and high temperatures, on the contrary, contribute to intense decomposition of substances. A higher temperature near the Earth's surface during the daytime causes the air to rise upwards, leading to additional turbulence. Once the air warms up to 10 degrees and above, the volumes of harmful substances begin to accumulate in the atmosphere. At night the temperature near the ground surface is lower, so the turbulence decreases. This phenomenon reduces the dispersion of exhaust gases. Therefore, the construction of the model will take into consideration the average air temperature and rainfall per month.

Paper [8] reports the results of research according to which morbidity indicators are affected by the quality of health care of the population. Therefore, the main metrics, which we should take into consideration in the construction of a model of morbidity rate, is an indicator of the number of physicians (all specializations) in a region. We shall also use an indicator of the number of hospital beds at stationary departments of medical establishments in the region as a quantitative indicator of medical services.

It was established in work [9] that the distribution of morbidity in different regions is statistical, so the number of people in the region should be considered for modeling such a dependence.

Medical data [3, 5] testify to that the general morbidity of population has different indicators for different age groups (it usually increases with age). A tendency was found that older people are more likely to develop cardiovascular disease, tuberculosis, and cancer than young adults. That is, high rates of morbidity are characteristic of regions with high proportions of elderly people. The regions with the highest average age of residents are potentially regions with unfavorable population morbidity. Therefore, it is also advisable to take into account the average age of the population in the region.

Thus, the generalized model of the dependence of health indicators on the volume of emissions can be reduced, under certain assumptions, to form (3):

$$K_{morb} = f \left(\begin{matrix} x_{emiss}, x_{popul}, x_{temp}, \\ x_{rainfall}, x_{docs}, x_{beds} \end{matrix} \right), \quad (3)$$

where x_{popul} is the indicator that characterizes the impact of population quantity, x_{temp} is the average air temperature,

$x_{rainfall}$ is the rainfall quantity, x_{docs} is the indicator that characterizes the impact of the number of doctors, x_{beds} is the indicator characterizing the impact of total beds at stationary clinics.

Article [10] proposes a classic regression analysis to derive a mathematical dependence of health indicators on pollutant emissions. It is shown that the classical method of stochastic forecasting of the morbidity explores interrelationships between indicators of morbidity and factors that predetermine it, when the dependence between them is not strictly functional and distorted by the influence of foreign factors. It is also shown that different correlation and regression models of morbidity are constructed during correlation-based regression analysis. These models distinguish factor and effective indicators (attributes). The authors of the cited work described a regression analysis, which shows the choice of a communication form and a model type to determine the estimated values of the dependent variable (effective attribute). The work developed non-adaptive regressive models, which consider the entire background of morbidity over the studied area. However, in order to build them, all existing data and observations of recent years were used, which have similar characteristics. Thus, once the properties of the morbidity process changed, the outdated data would no longer help to refine the forecast. Therefore, there remained the unresolved issue associated with the fact that the non-adaptive models make it possible to obtain projection of the long term morbidity. Such models ignore local fluctuations in the epidemic parameters and are poorly suited for short-term forecasting. The option to overcome the appropriate difficulties may be to calculate the medium-term morbidity estimate at a sufficiently large sliding window width. Consequently, the developed model must be sufficiently sensitive in order to respond to the current morbidity tendencies for the formation of forecasts a few weeks ahead.

In [11], the results of studies into the use of Bayesian networks for the prediction of morbidity were reported. Bayesian networks have been shown to be an effective, compact, and intuitive way of representing the uncertainty-related knowledge. The Bayesian Network (BN) was presented as a graphical model that reflects the probabilistic dependences of a set of variables and allows the probabilistic inference to be derived from these variables. It was shown that in medical diagnostics, the most probable diagnosis is defined as the value of the set of possible diagnoses, which has the maximum probability of having the disease under the condition of a specific data set. These data include symptoms, test results and other attributes. Construction of the authors' BN is carried out at both large and small volumes of initial data, but algorithms for estimating model parameters are difficult to calculate. Therefore, the authors analyzed the BN based on a narrow sliding window of observations. The cited work did not address the issue related to that Bayesian networks provide only short-term disease prediction.

Paper [12] gave a description of the use of artificial neural networks (ANN) to establish the dependence of population health indicators on diseases caused by external factors. The authors show that ANNs make it possible to simulate various kinds of dependences, which can be based on linear models, generalized linear models, and nonlinear models. It is the ability of ANN to generalize and highlight the hidden dependences between input and output data that underlies the obtaining of reliable statistical forecasts. The

paper shows that the potential prognostic ability of the neural networks is better due to the more qualitative division of classes predetermined by the use of smooth transformation functions. The functions ensure the preservation of information in the final decision-making phase.

The main drawback of the cited work is that the use of neural networks requires long-term time costs to perform a training procedure that often do not make it possible to use ANN in real-time systems [13]. Thus, after analyzing the cited work, one can conclude that ANN can be a very effective mathematical basis for forecasting the dependence of population health indicators on the emissions of pollutants into the air.

Thus, at present, there is no universal technique to forecast morbidity, the result being the researchers are forced to choose prognostic models, based on comparison of the results obtained with the help of different methods on the basis of empirical data.

Having analyzed works [10–13], we established that the set task can be resolved by the effective use of ANN because models based on artificial neural networks provide the possibility of processing multidimensional data of various types (thereby implementing the function of many variables), high adaptability to external changes, provide the possibility of synthesis of models with high approximating and generalizing properties. Therefore, it is necessary to develop a method for constructing neural network models based on empirical data, which would make it possible to synthesize the models of dependence of health indicators on the volumes of pollutant emissions.

The use of traditional statistics methods to predict the dependence of health indicators [14], as well as mathematical models proposed in [14], is associated with certain limitations and requirements for target functions. When using such methods, it is impossible to increase the accuracy of the forecast when parameters change, for example, in forecasting the dependence of indicators of public health on pollutant emissions in the air. These restrictions, when searching for optimal solutions, do not make it possible to increase the accuracy of the forecast to the desired value. The application of genetic algorithms (GA), based on mechanisms of natural selection and inheritance, avoids a series of constraints, and thereby increases the accuracy of the prognosis [15].

The evolutionary approach is used by GA [15] when the search for an extremum of the target function is carried out simultaneously in many areas by using a population of possible solutions. The transition from one population to another avoids getting into the local optimum; in this case, GA is characterized by the polynomial complexity of computation.

The application of GA solves the problem by using a process similar to the biological development. It works as recombination and mutation of genetic sequences. Recombination and mutation are the genetic operators, that is, they control genes (sequence of codes) containing all the information necessary to create a functional organism with certain characteristics (a genotype) [16].

For the case of genetic optimization used to solve forecasting-related tasks, the sequence of codes usually takes the form of a series of numbers. Similar to the process of biological selection (where less suitable populations leave less offspring), the less suitable solutions are removed. In this case, more suitable solutions multiply, creating a different generation of solutions, which can contain several better solutions than the previous ones. The process of recomb-

nation, accidental mutation, and selection is an extremely effective mechanism for solving a given task.

The main purpose of the work is to study the possibility of application of GA to solve the task of forecasting the population's health indicators depending on the volumes of pollutant emissions in the air, at minimum time costs.

The analysis of methods and tools of statistical prediction [7–11] indicates that the application of GA in a given aspect does not contradict the logic and mathematical basis laid down in these methods. In this regard, it is expedient to develop a forecasting model of the dependence of population health indicators on the volumes of emissions of pollutants in the air using GA and a modification of one of the operators of the genetic method [16].

In recent years, various methods and software tools [6–16] have been proposed that employ artificial neural networks to predict morbidity. However, known models often do not make it possible to provide for the acceptable reliability of forecasting results. This situation is primarily due to the fact that the architecture of a neural network model, its topology, the parameters values are chosen on the basis of an expert evaluation or empirically. These parameters include the number of layer nodes, a network optimization method, the size of subsampling, the number of epochs of network training, etc. To find the optimal values for these parameters, one can use such stochastic methods as a particle swarm method and the genetic algorithms. The combination of genetic algorithms and neural networks is known in the literature under the abbreviation COGANN (Combinations of Genetic Algorithms and Neural Networks) [16]. The use of GA for the training of neural networks has the following advantages: genetic algorithms are insensitive to the increase in the dimensionality of an input data set, such methods do not require the differential target function; at each iteration, they work with a set of solutions that make it possible to explore the search space more thoroughly and to go beyond regions of local extrema. Therefore, to overcome the specified issue, one can develop a genetic algorithm in order to select the neural network parameters.

3. The aim and objectives of the study

The aim of this study is to create a method of synthesis of neural network models based on a genetic approach in order to forecast the indicators of population health.

To accomplish the aim, the following tasks have been set:

- to develop the basis of a neural network model on the dependence of health indicators on the volume of pollutant emissions;
- to construct a method for building neural network models based on a long short-term memory;
- to perform an experimental study of the proposed genetic method when synthesizing neural network models of the dependence of population health indicators.

4. Development of the basis for a neural-network model of the dependence of health indicators on the volume of pollutant emissions

In order to construct a mathematical model of the dependence of population health indicators on the volume of pollutant emissions, we used artificial neural networks based on a multi-layer perceptron [16]. The applied input parameters were

the parameters of the volume of pollutant emissions into atmospheric air, number of people, the median age of population, average temperature, rainfall, number of doctors in a region, and the number of beds at healthcare facilities in the region [1–3].

Selecting a neural network's parameters, specifically the number of neurons in a hidden layer, is in most cases a rather complex task and is usually performed based on an expert evaluation. However, there are several recommendations on this matter. Thus, Hecht-Neilson [17], in order to compute the upper network of the number of hidden elements, used the Kolmogorov theorem [17], whereby any function of n variables could be represented as the superposition $2i+1$ of one-dimensional functions. This network h equals twice the number of input elements plus unity (4):

$$h \leq 2i + 1, \tag{4}$$

where i is the number of input elements.

Consequently, the dependence model has an input layer of the network containing seven neurons (based on the number of input parameters). The developed model of dependence (3) that employs the Kolmogorov theorem takes the form [17]:

$$K_{morb} = \sum_{j=1}^{2n+1} p_j \left(\sum_{i=1}^n d_{ij}(x_i) \right), \tag{5}$$

where n is the number of input parameters, p_i and d_{ij} are the continuous functions, and d_{ij} does not depend on K_{morb} . This formula shows the implementation of multi-variable functions as the summing operation and the composition of one variable function.

Of course, it is quite difficult to apply formula (5) in practice. However, this formula shows the possibility of implementing a complex dependence using a relatively simple neural network termed a multilayer perceptron. Therefore, we shall build a three-layer perceptron, which has an input layer, an output layer, and a hidden layer of neurons that implements the activation function. This network implements the following representation [17]:

$$y = \sum_{i=1}^h \varphi_i f(\omega_{i,0}x_1 + \omega_{i,1}x_1 + \omega_{i,2}x_2 + \dots + \omega_{i,n}x_n), \tag{6}$$

where φ_i is the matrix of weights of links between the outputs of the neurons in the hidden layer and the output neuron of the network, $\omega_{i,n}$ is the matrix of connection weights between the input neurons and the neurons in the hidden layer, which actually implement the activation function, f is the neuron activation function of the hidden layer.

The network's input vector is defined as a set of incidence values that come to the input neurons over one iteration of training. The network's output vector is the set of incidence values on the output neurons. To calculate the number of neurons in hidden layers, we used the formula for assessing the number of semantic weights U_s for multi-layer perceptron with sigmoidal transfer functions [17]:

$$\frac{mN}{1 + \log_2 N} \leq U_s \leq m \left(\frac{N}{m} + 1 \right) (n + m + 1) + m, \tag{7}$$

where n is the input signal dimensionality, m is the output signal dimensionality, N is the number of elements in the training sample.

The number of neurons in the hidden layer is calculated from formula:

$$U = \frac{U_s}{n+m} \tag{8}$$

Thus, the developed dependence model has a hidden layer containing 12 neurons ($12 < 2 \cdot 7$) and the output layer containing one neuron [18].

One of the most important aspects of neural networks is the activation function, which introduces non-linearity to the network, making them universal approximation functions [19].

The activation function is the technique to normalize input data. That is, once we have a large amount of data at the input, then treating them using the activation function produces the data in the required range at the output. In the network under construction, the neurons from the input and hidden layer employ the ReLU [20] (Rectifier activation function) as the activation function. The advantage of using the ReLU activation function is that it is devoid of resource-intensive operations, there is no overgrowth or fading of the gradient and it provides rapid learning.

Thus, the first model will consist of an outer layer (seven neurons), one hidden fully-connected layer (12 neurons), and an output layer (one neuron). The scheme of the constructed network is shown in Fig. 1.

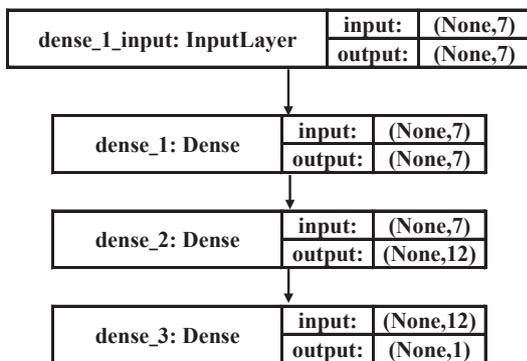


Fig. 1. Scheme of the neural network with a single hidden layer

In the course of our work, several multi-layer neural network models of direct propagation were constructed. Then, we added to the previously constructed model another fully-connected hidden layer of 12 neurons, which also uses ReLU as the activation function. The scheme of the constructed network is shown in Fig. 2.

Thus, the model of a neural network with two hidden layers was constructed. The model consists of an input layer containing 7 neurons (an input signal is sent to each of them), 2 hidden fully-connected layers (each containing 12 neurons), and an output layer consisting of one neuron (Fig. 2).

Retraining is one of the significant problems that complicate the practical application of neural networks. One technique to prevent the neural network retraining is the Dropout method, which implies excluding certain neurons of the network in the learning process [21].

The main idea of Dropout is that instead of training a single neural network one trains an ensemble of several deep

neural networks (Deep Neural Network, DNN), to subsequently average the results.

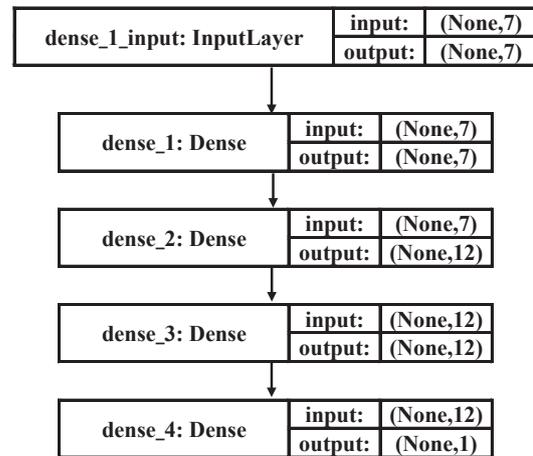


Fig. 2. Scheme of the neural network with two hidden layers

Training networks are obtained by dropping out neurons with probability p from the network, so the probability that the neuron would remain in the network is $q=1-p$. Dropping a neuron out means that at any input data or parameters, it returns 0.

The dropped-out neurons do not contribute to the learning process at any stage of the algorithm of error back propagation, so dropping out at least one of the neurons is equivalent to training a new neural network.

We added to the constructed model with two hidden layers the dropping out after the first hidden layer (50 % neurons dropped out). The scheme of the constructed network is shown in Fig. 3.

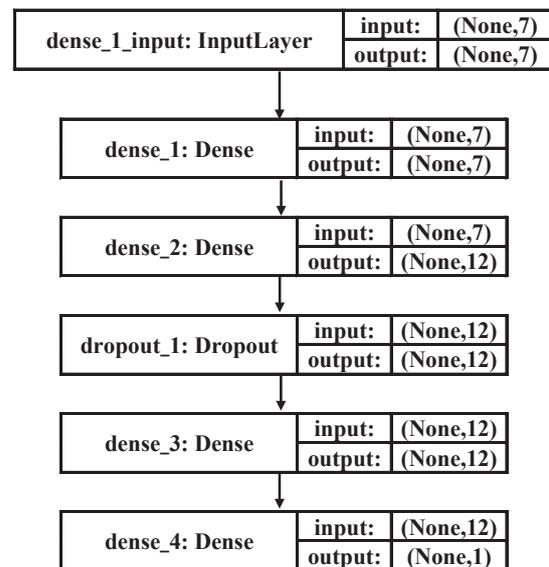


Fig. 3. Scheme of the neural network with two hidden layers and a dropout

Thus, we constructed a model of the neural network with two hidden layers and a dropout. The input layer contains 7, a fully-connected hidden layer (12 neurons), a dropout (disables half of all neurons of the layer), a fully-connected hidden layer (12 neurons and an output layer consisting of one neuron). The ReLU is used as the ac-

tivation function. The primary initialization of the synaptic scales corresponds to the normal distribution.

5. Development of a method for constructing neural network models based on a long short-term memory

The constructed models (Fig. 1–3) do not solve the issue of the long-term dependence of data sent to the input because the presented data can be considered a time series as the values of the examined parameters change over time. To analyze and predict a time series, one can use models based on neural networks with a long short-term memory (LSTM) [21].

Let us consider the structure of the LSTM layer in detail. The main element of such a network is a memory block, which at the same time as the h network status is computed at every step using the current input value x^t and the unit value in the previous step r^{t-1} . The input filter i^t determines how much the memory unit value in the current step should affect the result. The filter values vary from 0 (completely ignore the input value) to 1, which is provided by the region of values for the sigmoidal function:

$$i^t = \sigma(Y^i x^t + C^i h^{t-1}), \quad (9)$$

where C , Y are the training parameters of a neural network.

Forget gate makes it possible to exclude the memory values of the previous step in the calculation:

$$f^t = \sigma(Y^f x^t + C^f h^{t-1}). \quad (10)$$

Based on all the data that come at the time t , one calculates the status of the memory unit r^t in the current step, using filters:

$$\tilde{r}^t = \tanh(Yx^t + Ch^{t-1}). \quad (11)$$

$$r^t = f^t \cdot r^{t-1} + i^t \cdot \tilde{r}^t. \quad (12)$$

The output gate is similar to two previous ones and takes the form:

$$o^t = \sigma(Y^o x^t + C^o h^{t-1}). \quad (13)$$

The final value of the LSTM layer is determined by the output gate (13) and the nonlinear transformation over the state of the memory unit:

$$h^t = o^t \cdot \tanh(r^t). \quad (14)$$

The network receives eight parameters at the input – data for the previous period: morbidity rate, volume of emissions of pollutants into the atmospheric air from stationary sources, the number of people, people's average age, average temperature and average amount of precipitation, the number of doctors in a region, the number of beds at stationary health care establishments in the region. The hidden LSTM layer is made up of twenty neurons, and the original layer is from one neuron. The adam algorithm was used as an algorithm for optimization [21]. Adam is the optimization algorithm that can be used instead of a classical procedure to reduce a random gradient, to update the iterative weight of the network based on training data [22]. The algorithm

combines the advantages of such classic gradient descent extensions as the adaptive gradient algorithm (AdaGrad) and the moving average of squared gradients (RMSProp).

The main feature of the algorithm is the average values of both the gradients and the second moments of the gradients. Updating synaptic network weights using the adam algorithm is as follows:

$$m_{\omega}^{(t+1)} = \beta_1 m_{\omega}^{(t)} + (1 - \beta_1) \nabla_{\omega} L^{(t)}, \quad (15)$$

$$v_{\omega}^{(t+1)} = \beta_2 v_{\omega}^{(t)} + (1 - \beta_2) (\nabla_{\omega} L^{(t)})^2, \quad (16)$$

$$\hat{m}_{\omega} = \frac{m_{\omega}^{(t+1)}}{1 - \beta_1^t}, \quad (17)$$

$$\hat{v}_{\omega} = \frac{v_{\omega}^{(t+1)}}{1 - \beta_2^t}, \quad (18)$$

$$\omega^{(t+1)} = \omega^{(t)} - \eta \frac{\hat{m}_{\omega}}{\sqrt{\hat{v}_{\omega} + \epsilon}}, \quad (19)$$

where β_1 , β_2 are the hyperparameters indicating the exponential rate of decay at the time of evaluation; η is the initial level of training; ϵ is the small constant, introduced for numerical stability; m_{ω} is the exponential movable mean of the gradient; v_{ω} is the exponential mean of gradient square; $\nabla_{\omega} L^{(t)}$ is the gradient value over time t ; ω is the vector of gradient descent parameters [23].

Typically, the architecture of a neural network model, its topology, and the values of macro parameters are chosen based on an expert evaluation or empirically. For networks, these parameters can include the number of nodes in a long short-term memory layer, an optimizer, the sampling size, and the number of learning epochs.

To solve this problem, we developed a modification of the genetic algorithm to optimize the parameters of the constructed neural networks.

A forecasting model is based on the accumulated data about the following factors: main m_1, m_2, \dots, m_n (the volume of emissions) and auxiliary a_1, a_2, \dots, a_n (the number of people, precipitation, doctors, beds at stationary branches), where n is the length of the current part of the series (the number of observations of a time series), which is 20–30 values. We shall represent these data as the fuzzy time series $F_1(t)$ and $F_2(t)$, where $F_1(t)$ corresponds to the main, and $F_2(t)$ to auxiliary, factors in prediction [23]. Then the dependence in the following form:

$$F_1(t) = ((F_1 - k), F_2(t - k)), \dots, ((F_1(t - 2)), F_2(t - 2)), ((F_1(t - 1)), F_2(t - 1)) \quad (20)$$

is termed the factor prediction model of the k -th order based on fuzzy time series [23].

As it follows from the analysis of sources [16–23], finding the optimal solution using a GA requires that about two or three million individuals should be born. However, a high resource cost of determining the target function value for each individual can greatly prolong the time of an optimum search.

To solve this problem, it was decided to develop a modification of GA, which could significantly reduce the optimi-

zation time. In the developed modification, it was proposed to use the altered operators of interbreeding, selection, and mutation, as well as the new genetic selection operator of the second order based on the magnitude of mutation probability.

The proposed modification of the genetic method implies adding to the karyotype of each individual another chromosome with the same gene composition, that is, to use the diploid set consisting of two homological chromosomes. Both chromosomes are exposed to the same operators with the same parameters. Thus, when interbreeding, the karyotype of a descendant would also consist of two homological chromosomes, similar to his parents. The dominant gene in the proposed modification is chosen randomly from two allelic genes and is used to calculate the value for the adaptability function – a fitness function, that is, speaking in terms of biology, it determines the phenotype of an individual [23].

Denote an individual via a_n^t , where n is the number of an individual, t is some point in time in the evolutionary process. The accepted vector of controlling variables is $x=(x_1, x_2, \dots, x_m)$ – the smallest indivisible unit that characterizes in mathematical model (3) internal parameters at each t -th step of finding an optimal solution.

To describe individuals, we introduce two types of variable attributes that reflect qualitative and quantitative differences between individuals based on the degree of their expression. The qualitative attributes of individuals a_n^t are determined from the generalized model (3) as $s(\bar{x})$, where each point x corresponds to a_n^t . A gene is denoted by the combination $s_i(a_i)$, which determines the fixed value of controlling variable x_i . Each individual is characterized by m genes, and $s(\bar{x})=(s_1, s_2, \dots, s_m)$ can be interpreted via a chromosome containing n linked genes, which follow one another in a strictly defined sequence. The chromosome of individual a_n^t is to be denoted via x_n^t [23], that is,

$$\begin{aligned} x_n^t &= x(a_n^t) = (x_1(a_n^t), x_2(a_n^t), \dots, x_m(a_n^t)) = \\ &= s(\bar{x}) = (s_1, s_2, \dots, s_m). \end{aligned} \tag{21}$$

Quantitative attributes are the attributes that reflect variability; in this regard, the degree of their expression can be characterized by a number and is calculated in the work from formula:

$$d(x_i^t, x_j^t) = \sum_{n=1}^m x_n(a_i^t) \cdot x_n(a_j^t), \tag{22}$$

where a_i^t, a_j^t are the individuals, x_i^t, x_j^t are the genes equal in their values, m is the number of positions [23].

At the first stage, population is initialized. The gene composition of each of the two homological (H, H') chromosomes is selected randomly. To determine the phenotype of an individual, we select from each allele a gene that would be denoted as dominant and would determine the phenotype of the individual, that is, involved in the calculation of the function of the individual adaptability. Determining an individual's phenotype can be represented in the form of formula [21–23]:

$$F_j = \sum_{i=1}^m \text{rand} [H_j g_i; H'_j g_i], \tag{23}$$

where F_j is the phenotype of the j -th individual, m is the number of genes in chromosomes and $H_j g_i$ is the i -th gene in a pair of homological chromosomes of the j -th individual.

Therefore, the arguments of the individual's fitness function are defined. After calculating the functions of adaptability and selecting the individuals within a population, the interbreeding is performed. The genotype of an individual-descendant has the same structure as the genotype of the parents, that is, it consists of two homological chromosomes. The mutation operator is applied to the offspring. At the same time, any allele in a pair of homological chromosomes can mutate, but only one gene mutates in each allele [24].

Hereafter, the evolution of population P^t is represented as the alternation of generations, during which individuals change their variable attributes:

$$\eta_{cep}(t) = \frac{1}{t} \sum_{n=1}^m \eta(a_n^t), \tag{24}$$

where the totality of m genotypes of all individuals ($a_1^t, a_2^t, \dots, a_m^t$) forming the population P^t and a chromosomal set ($x_1^t, x_2^t, \dots, x_m^t$), which contains complete genetic information about the populations P^t in general.

The procedure for selecting the “best” solution from the population P^t takes into consideration not only the value of the fitness function F_j but also the chromosome structure x_i^t , thus it can be represented as [25]:

$$d(a^t, a_i^t) = \min_{i=1, m} d(x(a^t), x(a_i^t)) \tag{25}$$

provided $\eta(a_i^t) < \eta(a^t)$, where a_i^t is the “best” individual in the population P^t , a_i^t is the individual excluded from the population P^t , $d(x(a^t), x(a_i^t))$ is the measure of “proximity” of the genotype of individuals.

Next, similar to the classical method, the cycle repeats until the end of conditions of completion of optimization.

Summing up, we can say that the proposed method differs from the classical genetic method by using not a single chromosome but a pair of homological chromosomes, and by the addition of a phase for determining those genes in the allele that would take part in defining the value of an individual's fitness function. The result of this modification is the maintenance of a sufficiently high variability of attributes (genes) in the population (gene pool of population) during evolution, which at the same time may have a slight effect on the phenotype of individuals.

The specified modification method was used to optimize the LSTM [26] neural network: the number of network nodes, an optimization function at training, the size of sub-sampling, and the number of learning epochs.

Another proposed modification of the genetic method is a modification of the mutation operator. In contrast to the classical application of this operator, when the mutation is applied to all individuals in a generation with a certain probability, it is proposed to introduce the concept of the mutation stability of an individual, which is carried out in accordance with the following distribution:

$$x_i^1 = \begin{cases} x_i', & P(x_i') = \frac{\eta(x')}{\eta(x') + \eta(x'')}; \\ x_i'', & P(x_i'') = 1 - P(x_i'), \end{cases} \tag{26}$$

where x_i^1 is the descendant, $\eta(x'), \eta(x'')$ is the value of the adaptability function, according to the parent encoding x' and x'' [27–29].

The calculated value of an individual’s fitness function can be interpreted as the value of an individual’s mutation resistance. Thus, it is proposed at each iteration of the method, after calculating the function of adaptability, to rank the individuals from the received generation based on the value of mutation resistance. In contrast to the classical operator, we indicate at the beginning not the likelihood of a mutation but the proportion of individuals who subjected to operator (25).

$$K_{mut} = H_{gen} \cdot R_{mut}, \tag{27}$$

where K_{mut} is the number of individuals subjected to mutation, H_{gen} is the number of individuals in the received generation, R_{mut} is the share of individuals within a generation, who are subjected to mutation [29–31].

In fact, it is proposed to apply the operator only to individuals with the lowest value of the function of adaptability. In this case, when the population enters the region of the local extremum of the function, the mutation operator used must ensure the exit from such region. At the same time, it does not change the best individuals obtained at the moment of application of values, but conducts a search only at the expense of weaker individuals. The identified proportion of individuals subject to operator action should be sufficient to provide the potential for further evolution of the entire population.

Such mutations should be “softer” in the sense of preserving the best values found in previous iterations of the algorithm and should eliminate the risk of loss of the extremum of the function when applied without stopping the search for new better values.

Thus, a modified genetic method was developed for the parametric synthesis of a model based on a neural network of long short-term memory, which uses a modification of the mutation operator. The modified mutation operator allows one to search for optimal values, eliminating the loss of the best solutions found in the search.

6. Experimental study of the modified genetic method when synthesizing models of the dependence of population health indicators

To develop and test the model of the dependence of health indicators on the volume of pollutant emissions, we used statistical information about the amounts of pollutant emissions and carbon dioxide into the atmospheric air from stationary sources of pollution. We also used information on morbidity rate based on such indicators as the number of cases of circulatory system diseases (registered in outpatient establishments), the number of new cases of tuberculosis and the number of registered cases of cancer. Given the fact that the acquired data are expressed in absolute values, it is advisable to make a correction for the number of people in the region. Therefore, we used data on the number of people in the regions by years [2].

The developed models (Fig. 1–3) employ statistical data on the average temperature in the region and precipitation level, the number of doctors in the region, the number of beds at stationary health care establishments.

To solve the task, we chose programming environment based on the Python programming language; to accelerate the computation, we used the NVIDIA GeForce GTS 450,

with CUDA architecture support [41]. The NumPy library was used for convenient work with data arrays and the formation of datasets, the Python software package for scientific computing. To construct neural network models and work with them, the Keras library [42] and the Theano library [43] were selected.

Mean Absolute Error, the average absolute error, was used in the estimation of the forecast models [15]. Initial data processing was carried out prior to the beginning of the model creation and testing. Taking into consideration different data dimensionality, the input data were standardized. The data were transformed so that their average value was 0, and variance 1. In the course of our work, several models based on artificial neural networks were constructed and investigated. The result of training and operation of the first constructed network is shown in Fig. 4.

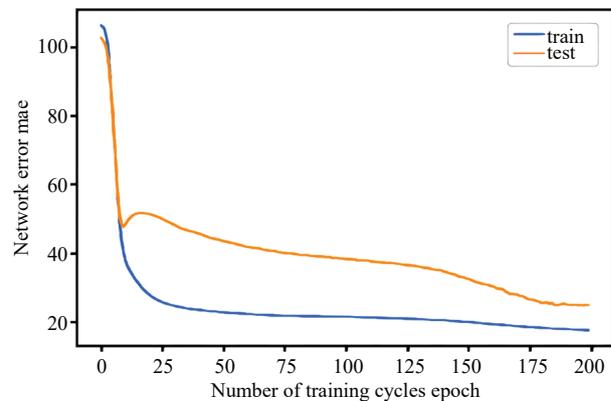


Fig. 4. Value of the network metric (mae) for indicator “The number of new cases of tuberculosis”

Fig. 1 shows that there is a gradual decrease in the error values during network training. Apparently, in the region of 10–15 epochs the training reaches a local extremum. The locality of a minimum of the error is indicated by further gradual decrease in the network error. Consequently, it is advisable, in this case, to keep on training the model.

To improve the metric of neural networks, their convergence, training costs, etc., there are several approaches associated with a search for an optimal network topology and learning methods. Thus, we added to the previously constructed model another fully-connected hidden layer of 12 neurons.

The second model consists of an input layer containing 7 neurons, 2 hidden fully-connected layers (each containing 12 neurons), and an output layer consisting of one neuron. Thus, the network has 264,505 parameters (synaptic weight) that can be trained. We trained the network with two hidden layers over 100 epochs (the subsampling size is 75) and by splitting a validation sample, which was equal to 0.1. The results of training the network are shown in Fig. 5.

In this case, the pattern is similar to the previous model – a gradual reduction in the error values during network training. There is a network convergence at the end of the training. Given the previous experience of the MLP network training, the training lasted over 100 epochs. Similarly to the previous case, the metric “The number of new cases of tuberculosis” demonstrates achieving a local minimum of the error.

One of the techniques to prevent a neural network re-training effect is the dropout method [31], which implies

excluding some neurons of the network during the learning process.

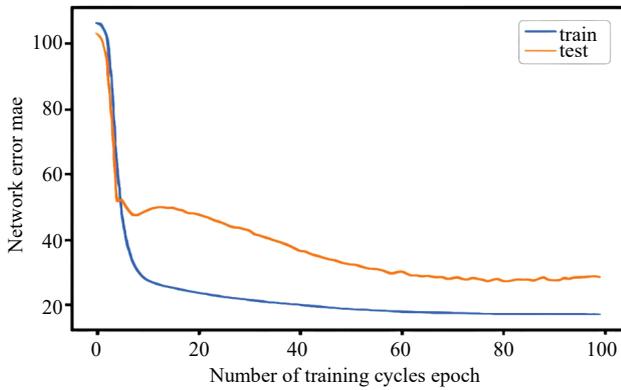


Fig. 5. Value of the network metric (mae) for indicator “The number of new cases of tuberculosis”

The third model consists of an input layer containing 7 neurons, 2 hidden fully-connected layers (each containing 12 neurons), a dropout, and a source layer consisting of one neuron. Thus, the network has 266,273 parameters (synaptic weights) that can be trained. The results of training the network are shown in Fig. 6.

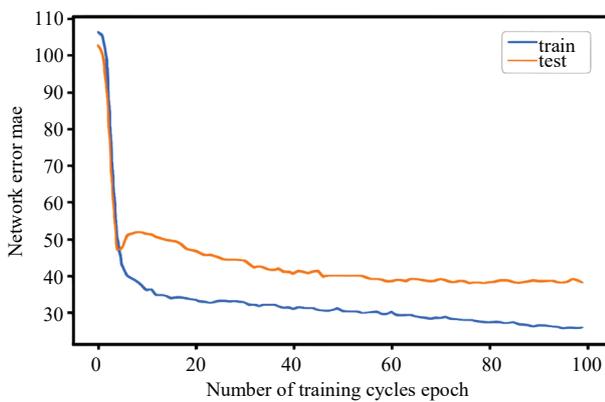


Fig. 6. Value of the network metric (mae) for indicator “The number of new cases of tuberculosis”

Similar to the previous cases, there is a network convergence during training, but at an earlier stage – approximately over 60–70 epochs of learning. In all three cases, the indicator “The number of new cases of tuberculosis” is found to have a local minimum of the network error, which is obviously a feature of the multilayer perceptron model for this morbidity rate. In addition, owing to the use of a dropout, the training curve loses its smoothness and turns into a polygon.

The reported data can be regarded as a time line, meaning the values of the examined parameters change over time. To analyze and predict the time series, one can use the models based on the neural networks of a long short-term memory [16].

A network using the LSTM layer receives eight parameters at the input. The hidden LSTM layer is made up of twenty neurons, and the output layer – one neuron. The results of testing the model are shown in Fig. 7.

Fig. 7 shows a change in the LSTM Network (mae) error value during training for indicator “The number of cases of TB disease”. For this indicator, we observed achieving a local minimum that is subsequently left. At the end of training, there is no further reduction of the model error value, so we can assume that during training a global minimum of errors was reached, and the network is considered to be trained. Table 1 gives comparative results of the mean absolute error (mae) values obtained when testing different types of models (logistic regression, multi-layered neural models, etc.), constructed in the course of our study.

Thus, Table 1 shows that the model based on the artificial neural network of a short long-time memory with 50 LSTM nodes in the layer produces the smallest error compared to the specified methods. Namely, the error in the prediction of the number of new cases of tuberculosis (MAE) is 6.139 and in the number of diseases of the circulatory system (MAE) is 441.889, which is an acceptable indicator. And to predict the number of all registered cases of cancer, the smallest average absolute error is 156.387, corresponding to a random forest.

In the course of our work, in order to optimize a long short-term memory network, we used a method of particle swarm. The results of algorithm implementation (determining the smallest network error value at each iteration of the algorithm) are shown in Fig. 8.

Table 1

Values of the mean absolute error when testing models constructed during our study

Predicted parameter	Predicting model type								
	Logistic regression	Support vector method	Least square method	Random forest	Nearest neighbor method	Multi-layer perceptron with one hidden layer (128 neurons)	A multilayered perceptron with two hidden layers (128, 1024 and 128 neurons)	Multilayer perceptron with two hidden layers and dropouts (128, dropout (0.5), 1024, dropout (0.5), 128 neurons)	A network of long-term short-term memory with 50 nodes of LSTM layer
Number of new cases of tuberculosis	29.764	40.271	22.957	7.671	8.236	23.445	21.675	22.3047	6.139
The number of cases of diseases of the circulatory system	3814.174	2794.433	1789.727	571.018	573.004	1766.470	1752.416	1620.676	441.889
The number of all reported cancer cases	1400.357	1050.690	367.381	156.387	210.157	336.419	338.953	272.465	226.096

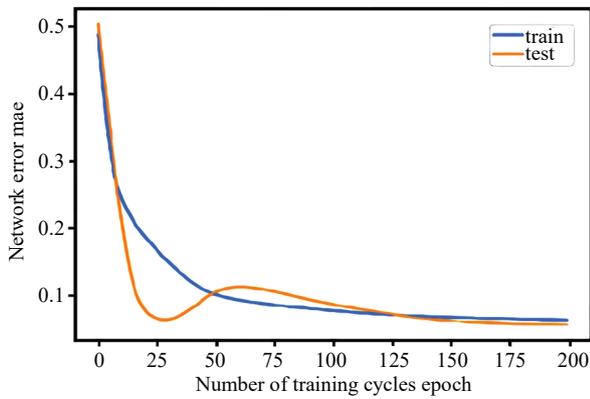


Fig. 7. Value of the LSTM network metric (mae) during training for indicator “The number of cases of tuberculosis diseases”

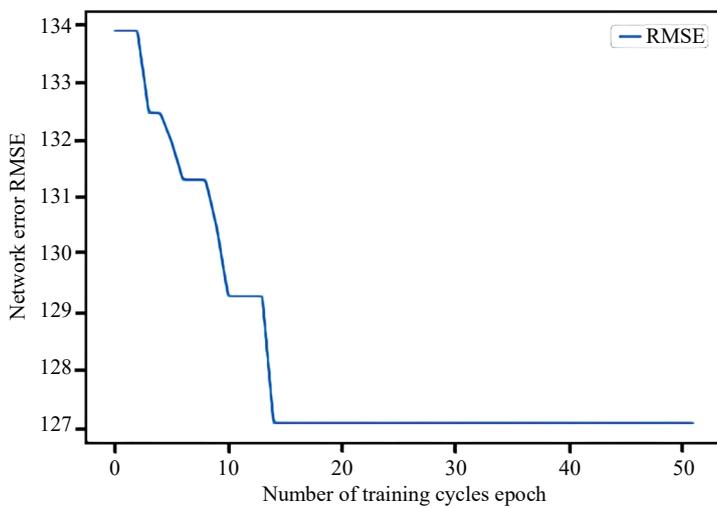


Fig. 8. Result of the particle swarm algorithm operation to optimize a long short-term memory network

Note that the constructed long short-term memory network made it possible, on the test sample, to obtain the RMSE error value of 127.087, which is an acceptable indicator for the practical task being solved.

Thus, the results of our study have shown that a model that can be used for the dependence of health indicators on the volume of pollutant emissions is the model based on the artificial neural network of a long short-term memory. The use of the modified genetic method should be used to select the parameters such as the number of the LSTM layer nodes, a network optimization method, the size of subsampling, and the number of epochs of network learning.

7. Discussion of results of studying the modified genetic method

Our comparative analysis of the constructed models (Table 1) reveals that the best results for the mean absolute error in predicting the number of new cases of tuberculosis were demonstrated by the long short-term memory network. Specifically, MAE is 6.139, which is an acceptable indicator compared to the method of supporting vectors, whose error is 40.271. To predict the number of cases of circulatory system diseases, the best results were obtained when using the

network of a long short-term memory (MAE is 441.889). In predicting the number of all registered cases of cancer, we received the smallest error in the random forest, which is 156.387, compared to logistic regression, which is 1400.357.

The results of analyzing the operational stability of the modified genetic algorithm are shown in Fig. 1–4. We performed 20 algorithm launches with different number of iterations. The above charts show that during the training of the network, the value of absolute error decreases and at the end of training there is a network convergence, which leads to a local minimum and a further exit from it. At the end of the training (Fig. 4), there is no further reduction in the error value of the model, so we can assume that a global minimum of the error is achieved during training, and the network is considered to be trained. In addition, due to the use of a dropout, the training curve loses smoothness and turns into a broken line.

Fig. 5 shows that the particle swarm algorithm operation, used to optimize a long short-term memory network, yielded the smallest error value (RMSE), 127.08, which is an acceptable indicator.

Thus, the proposed modified genetic method makes it possible to increase the accuracy of forecasting and reduce the time of training when synthesizing the models of the dependence of population health indicators on the volume of pollutant emissions. This is achieved due to that the developed modified methods employ new heuristic procedures, including the use of the diploid set of chromosome of the population that evolves. Such modification makes the dependence of the phenotype of the individual on the genotype less deterministic and, thus, helps preserve the diversity of the gene pool of the population and the variability of the attributes of the phenotype during the execution of the algorithm. In addition, we have proposed a modification of the genetic mutation operator. Unlike the classical method, individuals that are subjected to the action of the mutation operator are selected not randomly but

in accordance with their mutational stability, which corresponds to the value of the fitness function of the individual. This has made it possible to increase the accuracy indicator compared to the basic version of the genetic algorithm.

The disadvantage of the proposed modified genetic method, developed and investigated in this work, is the need to spend a great deal of time processing large data sets, which is unacceptable when solving some practical tasks. Thus, the limitations on the use of the proposed modified genetic method are small amounts of processed data.

The development of this study may be related to the elimination of the specified shortcomings, due to the practical threshold of using the proposed modified genetic method for constructing models based on the neural network of a long short-term memory. For this purpose, it is advisable to develop its parallel implementation, which would significantly (by times) increase the speed of the method operation. The associated problems that may arise when designing parallel modifications to the genetic method to build models based on the neural network of long short-term memory are related to the need for scheduling resources in a parallel computer system. They lead to increased requirements for hardware involved in the process of genetic optimization.

8. Conclusions

1. Models of dependence of health indicators on pollutant emissions based on artificial neural networks have been developed. The first model built consists of one hidden layer. When testing this model, a mean absolute error of 708.78 was obtained. Next, the model was created with two hidden layers. The second model during the test showed a mean absolute error of 721.01. We also created a model with two hidden layers and dropouts. When testing this model, a mean absolute error of 638.5 was obtained. The model was then built using a long short-term memory with 50 nodes of the LSTM layer. The mean error values of 647.13 were obtained when testing this model. Comparing the obtained results with known methods such as logistic regression, supporting vector methods, best square method, we can see that the developed models, reported in this work, yield the best result.

2. A method for constructing neural network models based on a long short-term memory has been developed. The proposed method uses a genetic approach for the parametric synthesis of neural models based on a long short-term memory. The fundamental difference between the proposed genetic algorithm and the existing modifications is the use of the diploid set of chromosomes in the evolving individuals. Such modification makes the dependence of the phenotype of the individual on the genotype less deterministic and, ultimately, helps preserve the diversity of the gene pool of the population and the variability of features of the phenotype during the execution of the algorithm. The result of such modification is to maintain a sufficiently high variability of the traits (genes) in the population (population gene pool) during evolution, which, at the same time, may have little effect on the phenotype of individuals. The proposed method

uses a modified genetic mutation operator, in which, unlike existing approaches to the implementation of such operators, individuals who are exposed to the mutation, are selected not in a random manner but in accordance with their mutational stability, which corresponds to the value of the fitness function of the individual. Thus, the “weaker” individuals are mutated while the genome of “strong” individuals remains unchanged. In this case, the likelihood of loss of the function reached during the evolution of the extremum due to the action of the mutation operator decreases, and the transition to a new extremum occurs if enough specific weight in the population is accumulated. This modification of the operator makes it possible to search for optimal values, excluding the loss of the ones found when looking for better solutions.

3. An experimental study of the proposed genetic method for the synthesis of neural network models of dependence of population health indicators has been performed. The results of our study have shown that the model developed gives the smallest error in predicting the number of new cases of tuberculosis, which is 6,139 and the number of diseases of the circulatory system, which is 441,889. While creating and training a model based on a long short-term network memory, the possibility of using a particle swarm method to optimize network parameters was explored. The particle swarm algorithm obtained the lowest error value (RMSE), 127.08, which is an acceptable indicator. The practical significance of this work is that the relevant task of synthesis of models of dependence of population health indicators on the basis of artificial neural networks has been solved, which would allow timely correction of the planned medical-diagnostic, preventive measures, advance determination of the necessary resources for localization and elimination of diseases in order to preserve the health of the population.

References

1. Paustenbach, D. (Ed.) (2002). *Paustenbach Human and ecological risk assessment. Theory and practice*. New York, 635.
2. Vykydy zabrudnuiuchykh rechovyn v atmosferne povitria. Derzhavna sluzhba statystyky Ukrainy. Available at: https://ukrstat.org/uk/operativ/operativ2009/ns_rik/ns_u/dvsr_u2008.html
3. Stan zabrudnennia pryrodnoho seredovyscha na terytoriyi Ukrainy. Available at: http://cgo-sreznevskiy.kiev.ua/index.php?fn=u_zabrud&f=ukraine
4. Tuberculosis. World Health Organization. Available at: <https://www.who.int/news-room/fact-sheets/detail/tuberculosis>
5. Ghazvini, K., Yousefi, M., Firoozeh, F., Mansouri, S. (2019). Predictors of tuberculosis: Application of a logistic regression model. *Gene Reports*, 17, 100527. doi: <https://doi.org/10.1016/j.genrep.2019.100527>
6. Mei, B., Xu, Y. (2019). Multi-task least squares twin support vector machine for classification. *Neurocomputing*, 338, 26–33. doi: <https://doi.org/10.1016/j.neucom.2018.12.079>
7. Rubal, Kumar, D. (2018). Evolving Differential evolution method with random forest for prediction of Air Pollution. *Procedia Computer Science*, 132, 824–833. doi: <https://doi.org/10.1016/j.procs.2018.05.094>
8. Dembinski, H., Schmelling, M., Waldi, R. (2019). Application of the iterated weighted least-squares fit to counting experiments. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 940, 135–141. doi: <https://doi.org/10.1016/j.nima.2019.05.086>
9. Soebiyanto, R. P., Kiang, R. K. (2000). Modeling Influenza Transmission Using Environmental Parameters. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science*, XXXVIII, 330–334.
10. Yi, H.-C., You, Z.-H., Zhou, X., Cheng, L., Li, X., Jiang, T.-H., Chen, Z.-H. (2019). ACP-DL: A Deep Learning Long Short-Term Memory Model to Predict Anticancer Peptides Using High-Efficiency Feature Representation. *Molecular Therapy - Nucleic Acids*, 17, 1–9. doi: <https://doi.org/10.1016/j.omtn.2019.04.025>
11. Speiser, J. L., Miller, M. E., Tooze, J., Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications*, 134, 93–101. doi: <https://doi.org/10.1016/j.eswa.2019.05.028>
12. Alam, S., Dobbie, G., Koh, Y. S., Riddle, P., Ur Rehman, S. (2014). Research on particle swarm optimization based clustering: A systematic review of literature and techniques. *Swarm and Evolutionary Computation*, 17, 1–13. doi: <https://doi.org/10.1016/j.swevo.2014.02.001>

13. Kumar, J., Goomer, R., Singh, A. K. (2018). Long Short Term Memory Recurrent Neural Network (LSTM-RNN) Based Workload Forecasting Model For Cloud Datacenters. *Procedia Computer Science*, 125, 676–682. doi: <https://doi.org/10.1016/j.procs.2017.12.087>
14. Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, Inc.
15. Viktorova, O. (2012). Use of fuzzy neural networks for technical diagnostics of road machinery. *Vestnik Har'kovskogo natsional'no-go avtomobil'no-dorozhnogo universiteta*, 56, 98–102.
16. McClure, N. (2017). *TensorFlow Machine Learning Cookbook*. Packt Publishing, 370.
17. Kolesnikov, K. V., Karapetian, A. R., Tsarenko, T. A. (2013). Henetychni alhorytmy dlia zadach bahatokryterialnoi optymizatsii v merezhakh adaptivnoi marshrutyzatsiyi danykh. *Visnyk Nats. tekhn. un-tu "KhPI"*, 56 (1029), 44–50.
18. Oliinyk, A., Fedorchenko, I., Stepanenko, A., Rud, M., Goncharenko, D. (2018). Evolutionary Method for Solving the Traveling Salesman Problem. 2018 International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T). doi: <https://doi.org/10.1109/infocommst.2018.8632033>
19. Lin, B., Sun, X., Salous, S. (2016). Solving Travelling Salesman Problem with an Improved Hybrid Genetic Algorithm. *Journal of Computer and Communications*, 04 (15), 98–106. doi: <https://doi.org/10.4236/jcc.2016.415009>
20. Haupt, R. L., Haupt, S. E. (2003). *Practical Genetic Algorithms*. John Wiley & Sons. doi: <https://doi.org/10.1002/0471671746>
21. Shkarupylo, V., Skrupsky, S., Oliinyk, A., Kolpakova, T. (2017). Development of stratified approach to software defined networks simulation. *Eastern-European Journal of Enterprise Technologies*, 5 (9 (89)), 67–73. doi: <https://doi.org/10.15587/1729-4061.2017.110142>
22. Fedorchenko, I., Oliinyk, A., Stepanenko, A., Zaiko, T., Shylo, S., Svyrydenko, A. (2019). Development of the modified methods to train a neural network to solve the task on recognition of road users. *Eastern-European Journal of Enterprise Technologies*, 2 (9 (98)), 46–55. doi: <https://doi.org/10.15587/1729-4061.2019.164789>
23. Oliinyk, A., Zaiko, T., Subbotin, S. (2014). Training sample reduction based on association rules for neuro-fuzzy networks synthesis. *Optical Memory and Neural Networks*, 23 (2), 89–95. doi: <https://doi.org/10.3103/s1060992x14020039>
24. Fedorchenko, I., Oliinyk, A., Stepanenko, A., Zaiko, T., Korniienko, S., Burtsev, N. (2019). Development of a genetic algorithm for placing power supply sources in a distributed electric network. *Eastern-European Journal of Enterprise Technologies*, 5 (3 (101)), 6–16. doi: <https://doi.org/10.15587/1729-4061.2019.180897>
25. Fedorchenko, I., Oliinyk, A., Stepanenko, A., Zaiko, T., Shylo, S., Svyrydenko, A. (2019). Development of the modified methods to train a neural network to solve the task on recognition of road users. *Eastern-European Journal of Enterprise Technologies*, 2 (9 (98)), 46–55. doi: <https://doi.org/10.15587/1729-4061.2019.164789>
26. Oliinyk, A. O., Zayko, T. A., Subbotin, S. O. (2014). Synthesis of Neuro-Fuzzy Networks on the Basis of Association Rules. *Cybernetics and Systems Analysis*, 50 (3), 348–357. doi: <https://doi.org/10.1007/s10559-014-9623-7>
27. Oliinyk, A., Fedorchenko, I., Stepanenko, A., Rud, M., Goncharenko, D. (2019). Combinatorial Optimization Problems Solving Based on Evolutionary Approach. 2019 IEEE 15th International Conference on the Experience of Designing and Application of CAD Systems (CADSM). doi: <https://doi.org/10.1109/cadsm.2019.8779290>
28. Sharifzadeh, M., Sikinioti-Lock, A., Shah, N. (2019). Machine-learning methods for integrated renewable power generation: A comparative study of artificial neural networks, support vector regression, and Gaussian Process Regression. *Renewable and Sustainable Energy Reviews*, 108, 513–538. doi: <https://doi.org/10.1016/j.rser.2019.03.040>
29. Buduma, N., Locascio, N. (Eds.) (2017). *Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms*. O'Reilly Media, 298.
30. Lapan, M. (2018). *Deep Reinforcement Learning Hands-On: Apply modern RL methods, with deep Q-networks, value iteration, policy gradients, TRPO, AlphaGo Zero and more*. Packt Publishing, 546.
31. Aggarwal, C. C. (2018). *Neural Networks and Deep Learning*. Springer. doi: <https://doi.org/10.1007/978-3-319-94463-0>