

Development of Genetic Methods for Predicting the Incidence of Volumes of Emissions of Pollutants in Air

Andrii Oliinyk¹[0000-0002-6740-6078], Ievgen Fedorchenko²[0000-0003-1605-8066],
Alexander Stepanenko³[0000-0002-0267-7440], Alexander Katschan⁴[0000-0001-8194-5572],
Yuliia Fedorchenko⁵[0000-1111-2222-3333], Anastasia Kharchenko⁶[0000-0002-4313-555X],
and Dmytro Goncharenko⁷[0000-0001-8973-3742]

Department of Software Tools, National University “Zaporizhzhia Polytechnic”, 64 Zhukovskoho str., Zaporizhzhia, Ukraine, 69063

¹olejnikaa@gmail.com, ²evg.fedorchenko@gmail.com,
³alex@zntu.edu.ua, ⁴alexander.j.katschan@gmail.com,
⁶kharchenko.13.08@gmail.com

Abstract. Air pollution essentially influences health of people, after all breath is a basis of life activity of any organism. Due to constant and repeated impacts on people through the air, they can change the quality of life and health of the population. The results of the analysis of the effect of polluted air on the incidence of the population are presented. Methods are proposed for constructing a mathematical model of the dependence of the incidence rates on the emissions of pollutants. A modified genetic method has been developed for optimizing model parameters based on a long short-term memory neural network. A modification of one of the operators of the genetic method, namely, the mutation operator, which allows the search for optimal values, excluding the loss of the best solutions acquired by the search, is proposed. Practical use of the developed methods will allow timely adjustment of the planned therapeutic, diagnostic, preventive measures, pre-determine the necessary resources for the localization and elimination of diseases in order to preserve the health of the population.

Keywords: modeling, emissions, pollutants, stationary sources, circulatory system diseases, cancer, neural networks, genetic algorithm, particle swarm method, python, keras, theano, cuda.

1 Introduction

In recent decades, the problem of air pollution by harmful chemicals has been considered in close connection with changes in health and mortality in many countries around the world. The problem of the subject area under study is very topical and attracts a lot of attention from the world scientific community in the field of health care, and from the side of researchers in the field of Data Science.

According to the World Health Organization (WHO), air pollution is currently the largest environmental health risk factor. According to this estimate, about 3.7 million additional deaths in 2012 were related to air pollution, and 4.3 million to indoor air pollution. Since many people are exposed to both indoor and outdoor air pollution, the number of deaths and diseases caused by these sources cannot simply be summed up, WHO proposes to estimate the final number of victims of pollution in 2012 at around 7 million (WHO, 2014). The biggest health problems caused by direct exposure to air pollution are related to circulatory diseases, respiratory diseases, cancer, neuropsychiatric disorders and some others. Consequently, the health status and morbidity of the population of the region can be considered as a derivative of the environment. Therefore, it is necessary to determine this impact, as well as the model of dependence of the number of patients on the types and volumes of emissions of pollutants into the air.

At the same time, reproduction of the model of such dependence is not a simple task, because the level of pollution is not the only factor affecting the level of morbidity and the dependence of morbidity on emissions is not linear. Therefore, such modeling requires the use of modern methods such as artificial neural networks, genetic algorithms and the like.

In the given work the research of dependence of indicators of morbidity of the population by diseases of blood circulation system, tuberculosis and oncological diseases of volumes of emissions of polluting substances into the air as a result of activity of stationary sources of pollution in various regions of Ukraine is carried out, and also development of model of dependence of indicators of morbidity on volumes of emissions of polluting substances.

Traditional mathematical models used for dependence analysis, as well as modern approaches to modeling using methods such as artificial neural networks, genetic algorithms, multi-agent systems (particle swarm algorithm) and their combined variants are considered.

In the course of the work, models of dependence of morbidity on the volume of pollutant emissions were built. A modified genetic method was developed to optimize the parameters of the model based on a neural network of long term memory. In addition, it was proposed to modify one of the operators of the genetic method, namely the mutation operator, which allows to search for optimal values, excluding the loss of the best solutions.

2 Formulation of the problem

Mathematical dependence of morbidity from pollutants amount can be defined as a function where the independent variable is the amount of pollutant emissions and dependent one – morbidity (1)

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

where K_{morb} – morbidity, x_{emiss} – indicators that describe the impact of emissions' amount.

Based on these data and statistical analysis [2], we can conclude that the desired mathematical model won't be deterministic, but rather stochastic.

Many other factors are influencing morbidity apart from pollutants amount, and their exact number is quite problematic to determine. If these factors are marked as x_1, x_2, \dots, x_n , then generalized model of relationship (1) can be represented in the form (2):

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

When analyzing statistical data it was determined [3] that the main factor of emissions' influence on human health is the presence of toxic substances in their composition. In turn, the nature and extent of exposure to toxic substances, their ability to induce pathological conditions in humans vary depending on the combination of meteorological and climatic factors such as temperature and rainfall.

In addition, of course, the quality of medical services affects morbidity rates. As the main metrics, which should be considered when building a morbidity dependance model, the number of physicians (all specialties) in the region and the number of hospital beds in inpatient health care facilities of the region as a quantitative indicator of health care were used. Finally, as the morbidity distribution in different regions is statistical, the population of the region should be taken into account to model such dependency. Since, according to the medical statistics data [1 – 3], total morbidity has different rates in different age groups (usually increases with age), the average age of the population in the region should also be taken into account.

Thus, a generalized model of morbidity dependency on emissions with certain assumptions can lead to type (3):

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

where x_{popul} – an indicator characterizing the impact of population size, x_{temp} – average air temperature, $x_{rainfall}$ – rainfall amount, x_{docs} – an indicator characterizing the influence of doctors' quantity, x_{beds} – an indicator characterizing the impact of the total number of beds in hospital wards.

3 Analysis of published data and problem definition

Methods of morbidity prediction are actively developed since the beginning of the XX century. In recent years, the number of works on this subject is growing rapidly due to development of information systems and accumulation of large amounts of statistics available for analysis.

The authors of the robot [4] presented the results of a comparison of long-term data on air emissions and mortality. In particular, it turned out that the reduction in the number of diseases of the circulatory system (per 1000 people) is due to a reduction in emissions from manufacturing enterprises and, in part, from housing and communal services. To build this model, the authors used dynamic Bayesian networks [5], which ensures that the data correlation structure is taken into account. Bayesian networks

allow estimating the probability of a certain event when observing some sequence of phenomena. The construction of the CMM is possible with both large and small amounts of source data, but the algorithms for evaluating the parameters of the model are difficult to calculate, so the CMM is usually analyzed on the basis of a narrow sliding observation window. For this reason, Bayesian networks currently provide only short-term morbidity predictions. Moreover, SMMs are often only used to detect elevated morbidity.

In the work [6] researchers have established a correlation between the number of completed suicides and the concentrations of suspended substances in the atmosphere, determined two days before the suicide. For this purpose, they used regression analysis [7]. The regression task is to find estimates of unknown parameters and to form a functional relationship between morbidity and factors that cause it. If the sliding window width is large enough, mid-term morbidity estimates can also be calculated, but achieving high quality is problematic.

In the works [4–7] remained unresolved issues related to the fact that the models are designed to take into account the entire history of morbidity in the analyzed area. All available data, or at least observations from recent years with similar characteristics, are used to build them. That is, if the properties of the morbidity process have changed, it is likely that outdated data will not help to clarify the forecast. To solve this problem it was accepted to use a neural network of long short-term memory.

4 The purpose and objectives of the study

The object of study is the dependence of the health indicators of pollutants' amounts. The purpose of the study is to create a model of health indicators dependency on the amounts of pollutants.

The research methods used include traditional models (logistic regression, support vector method, the least squares method, random forest, nearest neighbor method), neural networks, combined methods (neural networks and genetic algorithms, neural networks and multi-agent systems).

5 Development of modified genetic method based on long short-term memory neural network

For solving this problem it was decided to use the method of artificial neural networks, namely the use of multilayer perceptron [2]. For solving this problem it was decided to use the method of artificial neural networks, namely the use of multilayer perceptron [8].

The essence of proposed classic method optimization is to add a chromosome with the same genetic composition as each individual's karyotype, i.e. use diploid, consisting of two homologous chromosomes. Both chromosomes go through the same operators with the same parameters. Thus, the karyotype of crossbred offspring will also consist of two homologous chromosomes, like its parents. The dominant gene in

the proposed modification is chosen randomly from two allelic genes and is used to calculate the value of fitness function, that is, speaking in terms of biology, determines the phenotype of individuals.

Lets define the individual as ant, where n – the number of the person, t – arbitrary moment of evolution time. As a vector of control variables take $\bar{x} = (x_1, x_2, \dots, x_m)$ – its the smallest indivisible unit that describes internal parameters on each t -th step of finding the optimal solution in a mathematical model (3).

To describe the individuals we introduce two types of variable characteristics that reflect the qualitative and quantitative differences between individuals according to their severity. Qualitative characteristics of individuals a_n^t are determined from the generalized model (3) as $s(\bar{x})$, where each point \bar{x} corresponds to a_n^t . As the gene we take the combination $s_i(\alpha_i)$, which determines the value of fixed control variable x_i . Each individual is characterized by m genes and $s(\bar{x}) = (s_1, s_2, \dots, s_m)$ can be interpreted as chromosome containing n interlinked genes that follow each other in strictly defined sequence. a_n^t chromosome of the individual we will define as x_n^t [9], i.e.

$$x_n^t = x(a_n^t) = (x_1(a_n^t), x_2(a_n^t), \dots, x_m(a_n^t)) = s(\bar{x}) = (s_1, s_2, \dots, s_m). \quad (4)$$

Quantitative features show variation, and therefore the degree of their severity can be characterized numerically and calculated by formula below:

$$d(x_i^t, x_j^t) = \sum_{n=1}^m x_n(a_i^t) \cdot x_n(a_j^t), \quad (5)$$

where a_j^t, a_i^t – individuals, x_j^t, x_i^t – genes unequal in importance, m – number of positions [10].

The first step is the initialization of the population. Gene structure of each of the two homologous (H, H') chromosomes of individuals is chosen randomly. To determine the phenotype of individuals the gene is defined from each allele as a dominant and will define individuals phenotype, i.e. it will take part in fitness function evaluation of the individual. The definition of individual phenotype can be represented by the formula [10]:

$$F_j = \sum_{i=1}^m \text{rand} \left[H_j g_i; H'_j g_i \right], \quad (6)$$

where F_j – phenotype of j -th individual, m – number of alleles in a chromosome pair and H_{jgi} – i -th gene in a pair of homologous chromosomes of j -th individual.

Thus, in fact the arguments of individuals' fitness function are defined. After calculating the fitness function and selection of individuals in a population the crossbreeding is performed. Genotype of a descendant individuals has the same structure as the parent genotype, i.e. it consists of two homologous chromosomes. The descendants are subjected to mutations operator. At the same time, any allele can mutate in pairs of homologous chromosomes, but in each allele only one gene mutates.

Further evolution of the population P^t we will represent as generations alternation, during which individuals change their variable characteristics:

$$\eta_{cep}^t(t) = \frac{1}{t} \sum_{n=1}^m \eta(a_n^t), \quad (7)$$

where a set of m genotypes of all individuals $(a_1^t, a_2^t, \dots, a_m^t)$ which forms a population P^t and chromosome set $(x_1^t, x_2^t, \dots, x_m^t)$, which contains complete genetic information of the whole P^t population.

The procedure for selecting the "best" solution from P^t population takes into account not only the fitness function F_j value, but also the chromosome structure x_i^t , so it can be represented as [11]:

$$d(a^t, a_i^t) = \min_{l=1, m} d(x(a^t), x(a_l^t)) \quad (8)$$

provided that $\eta(a^t) < \eta(a^l)$, where a^l is the «best» individual in the P^t population, a_i^t – the individual which is excluded from the P^t population, $d(x(a^t), x(a^l))$ – a measure of genotype individuals "closeness".

Further, as in the classic method, the cycle is repeated until meeting the conditions for optimization completion.

Summarizing, we can say that the proposed method differs from the classic genetic method by using not one chromosome, but a pair of homologous chromosomes, and also by adding of the definition phase for those genes in alleles which will take part in determining the value of individual's fitness function. This modification results in maintaining a fairly high traits variability (genes) in the population (gene pool) during evolution, which, at the same time may have little effect on the phenotype of individuals.

Said method modification was used for neural network optimization LSTM [11]: the number of nodes, optimization function in learning, sub-sample size and the number of periods of learning.

Another proposed modification of the genetic method modifies mutations operator. Unlike the classic application of the operator, when all individuals in generation are subjected to mutations with a certain probability, it is proposed to introduce the concept of individuals' mutational persistense, which defined as following distribution:

$$x_i^1 = \begin{cases} x_i', & P(x_i') = \frac{\eta(x')}{\eta(x') + \eta(x'')}; \\ x_i'', & P(x_i'') = 1 - P(x_i'), \end{cases} \quad (9)$$

where x_i^1 – descendant, $\eta(x')$, $\eta(x'')$ – fitness function values, which evaluate parental encoding x' and x'' respectively [12].

The calculated value of individual's fitness function can be interpreted as a value of mutational stability of individuals. Thus, it is proposed at each method iteration after fitness function evaluation to rank individuals from obtained generation by mutational persistence value. Unlike classical operator, at a start the proportion of individuals who are exposed to the operator is specified instead of mutation probability (10).

$$K_{mut} = H_{gen} * R_{mut}, \quad (10)$$

where K_{mut} – the number of individuals exposed to mutation, H_{gen} – number of individuals in received generation, R_{mut} – the proportion of individuals in generation which are exposed to mutations [12].

In fact, it is proposed to apply the operator only to individuals with the lowest fitness function value. In this case, if the population falls into local extremum the mutation operator must allow to search the way out without changing the best received values at the time, but only at the expense of weaker individuals. The determined proportion of individuals exposed to the operator should be sufficient to ensure that the potential for further evolution of the entire population exists.

These mutations should be more "soft" in the sense of conservation of best values found in previous iterations of the algorithm and in sense of neutralizing the danger of losing function extremum when mutations are used, without stopping the search for new best values.

The modified genetic method for model parameters optimization based on long short-term memory neural networks, which demonstrates the morbidity dependence on emissions, was developed.

A modification of one of the operators of the genetic method, namely, the mutation operator, was also proposed. Such operator modification allows to search the optimal values, excluding the loss of best solutions acquired during the search.

6 Experiments of initialization methods for initial population in evolutionary algorithm

The statistical information on emissions of pollutants and carbon dioxide into the atmosphere from stationary sources and information on the morbidity on such indicators as the number of cases of cardiovascular diseases (registered in outpatient facilities), the number of new cases of tuberculosis and the number of registered cases of cancer was used to develop and test the models. Observation period - from 1990 to 2015, broken down by year and region of Ukraine [2].

Originally, a simple model was built, which has seven neurons in outermost layer (by the number of input parameters), one hidden fully connected layer with 12 neurons and output layer with one neuron.

As a network metrics MAE (Mean Absolute Error) was used - an average absolute error [13-25].

The results of learning and functioning of created network shown in Figure 1.

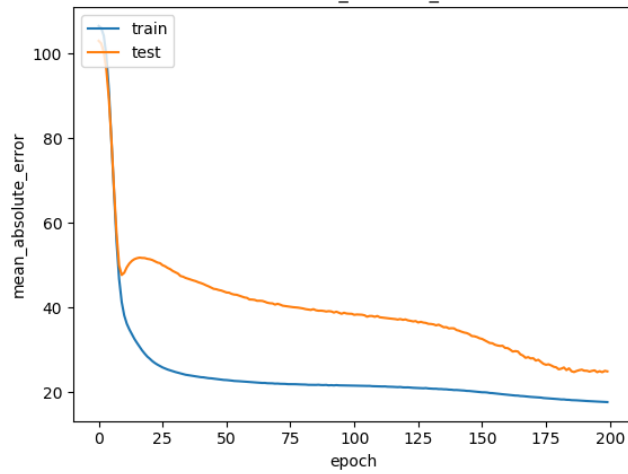


Fig. 1. The value of network metric (mae) for the indicator "Number of new cases of tuberculosis"

In the above graph it's clearly visible, that there is a gradual error decrease during network training. Apparently, the training reaches local extremum around 10-15 epoch. Further gradual decrease of network errors may indicate that the minimum is local. Therefore in this case the further model training is appropriate.

The learning of a network with two hidden layers was conducted for 100 epochs (sub-sample size 75) and validation sample split equal to 0.1. The results of network learning are shown in Figure 2.

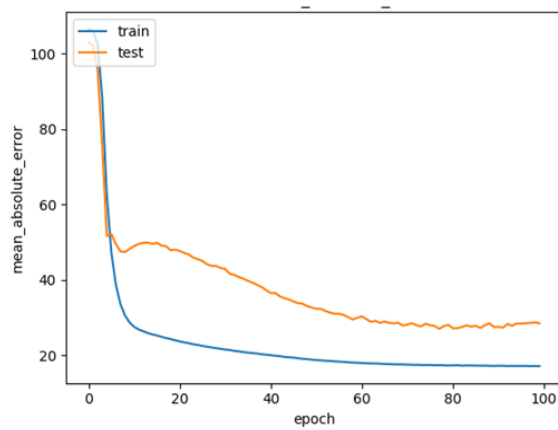


Fig. 2. The value of network metric (mae) for the indicator "Number of new cases of tuberculosis"

In this case, there is a similar pattern to the previous model – a gradual decrease of the error during network training. At the end of training there is a network conver-

gence. As in the previous case, for the "number of new TB cases" indicator there is a local minimum of error. One of the methods of preventing the effect of neural network overtraining is a method of exclusion (Dropout) [14], which is the exclusion of certain neurons from a network during the learning process. In previously created model the exclusions were added after the first hidden layer (50% of neurons excluded). The results of network learning are shown in Figure 3.

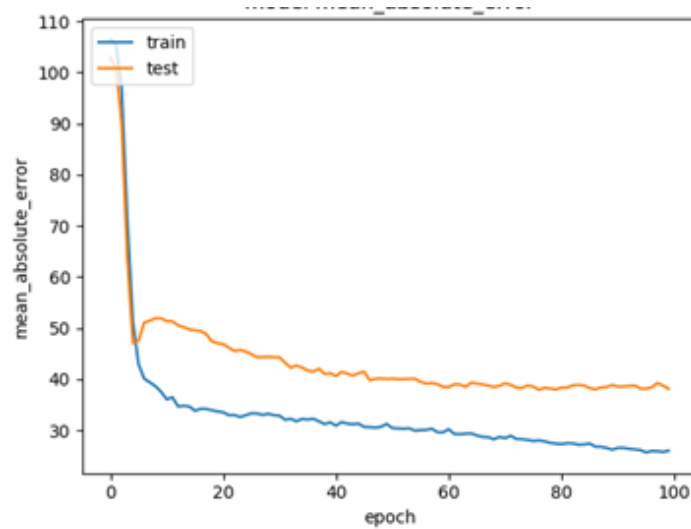


Fig. 3. The value of network metric (mae) for the indicator "Number of new cases of tuberculosis"

As in previous cases, during training there is a network convergence, but a little earlier - at about 60–70 epoch of learning. In all three cases for the indicator "number of new TB cases" there is a local minimum of network errors, which obviously is a feature of a model based on multilayer perceptron for this morbidity indicator. In addition, by using dropout the training curve loses smoothness and becomes jagged [15].

The presented data can be viewed as a time series, meaning that the parameter values are changing over time. For analysis and prediction of time series the models based on long short-term memory neural networks [15] can be used.

Network with LSTM layer receives eight input parameters. Hidden LSTM layer consists of twenty neurons and output layer has one neuron. The test results of the model are shown in Figure 4.

Figure 4 shows the change in value of network errors during training. For the indicator "The number of new cases of tuberculosis" and "number of cases of cardiovascular diseases" the local minimum is achieved with subsequent release of it. At the end of training there is no further reduction of model error value, so we can assume that during training the global error minimum was achieved and the network is considered as trained. Table 1 shows the comparison of resulting values of the average

absolute error (MAE), received during tests of different types of models (logistic regression, multilayer neural network models and so on.), which were created during the study. Thus, we can see that the smallest error of all prediction methods used is given by the model based on short long-term memory artificial neural network.

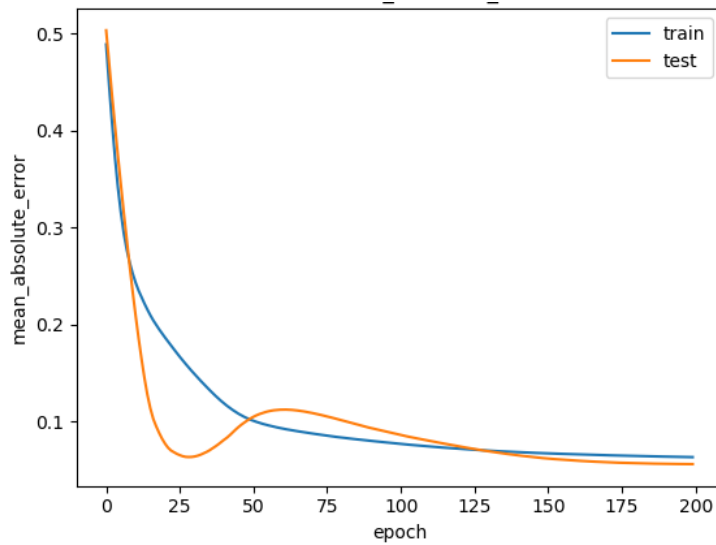


Fig. 4. The value of LSTM network metric (mae) during training for the indicator "number of cases of TB"

During the study the particle swarm optimization was used to optimize long short-term memory network.

As a result of such optimization the following optimal parameters of the network were obtained:

- the number of network nodes - 1000;
- Optimizer - Adadelata;
- subsample size - 1;
- number of training epochs - 100.

The long short-term memory network with these parameters using the test sample has an error value (RMSE) 127,08735307850266.

Table 1. The value of the average absolute error during testing of models created in the study of the problem

Predicted parameter	Type of predictive models								
	Logistic regression	The method of support vectors	The method of least squares	Random forest	Nearest Neighbor Algorithm	Multilayer perceptron with one hidden layer (128 neurons)	Multilayer perceptron with three hidden layers (128, 1024 and 128 neurons)	Multilayer perceptron with three hidden layers and dropout (128, 1024, dropout (0.5), 128 neurons)	Long short-term memory network of 50 nodes of LSTM layer
The number of new TB cases	29,764	40,271	22,957	7,671	156,387	23,445	21,675	22,3047	6,139
The number of cases of cardiovascular diseases	3814.174	2794.433	1789.727	571,018	573,004	1766.470	1752.416	1620.676	441,889
The number of reported cases of cancer	1400.357	1050.690	367.381	156,387	210,157	336,419	338,953	272,465	226,096

7 Conclusions

The available official statistical data on emissions of pollutants into the atmosphere from stationary sources and dynamics of population morbidity, the number of TB cases, the amount of cardiovascular diseases found in an outpatient setting and cancer morbidity in different regions of Ukraine was analyzed.

Methods of constructing mathematical models of the dependencies of these morbidity indicators on the volume of pollutant emissions were investigated.

Comparative analysis of the created models has shown that the best results were achieved in the models based on long short-term memory neural networks (LSTM).

When creating and training the model based on long short-term memory neural network the possibility of using particle swarm optimization and genetic algorithm to optimize network parameters was investigated and developed a modification of the classical method and modification of mutation operator.

Two modifications of genetic method were developed: first one using diploid set of chromosomes and second one with modified mutation operator. The first modification is the use of not one chromosome in population individuals karyotype, but a pair of homologous chromosomes, i.e. diploid set of chromosomes. The individual karyotype is a set of chromosomes, which is specific to certain type of individuals, that is characterized by a certain number of chromosomes and their structural features. Individual phenotype is determined by one of the allelic gene selected randomly. Using this modification allows to support a relatively large variability in population features during evolution, creating the potential to overcome the likely local extremums.

It was also developed the modification of mutation operator, which was never used before. Unlike the classical method, individuals which are exposed to the mutation operator are selected not randomly, but according to their mutational persistence, which corresponds to fitness function of individual. Thus the "weaker" individuals mutate and the genome of "stronger" ones remains unchanged. In this case, the probability of losing functions' extremum (which was achieved during evolution) due to mutation operator is reduced and the transition to the new extremum is carried out in case of accumulating sufficient weight of the "best" traits in a population.

Thus, these results lead to the conclusion that the proposed model based on long short-term memory neural network and modification of the classical method and mutation operator is feasible and effective solution to establish mathematical relationships of health indicators from pollutants. Practical use of the developed methods will allow timely adjustments of planned medical diagnostic, preventive measures, early determination of the necessary resources to contain and eliminate diseases to preserve population health.

References

1. Schmidhuber, J., Wierstra, D., Gagliolo, M., Gomez, F.: Training Recurrent Networks by Evolino. *Neural computation*, vol. 19(3), pp. 757–779 (2007). doi: 10.1162/neco.2007.19.3.757.

2. Lugovskaya, A. M.: Artificial neural networks in medical diagnostics. Computer systems and networks: proceedings of the 54th scientific conference of postgraduates, undergraduates and students 2018, vol. 1, pp. 182 – 183. BGUIR, Minsk (2018).
3. Siettos, C., Russo, L.: Mathematical modeling of infectious disease dynamics. *Virulence*, vol. 4, no. 4, pp. 295-306 (2013). DOI: 10.4161/viru.24041.
4. Cantu-Paz, E.: Solving Travelling Salesman Problem with an Improved Hybrid Genetic Algorithm. *Journal of Computer and Communications* 4(15), pp. 99–110 (2016).
5. Anisur, R., Zahidul, I.: A hybrid clustering technique combining a novel genetic algorithm with K-Means. *Knowledge-Based Systems*, vol. 71, pp. 345–365 (2014).
6. Haykin, S.: *Neural Networks and Learning Machines*. Prentice Hall Internat, Upper Saddle River (1999).
7. Pratyay, K., Prasanta, K.: A novel differential evolution based clustering algorithm for wireless sensor networks. *Applied Soft Computing*, vol. 25, pp. 414–425 (2014).
8. Shkaruplyo, V., Skrupsky, S., Oliinyk, A., Kolpakova, T.: Development of stratified approach to software defined networks simulation. *EasternEuropean Journal of Enterprise Technologies*, vol. 89, issue 5/9, pp. 67–73 (2017). Doi: 10.15587/1729-4061.2017.110142.
9. Fedorchenko, I., Oliinyk, A., Stepanenko, A., Zaiko, T., Shylo, S., Svyrydenko, A.: Development of the modified methods to train a neural network to solve the task on recognition of road users. *EasternEuropean Journal of Enterprise Technologies*, issue 9/98, pp. 46–55 (2019). DOI: 10.15587/1729-4061.2019.164789.
10. Oliinyk A., Fedorchenko I., Stepanenko A., Rud M., Goncharenko D.: Evolutionary method for solving the traveling salesman problem. *Problems of Infocommunications. Science and Technology: 5th International Scientific-Practical Conference PICST2018, Kharkiv, 9–12 October 2018, Kharkiv, Kharkiv National University of Radioelectronics*, pp. 331-339 (2018). Doi: 10.1109/INFOCOMMST.2018.8632033.
11. Sanches, D., Whitley, D.: Improving an exact solver for the traveling salesman problem using partition crossover. In: *Proceedings of the Genetic and Evolutionary Computation Conference, ACM New York*, pp. 337–344 (2017). DOI: 10.1145/3071178.3071304.
12. Hussain, A., Muhammad, Y.: Genetic Algorithm for Traveling Salesman Problem with Modified Cycle Crossover Operator. *Computational Intelligence and Neuroscience*, vol. 2017, pp. 1–7 (2017). DOI: 10.1155/2017/7430125.
13. Buse, M., Mutlu, A.: Logarithmic learning for generalized classifier neural network. *Neural Networks*, issue 12/60, pp. 133–140 (2014). DOI: 10.1016/j.neunet.2014.08.004.
14. Garcia, T., Cancelas, N., Soler-Flores, F.: The artificial neural networks to obtain port planning parameters. *Procedia-Social and Behavioral Sciences*, issue 19/162, pp. 168–177 (2014). DOI: 10.1016/j.sbspro.2014.12.197.
15. Alam, S., Dobbie, G., Koh, Y., Riddle, P., Rehman, S.: Research on Particle Swarm Optimization based clustering: a systematic review of literature and techniques. *Swarm and Evolutionary Computation*, issue 17/2, pp. 1–13 (2014). DOI: 10.1016/j.swevo.2014.02.001.
16. Oliinyk, A., Fedorchenko, I., Stepanenko, A., Rud, M., Goncharenko, D.: Combinatorial optimization problems solving based on evolutionary approach. In: *2019 15th International Conference on the Experience of Designing and Application of CAD Systems (CADSM)*, pp. 41-45 (2019). DOI: 10.1109/CADSM.2019.8779290.
17. Fedorchenko, I., Oliinyk, A., Stepanenko, A., Zaiko, T., Svyrydenko, A., Goncharenko, D.: Genetic method of image processing for motor vehicle recognition. In: *CEUR Workshop Proceedings 2353*, pp. 211-226 (2018). ISSN: 16130073.

18. Oliinyk, A., Zaiko, T., Subbotin, S.: Training sample reduction based on association rules for neuro-fuzzy networks synthesis. *Optical Memory and Neural Networks (Information Optics)*, vol. 23(2), pp. 89-95 (2014). doi: 10.3103/S1060992X14020039.
19. Oliinyk, A., Subbotin, S.: A stochastic approach for association rule extraction. *Pattern Recognition and Image Analysis*, vol. 26, no. 2, pp. 419-426 (2016). doi: 10.1134/S1054661816020139.
20. Oliinyk, A., Zayko, T., Subbotin, S.: Synthesis of Neuro-Fuzzy Networks on the Basis of Association Rules. *Cybernetics and Systems Analysis*, vol. 50, no. 3, pp. 348-357 (2014). DOI: 10.1007/s10559-014-9623-7.
21. Kolpakova, T., Oliinyk, A., Lovkin, V.: Improved method of group decision making in expert systems based on competitive agents selection 2017. 2017 IEEE 1st Ukraine Conference on Electrical and Computer Engineering, UKRCON 2017, pp. 939-943 (2017). doi: 10.1109/UKRCON.2017.8100388.
22. Stepanenko, O., Oliinyk, A., Deineha, L., Zaiko, T.: Development of the method for decomposition of superpositions of unknown pulsed signals using the second-order adaptive spectral analysis. *EasternEuropean Journal of Enterprise Technologies*, vol. 2, no 9, pp. 48-54 (2018). doi: 10.15587/1729-4061.2018.126578.
23. Oliinyk, A.O., Oliinyk, O.O., Subbotin, S.A.: Software-hardware systems: Agent technologies for feature selection. *Cybernetics and Systems Analysis*, vol. 48(2), 257-267 (2012). doi: 10.1007/s10559-012-9405-z.
24. Oliinyk, A.A., Subbotin, S.A.: The decision tree construction based on a stochastic search for the neuro-fuzzy network synthesis. *Optical Memory and Neural Networks (Information Optics)*, vol. 24(1), 18-27 (2015). doi: 10.1007/s10559-012-9405-z.
25. Oliinyk, A., Subbotin, S., Lovkin, V., Leoshchenko, S., Zaiko, T.: Development of the indicator set of the features informativeness estimation for recognition and diagnostic model synthesis. 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET 2018), pp. 903-908, (2018). doi: 10.1109/TCSET.2018.8336342.