ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ ПРОБЛЕМИ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ ПІДТРИМАННЯ ПРИЙНЯТТЯ РІШЕНЬ

UDC 004.02, 004.67, 004.891.3
DOI: 10.20535/SRIT.2308-8893.2024.2.04

A HYBRID MODEL OF ARTIFICIAL INTELLIGENCE INTEGRATED INTO GIS FOR PREDICTING ACCIDENTS IN WATER SUPPLY NETWORKS

Yu. ZAYCHENKO, T. STAROVOIT

Abstract. The search for an effective and reliable model for predicting accidents on water supply networks by determining their exact locations has always been important for effectively managing water distribution systems. This study, based on the adaptive neuro-fuzzy logical inference system (ANFIS) model, was developed to predict accidents in the city of Kyiv (Ukraine) water supply network. The ANFIS model was combined with genetic algorithms and swarm optimization (ACO) methods and integrated into a GIS to visualize results and determine locations. Forecasts were evaluated according to the following criteria: mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2). Depending on the amount and type of input data, ANFIS optimization with genetic algorithms and swarm optimization (ACO) can, on average, increase the accuracy of ANFIS predictions by 10.1% to 11%. The obtained results indicate that the developed hybrid model may be successfully applied to predict accidents on water supply networks.

Keywords: ANFIS, ACO, GA, spatial objects, geodatabase, metaheuristics, spatio-temporal analysis, water loss.

INTRODUCTION

Forecasting accidents in water distribution systems is important in the management of water resources, as it makes it possible to identify problem areas in the network and eliminate them in advance. Intelligent predictive systems are models and algorithms that provide valuable information about the future performance of a system as a decision support system. With the development of supervisory control and data acquisition (SCADA) systems, real-time monitoring of pressure and data flows is commonly used to detect pipe bursts. Machine learning [5] and cluster analysis models were developed for optimal assessment. Failures in the network can also be analyzed using hydraulic models [6].

The techniques mentioned above were successful in detecting accidents, but not in pinpointing their exact locations [7]. The model-based approach relies heavily on the accuracy of hydraulic models [8] and may not be suitable for larger water supply systems. Other methods that utilize pressure/flow measurements and GIS have also been proposed. For instance, [9] utilized triangle-based cubic interpolation to establish a pressure drop surface during network breaks to locate the source of the problem. In [10] the measuring zone's rupture location in the water supply network was identified by assessing the sensitivity of various pressure/flow measurements in relation to emergency leaks. [11] employed a multivariate graphical model that utilized data from multiple pressure gauges to identify potential accident locations, employing a combination of Gaussian and geostatistical methods. Typically, fluctuations in demand can make it difficult to detect hydraulic indicators resulting from accidents. Therefore, these methods can only provide a general idea of where network breaches may occur, with an error range of hundreds of meters and several pipes. Unfortunately, this is not precise enough to quickly locate and fix network issues, resulting in delayed system restoration.

A more accurate method is needed to locate pipe bursts, which involves gathering detailed information about the water system's behaviour in potential locations to detect anomalies. This can be achieved by placing accelerometer sensors and analyzing acoustic signals, which can automatically determine the location of the rupture or leak [12]. However, the reliability of this method depends on the characteristics of the leakage conditions, such as pressure and flow rate, and the detection range is limited by the clarity and correlation of the acoustic signals. Another approach is based on transient processes [13], which analyzes characteristic transient waves to determine the location of accidents. However, background noise or other activities in the system can interfere with transient signals caused by discontinuity, especially if the number of channels to be analyzed increases [7]. Hence, methods based on transient processes may not be suitable for locating pipe breaks with exact precision.

Many researchers have explored the use of machine learning in water resources research [14], but there is no consensus on the best model for predicting water supply network emergencies. To address this, a forecasting model was developed that can pinpoint the exact location of potential emergencies. Artificial neural networks are commonly used in water resource assessment due to their computational efficiency [15–17], but they may produce errors in some cases due to poor prediction or overtraining [15]. Therefore, it is necessary to optimize the ANN and look for new approaches and new classes of neural networks.

Studies [18–20] have proposed a high-precision hybrid model called ANFIS, which combines artificial neural networks (ANN) and fuzzy logic. The hybrid ANFIS model has better performance than the two separate models, but it has certain limitations in finding the best weight parameters, which greatly affect the prediction performance [15]. Furthermore, different optimization algorithms yield varying results based on the geoenvironmental factors of the area being studied. Therefore, developing new hybrid algorithms to determine the best weights and produce reliable results is fundamental for flow modeling processes.

The purpose of this work is the development of a new model of artificial intelligence and the study of its effectiveness in the tasks of predicting accidents on water supply networks with the determination of exact locations. This research is conducted for the first time on the water distribution system of the city of Kyiv.

MODEL DEVELOPMENT AND TRAINING METODOLOGY

Data set collection for spatial modeling

The proposed modeling method is applied to the GIS water supply system of the city of Kyiv (Ukraine). The length of the water supply networks in the city is increasing due to the inclusion of street and intra-quarter networks from enterprises. As of 2019, the total length of the networks was 4,284.8 km.

In the structure of the city's water supply networks, the main part is street networks — 2614.8 km or 61% of the total length of pipes; intra-quarter networks — 1275.1 km or 29.8%; water pipes — 394.9 km or 9.2%. The vast majority of pipelines, namely 65.9%, are made of cast iron; 30.5% — from steel and only 3.6% — from plastic materials.

21.4% of the pipes of the water supply network have been operated for more than 50 years and another 33.2% — about 50 years; the service life of 27.1% is up to 35 years, 12.3% — up to 25 years, 4.6% — up to 15 years, and only 1.4% — up to 5 years. According to the degree of wear, 50.4% of the pipes are worn by more than 90%; 24.3% of pipes — by 50–75%; 15.5% — by 75–90%; 6.3% — by 25–50%; 3.5% — less than 25%.

Pipelines made of cast iron have the longest average age — 46.8 years, pipelines made of steel — 45.4 years, the smallest — made of plastic — 15 years. According to the pipe depreciation indicator, the water distribution system is characterized as follows: the average degree of wear of steel pipes is 90%, cast iron pipes are 75%, and plastic pipes are 23%.

The accident rate, which is determined by the number of accidents per unit length of the network, has fluctuated in the range of 2.0–2.2 accidents/km in recent years, and the tendency to increase the number of accidents was observed specifically for water pipes.

The methodology of this study is shown in Fig. 1, and includes the following stages:

1) preparation of input data;

2) separation of data into training (70%) and test (30%) sets;

3) training of ANFIS neuro-fuzzy network;

4) optimization of the ANFIS model by genetic algorithms and the swarm optimization algorithm (ACO);

5) checking the accuracy of ANFIS, ANFIS-GA and ANFIS-ACO models.

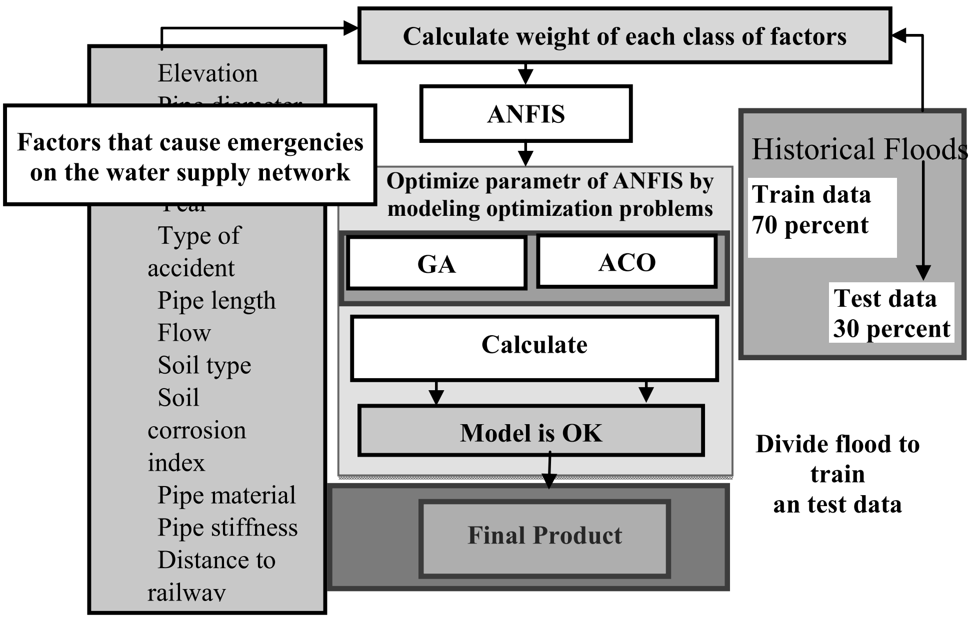

Fig. 1. Structural diagram of the development and optimization of the ANFIS model

It is important to consider how the problem occurs in relation to other factors to make accurate spatial predictions. Table 1 shows the data used in our predictive model, with some entered into the GIS and the rest determined through hydraulic modeling based on the GIS model.

Table 1. Factors and conditions used in the model that impact the emergence of issues in the water supply network

Factors/conditions	Units	Description
The degree of proximity of the location to railway tracks/	m	When trains are in motion, the ground vibrates, causing pipes to crack and gate valves to be damaged.
Age	year	Year of laying the pipe
Length	m	Length of a pipe
Diameter	mm	Size of a pipe
Soil type index	NA	Soil type
Geoposition	NA	Geospatial location
Accident date	year	Accident date on network
Pressure	bar	Pressure from hydraulic calculation results
Volume of consumption	m^3/hour	Volume of water consumption per hour
Volume of consumption	m^3/month	Volume of water consumption per month
Demand	NA	Water demand
flow rate	NA	Flow rate according to hydraulic calculation
Pipe materials rigidity	NA	rigidity coefficient
Consumers	NA	Individuals and legal persons

It is probable that certain factors may affect the occurrence of pipe ruptures or damages in specific parts of the network, while leaving other areas unaffected. One such factor could be the presence of railway tracks. The vibrations caused by freight trains passing by can lead to frequent failures in the water supply network, resulting in pipe ruptures or damage to fittings. Additionally, the type of pipe material used also plays a significant role in determining its lifespan. Steel pipes typically last for 25 years, while plastic or cast iron pipes can last up to 50 years.

PREPARATION OF DATA SET FOR TRAINING AND TESTING

In order to check if the model is practical, the data set for analysis should be split into two groups: one for building the model (called the training data set) and the other for testing it (called the test data set) [21]. To create the training data set, 70% of locations with and without previous accidents on the network (a total of 313 locations) were randomly chosen and combined.

The remaining 30% were then used to create the test dataset. Both data sets were originally in vector format but were converted to csv format for further analysis. For both data sets, the value 1 was assigned to indicate the presence of an accident on the network, while 0 was assigned to indicate the absence of accidents.

We conducted a statistical analysis to thoroughly examine the data and improve the intelligent model.

We performed a statistical analysis of spatial data in order to determine the parameters of the membership function for training the ANFIS network and its optimization (Fig. 2).

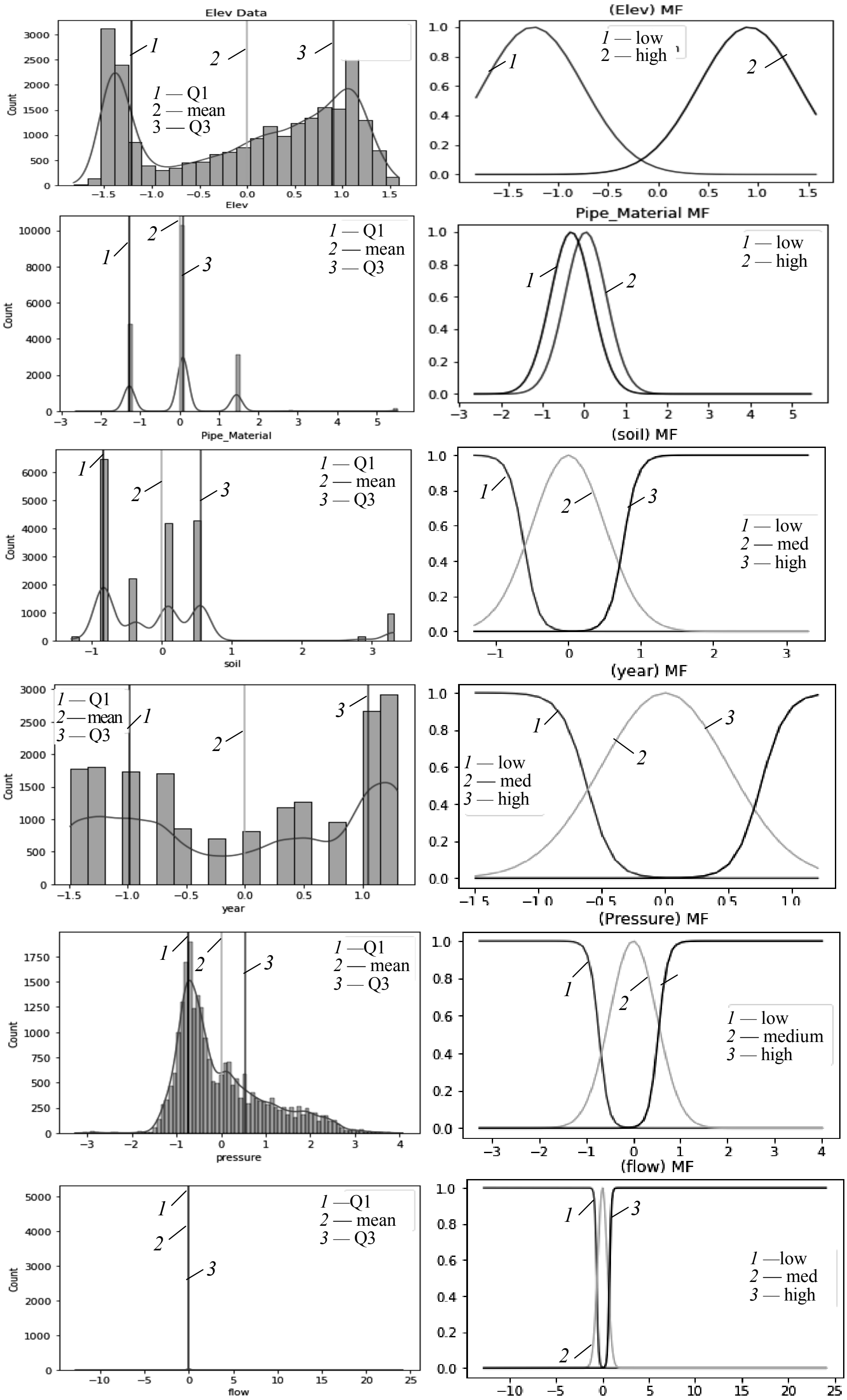

Fig. 2. Statistical analysis of the spatial data

DEVELOPMENT OF THE ANFIS MODEL AND ALGORITHMS FOR ITS OPTIMIZATION

Adaptive neuro-fuzzy logical inference system

ANFIS (Adaptive Network Based Fuzzy Inference System) is an adaptive fuzzy logical inference system proposed by Sugeno based on IF-THEN rules. It is a method that combines artificial neural networks (ANNs) with fuzzy ones. This neural network is used for membership function tuning and rule base tuning in a fuzzy expert system. Below is the Sugeno model of fuzzy logic inference (Fig. 3).

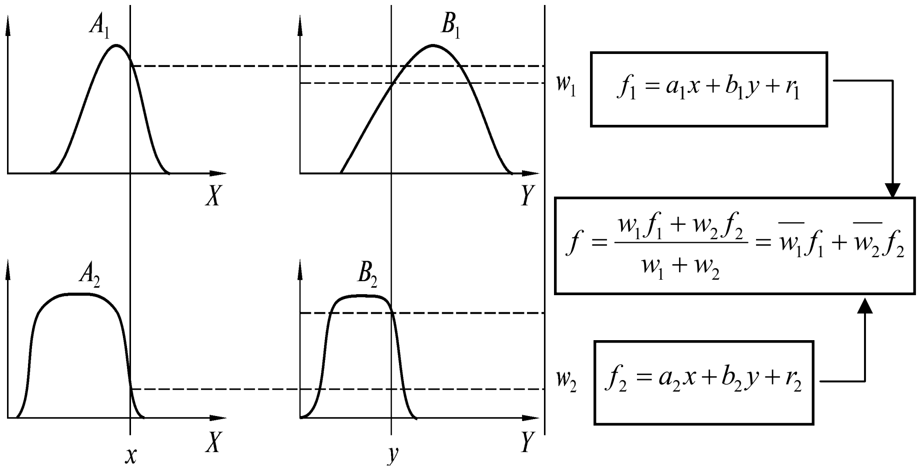

Fig. 3. Sugeno's fuzzy logic model

The layers of this fuzzy neural network perform the following functions.

Layer 1. Membership Function Layer

In this layer, each neuron uses a membership function (fuzzifier) to transform the input signal x or y. The most commonly used functions are the bell-shaped function or the Gaussian function:

$$\mu_{A_i}(x) = \frac{1}{1 + \left[\left(\frac{x - c_i}{a_i}\right)^2\right]};$$

$$\mu_{A_i}(x) = \exp\left[-\left(\frac{x - c_i}{a_i}\right)^2\right].$$

Layer 2. Antecedent Layer

Each neuron is represented by the symbol Π. It performs an intersection between sets of input signals, which simulates a logical AND operation. The neuron then sends an output:

$$w_i = \mu_{A_i}(x) \times \mu_{B_i}(y), \quad i = i = 1,2,..,n.$$

In fact, any *T*-norm operator that generalizes the AND operation can be used in these neurons.

Layer 3. Normalization Layer

Each neuron in this layer calculates the normalized strength of the rule:

$$\underline{w_i} = \frac{w_i}{\sum_i w_i}, \quad i = 1,2,..,n .$$

Layer 4. Consequent Layer

The values of output variables are formed in neurons:

$$O_i^4 = \underline{w_i} f_i = \underline{w_i}(a_i x + b_i y + r_i) .$$

Layer 5. Aggregation Layer

We receive the output signal of the neural network and perform defuzzification of the results:

$$O^5 = \text{overall output} = \sum \underline{w_i} f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} .$$

The neural network of the ANFIS architecture is trained using the gradient descent method.

OPTIMIZATION OF ANFIS WEIGHTING COEFFICIENTS AND OFFSETS BY THE ANT COLONY ALGORITHM

Ant Colony Optimization (ACO) is a probabilistic method for solving complex computational problems that find optimal parameters in a search environment. This algorithm, which was proposed by Marcus Dorigo in 1996, imitates the behaviour of ants in finding the optimal path from their nest to a food source. In [22; 23], the author optimizes the weighting coefficients of an artificial neural network using ACO and investigates the performance of the network. In the search space, a population of weights is created which is considered as an objective function and is found according to the formula:

$$\text{SEP} = 100 \frac{O_{\max} - O_{\min}}{n_o n_p} \sum_{p=1}^{n_p} \sum_{i=1}^{n_o} (t_i^p - o_i^p)^2 ,$$

where t_i^p and o_i^p are the expected and actual value of the output neuron and for the template p.

The terms $O_{\max}$, and $O_{\min}$ represent the highest and lowest values of the output signal from a specific neuron, while n_o and n_p refer to the number of output neurons.

The ACO algorithm is a tool for optimizing neural network parameters such as synaptic weights, number of layers, and number of hidden neurons. It begins by randomly selecting decisions from a predefined set of data, which are then evaluated and assigned to the decision space based on their fitness values. New solutions are created using information from previous iterations, with a higher likelihood of selecting values with a greater concentration of pheromones [23]. This process generates a matrix of size $M \times N$, where M represents the decision population size and N represents the number of decision variables.

$$\text{Population} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_j \\ \vdots \\ X_M \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1i} & \dots & x_{1N} \\ x_{21} & x_{22} & \dots & x_{2i} & \dots & x_{2N} \\ \dots & \dots & \dots & \vdots & \dots & \dots \\ x_{j1} & x_{j2} & \dots & x_{ji} & \dots & x_{jN} \\ \dots & \dots & \dots & \vdots & \dots & \dots \\ x_{M1} & x_{M2} & \dots & x_{Mi} & \dots & x_{MN} \end{bmatrix},$$

where $X_j = j$ -th solution, x_{ji} – i-th solution variable for the j-th solution, and M is the size of the number of solutions. The value x_{ji} is chosen randomly from the set V_i :

$$V_i = \{v_{i1}, v_{i2}, ..., v_{id}, v_i D_i\}\ ,\ i = 1, 2, ..., N\ ,$$

where V_i = set of predefined values for the i-th decision variable, $v_{id} = d$ -th possible value for the i-th decision variable, and D_i = total number of possible values for the i-th decision variable [23].

GENETIC ALGORITHMS

Genetic algorithms develop optimal solutions by sampling from all possible solutions. The best of these solutions are then combined using the genetic operators of crossover and mutation to generate new solutions. This process continues until a certain termination condition is met [4]. The diagram of the GA process is shown in Fig. 4. The first step is the initial state in which we want to find the Hamiltonian cycle with the smallest sum of weights. In the second step, the fitness function estimates the Hamiltonian cycles, with lower cost functions indicating the best individuals. Finally, in the third step, the most adapted individual is identified.

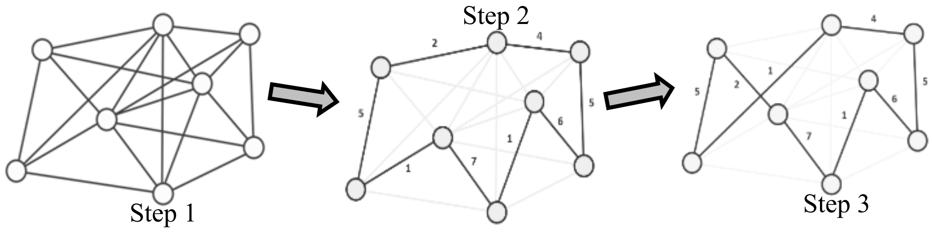

Fig. 4. Scheme of the process of genetic algorithms

GA can be used to optimize various parameters in water distribution systems. It uses the following mechanisms: crossover, mutation, selection. The goal of training is to minimize the root mean square error:

$$E(W) = \frac{1}{M}\sum_{k=1}^{M}(d_k - y_k(w))^2\ ;$$

$$W = [W_I, W_O]\ ;$$

$$W_I = \left\| w_{ij}^I \right\|;$$

$$W_O = \left\| w_{ij}^O \right\|.$$

We set the initial population in which any individual is represented by the corresponding weights of N individuals: $[W_I(0),...,W_i(0),..., W_N(0)]$.

We calculate the fitness index (Fitness Index) and evaluate the quality of forecasting:

$$FI(W_i) = C - E(W_i) \to \max,$$

where C — constant.

IMPLEMENTATION OF THE MODEL

The technique of forecasting with a combination of GIS and artificial intelligence methods were applied to predict accidents on the water supply network of the city of Kyiv. The Sugeno method was used, as it shows better accuracy. The optimal membership function was chosen by trial and error. The ANFIS method was optimized by GA and ACO to improve accuracy.

The performance of the ANFIS, ANFIS-GA, ANFIS-ACO models was determined from the resulting mean absolute error (MAE), which indicates a risk metric corresponding to the expected value of the absolute error loss or the loss rate:

$$MAE(y,\hat{y}) = \frac{1}{n_{samples}} \sum_{i=1}^{n_{samples}} |y_i - \hat{y}_i|.$$

The mean squared error indicates the risk indicator corresponding to the expected value of the squared error or loss:

$$RMSE(y,\hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}} (y_i - \hat{y}_i)^2.$$

The R^2 function calculates the coefficient of determination, which represents the proportion of variance (y) that was explained by the independent variables in the model:

$$R^2(y,\hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \underline{y})^2}.$$

The function explained variance calculates the estimate of the explained variance:

$$\text{explained\_variance } (y,\hat{y}) = 1 - \frac{\text{var}\{y,-\hat{y}\}}{\text{var}\{y\}}.$$

RESULTS

Spatial-temporal assessments and prediction of the occurrence of accidents on the water supply network of the city of Kyiv

Spatiotemporal GIS analysis and modeling are essential for studying and predicting future events. For modeling, we used the ESRI GIS package: ArcGIS Pro 2.7. The first step was data acquisition and preparation. The obtained information was summarized in the netCDF data structure, which was used for spatial statistical analysis and creation of a space-time cube (Fig. 5) [24] .

Fig. 5. The result of spatial forecasting

A space-time cube is a well-known model in ArcGIS that combines spatial data and time into a three-dimensional data structure of the netCDF (total network shape) format, containing an array of bins with absolute location and absolute time [24]. So, we aggregated incidents of accidents on the water supply network within a grid size of 500 × 500 m^2 (distance interval) with an absolute step interval of 1 month. This approach made it possible to investigate cases of accidents on the water supply network of the city of Kyiv (Ukraine).

We applied the space-time cube to a forest-based prediction model, which generated a 2D object class indicating the predicted locations within the original space-time cube. Each location is predicted individually (as shown in Fig. 5) and has its own schedule (as seen in Fig. 6).

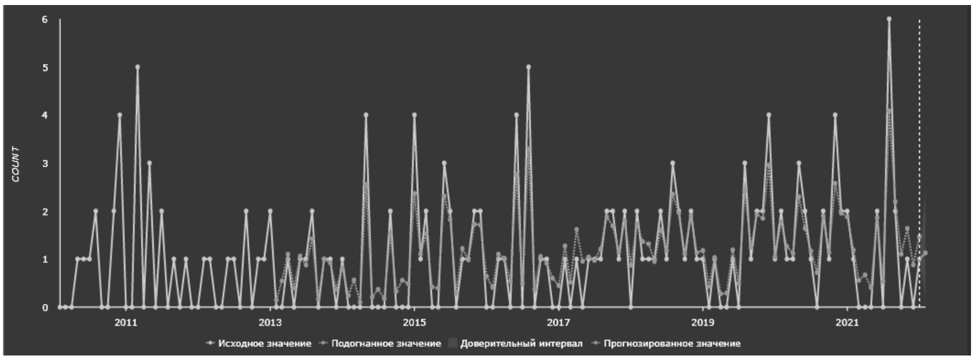

Fig. 6. Graph of the values of the locations of the space-time cube

In Fig. 6, the graph displays the input, data gaps restored as a result of calculations, predicted values and confidence intervals. Confidence intervals are created for each predicted time step, which are presented as fields of output objects.

The upper and lower bounds of the confidence intervals for the first predicted time step are calculated using quantile random forest regression. To predict values for a future time, observations from each leaf of the tree are averaged together. The confidence interval of the second forecast is calculated in a similar way, but is adjusted taking into account the confidence interval of the first forecast [25]. The real confidence interval of the second forecast is calculated by adding the lengths of the limits of the confidence interval of the two forecasts. Subsequent time steps are calculated by adding previous predictions. The real confidence level of these intervals is at least 90%, but in reality the accuracy may be higher [25].

The result of the assessment of the total accuracy of the forecast in different locations, using the forest-based method, is shown in Table 2.

Table 2. The result of the overall assessment of forecast accuracy in different locations

Category	Min	Max	Mean	Median	Mean sq. dev.
RMSE of the prediction	0.00	1.25	0.26	0.24	0.15
RMSE of validation	0.00	2.89	0.56	0.48	0.45

This forecasting method is best used for time series with a complex shape and trends that are difficult to model using simple mathematical functions. The correct selection of time steps during model validation is important. The more time steps that are excluded, the less time it takes to evaluate the validation model. However, if too few time steps are included, the RMSE value will be estimated using less data and may be misleading. Also, this tool can produce unstable and unreliable forecast results if the same value is repeated too often in time series [25]. To optimize and improve the accuracy of the predictive model, we combined GIS methods with hybrid artificial intelligence methods.

Configuration of hybrid models

In this study, we integrated the ANFIS model with GA and ACO algorithms, and compared the performance of the models. The algorithms are implemented in the Spyder environment (Anaconda 3). In order to test the model with different optimization algorithms, the data were organized into separate training and test datasets, which were divided into 70% and 30% (Fig. 7).

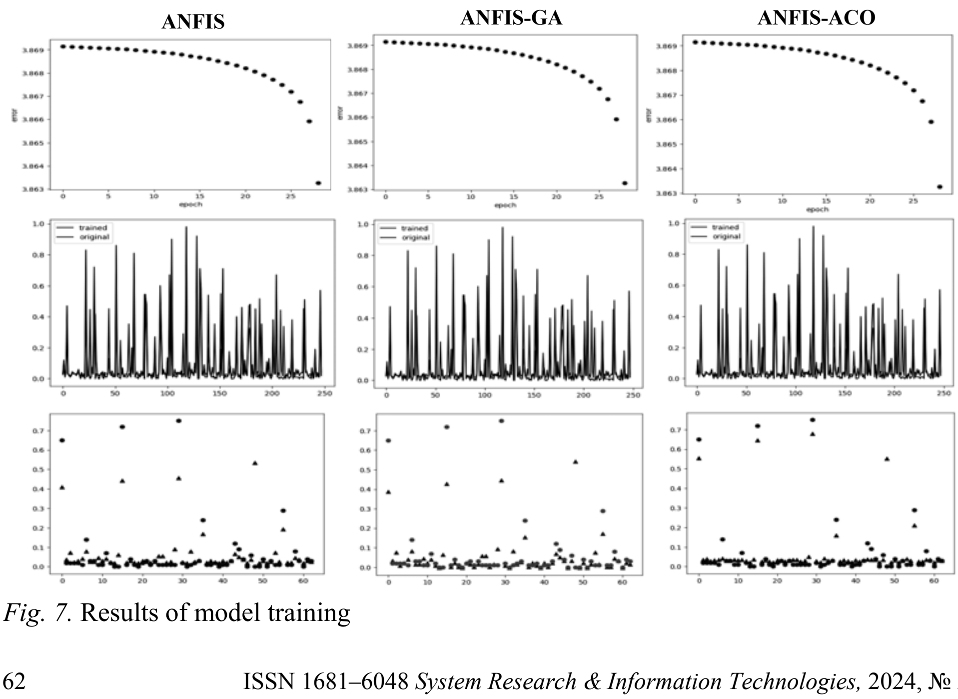

Fig. 7. Results of model training

Fig. 7 shows the result of model training: change in error frequency during training; comparing predicted values with actual data on the training set and comparing predicted values with actual data on the test set.

The first step was to import the training data into the ANFIS, ANFIS-GA, ANFIS-ACO models to reveal the hidden relationships between the factors affecting the emergency of the water supply network. As a next step, the validation data were used to test the performance and predictive capabilities of the models. MAE, RMSE, R^2, and explained_variance were used to measure accuracy. Table 3 shows the result of learning hybrid models (the first 5 iterations in GA and ACO).

T a b l e 3. Comparison and performance testing of models

Model	Test data				Train data			
	MAE	RMSE	R^2	Cov	MAE	RMSE	R^2	Cov
ANFIS	0.043	0.094	0.613	0.613	0.062	0.125	0.576	0.578
	Study time:				0:00:06.19			
ANFIS-GA	0.041	0.097	0.599	0.600	0.061	0.124	0.575	0.574
	Study time:				0:00:08.31			
	0.042	0.098	0.583	0.585	0.061	0.124	0.575	0.577
	Study time:				0:00:08.98			
	0.041	0.098	0.587	0.589	0.061	0.125	0.575	0.577
	Study time:				0:00:09.09			
	0.044	0.098	0.585	0.587	0.062	0.124	0.573	0.575
	Study time:				0:00:08.41			
	0.042	0.098	0.584	0.586	0.061	0.125	0.575	0.576
	Study time:				0:00:08.24			
ANFIS-ACO	0.041	0.096	0.593	0.595	0.061	0.124	0.573	0.575
	Study time:				0:00:11.96			
	0.042	0.097	0.585	0.587	0.061	0.124	0.574	0.576
	Study time:				0:00:11.86			
	0.043	0.098	0.585	0.586	0.062	0.125	0.572	0.575
	Study time:				0:00:12.22			
	0.041	0.098	0.586	0.588	0.061	0.125	0.576	0.577
	Study time:				0:00:11.92			
	0.042	0.097	0.585	0.587	0.062	0.125	0.573	0.575
	Study time:				0:00:12.21			

The MAE values for the ANFIS, ANFIS-GA, and ANFIS-ACO models were calculated for both the test and training data. The results show that the ANFIS-GA model had the best performance with a MAE value of 0.042 for the test data and 0.061 for the training data. The GA algorithm was found to be more efficient than the ACO algorithm, which had a similar performance but required twice as much training time. It's important to note that these results may vary based on the input data. Overall, the ANFIS-GA model is stable, efficient, and has a fast convergence rate.

CHECKING AND COMPARING MODELS

We used three different optimization models, namely ANFIS, ANFIS-GA, and ANFIS-ACO, which were developed and implemented in Spyder (Anaconda3).

The results obtained from these models were then visualized in ArcGIS Pro 2.7. To train these models, we divided the pointed objects of accidents into two categories: 30% for training and 70% for testing. We used the training data set to establish relationships between the occurrence of accidents (1) and the absence of accidents (0).

We checked the accuracy and performance of hybrid intelligent models by calculating the mean absolute error of MAE. Fig. 8 shows the membership functions of the input variables of the ANFIS model. Fig. 9 illustrates the graph of the change in the loss function depending on the number of iterations. Membership functions indicate the fuzziness of the inputs. A comparison of the accuracy scores in Fig. 8 shows that the ANFIS network performs well.

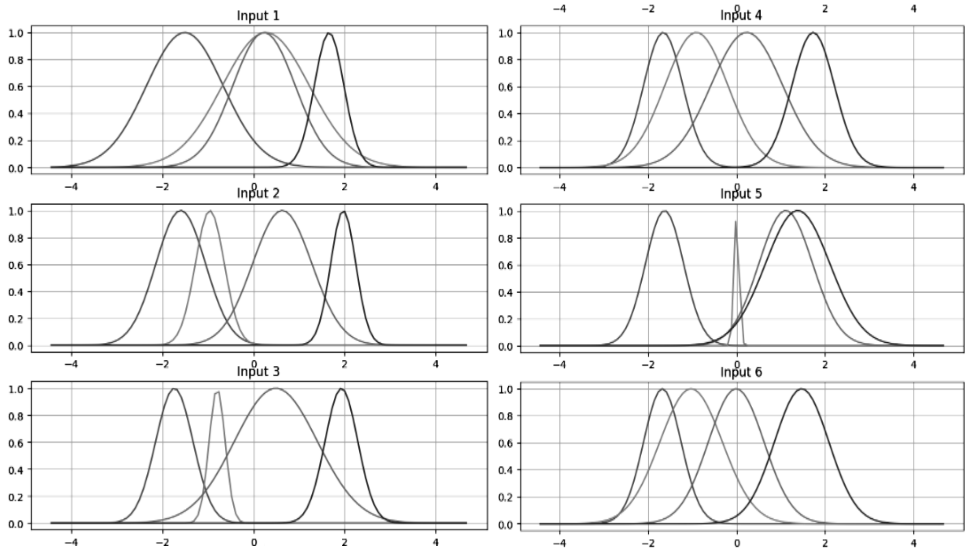

Fig. 8. Membership functions of the used input variables

As a result, the accuracy of the ANFIS model was 95.49%. The accuracy decreases when the number of inputs increases, so to increase the accuracy, it is necessary to improve the network with optimization algorithms.

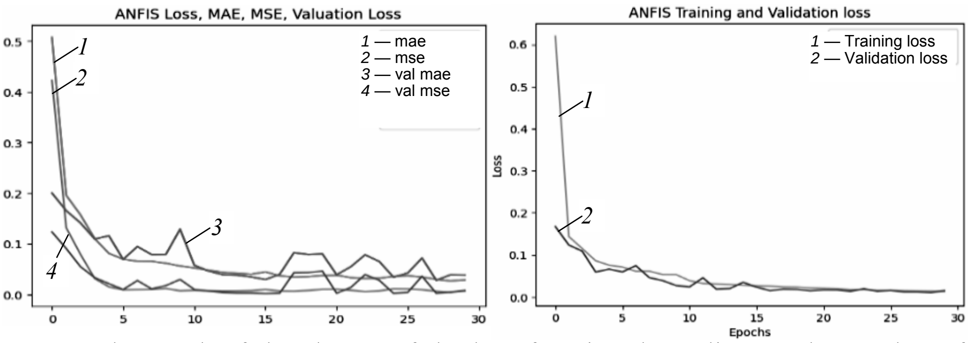

Fig. 9. The graph of the change of the loss function depending on the number of iterations

The result of training the ANFIS-GA and ANFIS-ACO models was not much better than the classic ANFIS, moreover, the ANFIS-ACO model required much more time. In ANFIS-GA, the training time was the same as in ANFIS (one iteration on average 0:00:06.24), while in ANFIS-ACO the total training time took 0:58:19.69 (0:00:12.38 one iteration). Overall, the predictions aligned well and matched the experimental data accurately. It's worth mentioning that the test results demonstrate the developed models' proficiency in forecasting data beyond the training range.

Compared to GIS forecasting methods, developed artificial intelligence models provide an opportunity to expand and increase forecast accuracy, and indicate specific problematic pipes. Also, the developed models can be easily integrated into ArcGIS Pro in the form of geoprocessing tools, and published on corporate geoportals.

CONCLUSIONS

Adaptive neural fuzzy logic inference system (ANFIS) and its hybrid learning methods: ANFIS-GA, ANFIS-ACO were used to predict water supply network accidents. This model was integrated into GIS (ArcGIS Pro) to visualize and determine the exact locations of possible accidents and was verified in practice (all predicted accident locations for the next three days coincided with accidents that occurred on the Kyiv water supply network). The following conclusions can be drawn from the forecasting model described above:

- Performance evaluation and model validation results of selected metrics: R2, RMSE, and MAE for both training and testing on a small amount of data showed that the hybrid models did not outperform ANFIS model.
- When the amount of input data increased, the accuracy of the ANFIS model decreased and it became necessary to optimize the ANFIS with genetic algorithms and the swarm optimization algorithm (ACO). This optimization increased the accuracy of ANFIS prediction by 10.1%, 11%.
- The results of ANFIS, ANFIS-GA, and ANFIS-ACO intelligent models combined with GIS indicate a large information potential that can support real-time operational control of water supply systems. Fuzzy models of emergency forecasts have a significant advantage as they require less information about water supply systems than conventional probabilistic models. In addition, this information may be vague and inaccurate. The ANFIS model is suitable for modeling complex problems, especially when the relationship between factors is unknown. It is especially useful for identifying threats and providing advance warnings about the likelihood of an accident.

REFERENCES

1. C.M. Fontanazza, V. Notaro, V. Puleo, P. Nicolosi, and G. Freni, "Contaminant intrusion through leaks in water distribution system: experimental Analysis," *Procedia Eng.*, 119, pp. 426–433, 2015. doi: 10.1016/j.proeng.2015.08.904.
2. "Urban Water Statistics Yearbook 2017," (in Chinese), *CUWA*.
3. Dung H. Tran, A.W.M. Ng, and B.J.C. Perera, "Neural networks deterioration models for serviceability condition of buried stormwater pipes," *Engineering Applications of Artificial Intelligence*, 20(8), pp. 1144–1151, 2007. doi: 10.1016/j.engappai.2007.02.005.
4. M. van Dijk, S.J. van Vuuren, J.E. van Zyl, "Optimising water distribution systems using a weighted penalty in a genetic algorithm", *Water SA*, 35(5), pp. 537–548, 2008. doi: 10.4314/wsa.v34i5.180651.
5. M. Romano, Z. Kapelan, and D.A. Savic, "Evolutionary algorithm and expectation maximization strategies for improved detection of pipe bursts and other events in

water distribution systems," *J. Water Resour. Plan. Manag.*, 140 (5), pp. 572–584, 2014. doi: 10.1061/(ASCE)WR.1943-5452.0000347.

6. J. Meseguer et al., "A decision support system for on-line leakage localization Environ," *Environmental Modelling & Software*, vol. 60, pp. 331–345, 2014. doi: 10.1016/j.envsoft.2014.06.025.
7. Y. Wu and S. Liu "A review of data-driven approaches for burst detection in water distribution systems," *Urban Water J.*, 14 (9), pp. 972–983, 2017. doi: 10.1080/1573062X.2017.1279191.
8. C.J. Hutton, Z. Kapelan, L. Vamvakeridou-Lyroudia, and D.A. Savic, "Dealing with uncertainty in water distribution system models: a framework for real-time modeling and data assimilation," *J. Water Resour. Plan. Manag.*, 140 (2), pp. 169–183, 2014. doi: 10.1061/(ASCE)WR.1943-5452.0000325.
9. S.R. Mounce, A. Khan, A.S. Wood, A.J. Day, P.D. Widdop, and J. Machell, "Sensor-fusion of hydraulic data for burst detection and location in a treated water distribution system," *Inf. Fusion*, 4 (3), pp. 217–229, 2003. doi: 10.1016/S1566-2535(03)00034-4.
10. B. Farley, S.R. Mounce, and J.B. Boxall, "Development and field validation of a burst localization methodology," *J. Water Resour. Plan. Manag.*, 139 (6), pp. 604–613, 2013. doi: 10.1061/(ASCE)WR.1943-5452.0000290.
11. M. Romano, Z. Kapelan, and D.A. Savic "Geostatistical techniques for approximate location of pipe burst events in water distribution systems" *J. Hydroinf.*, 15 (3), pp. 634–651, 2013. doi: 10.2166/hydro.2013.094.
12. J. Kang, Y.J. Park, J. Lee, S.H. Wang, and D.S. Eom, "Novel leakage detection by ensemble cnn-svm and graph-based localization in water distribution systems," *IEEE Trans. Ind. Electron.*, 65 (5), pp. 4279–4289, 2018. doi: 10.1109/TIE.2017.2764861.
13. A.F. Colombo, P. Lee, and B.W. Karney, "A selective literature review of transient-based leak detection methods," *J. Hydro Environ. Res.*, 2 (4), pp. 212–227, 2009. doi: 10.1016/j.jher.2009.02.003.
14. H.R. Maier, A. Jain, G.C. Dandy, and K.P. Sudheer, "Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. Environ," *Model. Softw.*, 25, pp. 891–909, 2010. doi: 10.1016/j.envsoft.2010.02.003.
15. D.T. Bui, B. Pradhan, H. Nampak, Q.-T. Bui, Q.-A. Tran, Q.-P. Nguyen, "Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibilitgy modeling in a high-frequency tropical cyclone area using GIS," *J. Hydrol.*, 540, pp. 317–330, 2016. doi: 10.1016/j.jhydrol.2016.06.027.
16. H. Ghalkhani, S. Golian, B. Saghafian, A. Farokhnia, and A. Shamseldin, "Application of surrogate artificial intelligent models for real-time flood routing," *Water Environ. J.*, 27(4), pp. 535–548, 2013. doi: 10.1111/j.1747-6593.2012.00344.x
17. M. Rezaeianzadeh, H. Tabari, A.A. Yazdi, S. Isik, and L. Kalin, "Flood flow forecasting using ANN, ANFIS and regression models," *Neural Comput. Appl.*, 25(1), pp. 25–37, 2013. doi: 10.1007/s00521-013-1443-6.
18. F.-J. Chang and M.-J. Tsai, "A nonlinear spatio-temporal lumping of radar rainfall for modeling multi-step-ahead inflow forecasts by data-driven techniques," *J. Hydrol.*, vol. 535, pp. 256–269, 2016. doi: 10.1016/j.jhydrol.2016.01.056.
19. Y.S. Güçlü and Z. Şen, "Hydrograph estimation with fuzzy chain model," *J. Hydrol.*, vol. 538, pp. 587–597, 2016. doi: 10.1016/j.jhydrol.2016.04.057.
20. A. Lohani, R. Kumar, and R. Singh, "Hydrological time series modeling: A comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques," *J. Hydrol.*, vol. 442–443, pp. 23–35, 2012. doi: 10.1016/j.jhydrol.2012.03.031.
21. C.-J.F. Chung and A.G. Fabbri, "Validation of spatial prediction models for landslide hazard mapping," *Nat. Hazards*, vol. 30, pp. 451–472, 2003. doi: 10.1023/B:NHAZ.0000007172.62651.2b.

22. K. Socha and C. Blum, "An ant colony optimization algorithm for continuous optimization: application to feed-forward neural network training," *Neural Computing and Applications*, vol. 16(3), pp. 235–247, 2007. doi: 10.1007/s00521-007-0084-z.
23. C. Blum and K. Socha, "Training feed-forward neural networks with ant colony optimization: An application to pattern classification," in *Fifth International Conference on Hybrid Intelligent Systems (HIS'05), IEEE, 2005, November*. doi: 10.1109/ICHIS.2005.104.
24. Omid Bozorg-Haddad, Mohammad Solgi, and Hugo A. Loáiciga, *Meta-heuristic and Evolutionary Algorithms for Engineering Optimization*. 2017. doi: DOI:10.1002/9781119387053.
25. "ArcGIS analysis workflows for public safety," *ESRI*.

Received 28.05.2023

INFORMATION ON THE ARTICLE

Yuriy P. Zaychenko, ORCID: 0000-0001-9662-3269, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: zaychenkoyuri@ukr.net

Tetiana V. Starovoit, ORCID: 0009-0008-6335-7679, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: starovoyt.tania@lll.kpi.ua

ГІБРИДНА МОДЕЛЬ ШТУЧНОГО ІНТЕЛЕКТУ ІНТЕГРОВАНА В ГІС ДЛЯ ПРОГНОЗУВАННЯ АВАРІЙ НА МЕРЕЖАХ ВОДОПОСТАЧАННЯ / Ю.П. Зайченко, Т.В. Старовойт

Анотація. Пошук ефективної та надійної моделі прогнозування аварій на мережах водопостачання з визначенням їх точних розташувань завжди був важливим для ефективного керування системами розподілу води. Дослідження, засноване на моделі адаптивної нейронечіткої системи логічного висновку (ANFIS), розроблено для прогнозування аварій на мережі водопостачання міста Києва (Україна). Модель ANFIS поєднано з генетичними алгоритмами та методами ройової оптимізації (ACO) та інтегрували в ГІС для візуалізації результатів і визначення їх розташування. Прогнози оцінювали за такими критеріями: середньої абсолютної похибки (MAE), середньої квадратичної похибки (RMSE) та коефіцієнтом детермінації (R^2). Залежно від кількості та вигляду вхідних даних оптимізація ANFIS генетичними алгоритмами та алгоритмом ройової оптимізації (ACO) може в середньому збільшувати точність передбачення ANFIS на 10,1%, 11%. Отримані результати свідчать про те, що розроблена гібридна модель може бути успішно застосована для прогнозування аварій на мережах водопостачання.

Ключові слова: геоінформаційні системи, ANFIS, ACO, GA, просторові об'єкти, просторово-часовий аналіз, втрати води.