

LONG-TERM MONITORING OF SURFACE WATER QUALITY AND GROUNDWATER POTENTIAL USING COMPUTATIONAL INTELLIGENCE, GIS TECHNOLOGIES, AND REMOTE SENSING

S.V. KLIMOV, T.V. STAROVOIT

Abstract. Water scarcity and declining water quality due to population growth, urbanization, industrialization, and climate change highlight the importance of effective water management. Advances in remote sensing, cloud computing, and computational intelligence underscore the need to utilize modern technologies for monitoring surface water quality. This research involves the development of hybrid intelligent models using Landsat and Sentinel-2 images and WISE data with hybrid deep learning networks to evaluate surface water quality and groundwater potential. Correlation analysis revealed strong connections between remote sensing data and water quality parameters (such as chlorophyll-a, dissolved oxygen, nitrogen, and phosphorus). The hybrid models surpassed traditional machine learning methods, demonstrating their effectiveness in real-world water management.

Keywords: computational intelligence, fuzzy logic, remote sensing, satellite imagery, surface water quality monitoring, groundwater potential assessment, hybrid neural networks, NEFCLASS-EM, TS-FNN, Fuzzy C-Means, K-Means.

INTRODUCTION

Water is crucial for human health, food security, economic growth, energy production, and ecosystems. However, factors such as population growth, urbanization, industrial development, increased demand, and water misuse have made water scarce and expensive, particularly in developing countries. To address this issue, various strategies have been developed to improve water quality and quantity by 2030 [1]. In Europe, the Water Framework Directive (WFD) [2] aims to achieve good status for water resources. To assess the status, it's essential to monitor biological, hydro morphological, and physicochemical water quality indicators. According to this directive, rivers with a catchment area of more than 10 km² and lakes with an area of more than 0.5 km² should be included in the assessment and monitoring of water status [2].

Water quality parameters are traditionally determined by collecting samples on-site and analyzing them in the laboratory [3]. This method provides high accuracy, but is also labor-intensive and time-consuming, requiring significant financial investment. In addition, the traditional method determines the concentration of the required indicators only at the point of sampling. Meanwhile, the water quality in water bodies is rarely constant due to unpredictable events, such as accidental or deliberate leaks from industrial facilities and other factors. This makes accurate water quality monitoring a challenging task.

To overcome these limitations, we used remote sensing technology (RS), which has the advantage of large spatial coverage and high temporal resolution, which has been used to identify and monitor water resources more efficiently and effectively [4–6]. Remote monitoring of water quality indicators is based on establishing a correlation between the monitoring data and the corresponding surface reflection. Spectral characteristics of water are functions of hydrological, biological, and chemical characteristics of water [7]. Specifically, the amount of wave radiation at different wavelengths reflected from the water surface can be used directly or indirectly to detect water quality indicators [8].

Pure water can reflect light with a wavelength of more than 600 nm, which provides a high blue-green reflectance while absorbing radiation in the near-infrared (NIR) spectrum and beyond. Increasing the chlorophyll concentration increases the absorption of red (R) light and strongly absorbs blue (B) light, while the peak of reflection is located in the green (G) part of the spectrum [9]. The transparency of water depends on the total concentration of suspended solids. This concentration is a measure of the weight of inorganic particles suspended in the water column and is responsible for most of the scattering. By affecting the scattering of light, the suspended solids concentration (SSC) in water directly controls the transparency and oxygen content of a water body [10]. An increased concentration of SSC causes a shift in the peak from the G to the R region and increases the reflectance of water in the NIR region.

The relationship between surface reflectance and the concentration of water quality parameters is indirect and non-linear. This makes their estimation problematic, especially when based on traditional empirical algorithms. Over the past decade, the advancement in computing power and the development of artificial intelligence and machine learning (ML) algorithms have led to an increased use of these technologies to solve this problem. The most common machine learning models used in water quality assessment tasks are Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN).

Studies [11–14] have demonstrated that Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) deliver excellent performance in monitoring both optically active and inactive water quality indicators. Generally, artificial neural networks, as a linear approximation method, offer greater flexibility for monitoring water quality indicators. However, the accuracy of machine learning models typically depends on the chosen model and the quality of the training data. Developing ANN models requires large training datasets and significant experience to construct the optimal architecture for artificial neural networks. Using too many layers can lead to overfitting, which involves fitting noise in the training data [15]. Conversely, a small number of layers can lead to underfitting, where the model cannot adequately represent the complexity of the data [15].

This study aims to develop and compare hybrid computational intelligence models, specifically neuro-fuzzy neural networks, using spatial data from remote sensing and geographic information systems (GIS). It proposes combining neuro-fuzzy neural networks with metaheuristic and remote sensing algorithms to assess water quality from satellite images and evaluate their effectiveness. These hybrid models, integrating remote sensing and GIS, offer innovative methods for assessing water quality and groundwater potential. They can accurately identify factors contributing to water quality deterioration and unexpected surface formations in inland water bodies, while also examining their long-term impact on ecological status.

MATERIALS AND METHODS

Research area

We chose the territory of Kyiv and the Kyiv region (Ukraine) as the research area (Fig. 1). The study area is located in south-eastern Europe. Most of the rivers belong to the Black Sea basin. The largest river is the Dnipro, which we chose for our research. We used data from the Water Information System of Europe (WISE) to train artificial neural networks and neuro-fuzzy neural networks.

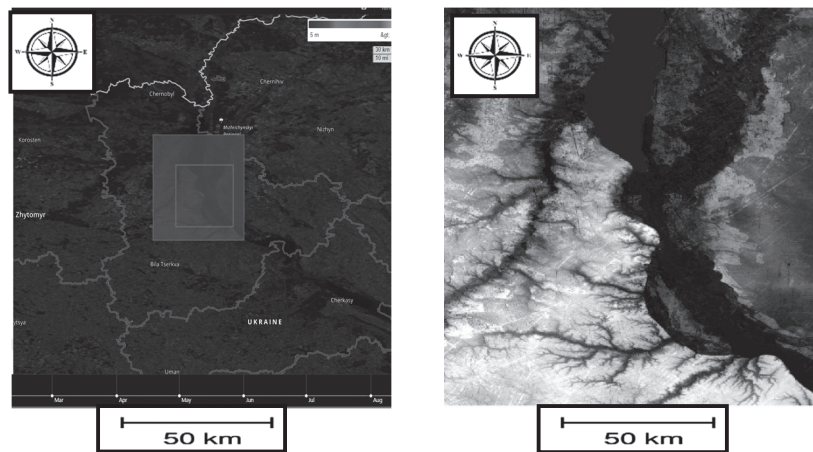


Fig. 1. Location of the study area in Kyiv and Kyiv region, Ukraine. The green rectangle marks the region of interest including key sampling sites along the Dnipro River

Data Preparation

Remote monitoring of water quality indicators is based on the correlation between on-site measurements and the corresponding surface reflectance. For this study, Landsat and Sentinel-2 satellite images of the surface over Europe from 2010 to 2024 were used. In total, more than two thousand images were analyzed to create time series and train monitoring models. Also used some of the materials from the DHI educational resource [16].

Landsat satellites achieve maximum surface coverage once every 16 days, with a spatial resolution of 30 m for multispectral bands. The Google Earth Engine API, integrated into Google Colab, was used as an access point to the imagery.

The surface reflectance values for each point were obtained from available Landsat Surface Reflectance Level 2A images. Cloud and shadow masking were performed to ensure clear water pixels. The resulting table included the identifier of the monitoring stations, the corresponding surface reflectance value, and the date of the survey. The surface reflectance was filtered by date to match the on-site data, with a maximum time interval of 3 days between the on-site sampling and the satellite overpass.

We used Pearson correlation analysis to explore the connection between remote sensing and in situ data, using the correlation coefficient (r). Based on the correlation, we identified a specific set of input data for each water quality indicator. The data was then standardized to a normal distribution with a mean of 0 and a standard deviation of 1, and then divided into training and test sets (80 % and 20 %, respectively).

Fuzzy neural network model

Fuzzy theoretical TS systems can apply fuzzy mathematical rules to generate more complex nonlinear functions. This allows the system to reduce the number of fuzzy rules needed when dealing with problems involving multiple variables [17]. The fuzzy theoretical system TS is typically defined using "if-then" logic, and its fuzzy conclusion is expressed as follows [18]:

$$R^i : \text{If } x_1 \text{ is } A_1^i, x_2 \text{ is } A_2^i, \dots, x_k \text{ is } A_k^i, \quad (1)$$

$$\text{Then } y_i = p_0^i + p_1^i x_1 + \dots + p_k^i x_k,$$

where A_j^i is a fuzzy

set of a fuzzy system; p_j^i ($j = 1, 2, \dots, k$) are the parameters of the fuzzy system; y_i is the initial value obtained by the fuzzy rule; input part i.e. *If* is fuzzy and output part i.e. *Then* is deterministic part. This vague conclusion indicates that the output is a linear combination of the inputs [18].

Suppose that for inputs $x = [x_1, x_2, \dots, x_k]$, the degree of membership of each input variable x_j is first calculated by a fuzzy rule [18]:

$$\mu_{A_j^i} = \exp(-(x_j - c_j^i)^2 / b_j^i), \quad (2)$$

$$(j = 1, 2, \dots, k; i = 1, 2, \dots, n),$$

where c_j^i and b_j^i are the center and width of the membership function, k is an input parameter, and n is the number of fuzzy subsets [18].

Fuzzy operations are performed on each of the above degrees of membership, and the fuzzy operator is used as a concatenated multiplicative operator [18]:

$$\omega^i = u_{A_1^i}(x_1) \times u_{A_2^i}(x_2) \times u_{A_k^i}(x_k), \quad (3)$$

$$(i = 1, 2, \dots, n).$$

According to the results of the fuzzy calculation, the initial value of the model y_i is obtained [18]:

$$y = \frac{\sum_{i=1}^n \omega^i (p_0^i + p_1^i x_1 + \dots + p_k^i x_k)}{\sum_{i=1}^n \omega^i}. \quad (4)$$

A Fuzzy Neural Network (FNN) is a hybrid model that integrates the learning capabilities of artificial neural networks with the interpretability of fuzzy logic systems. The network typically consists of four layers: an input layer, a fuzzification layer, a rule base, and an output layer.

The algorithm of this type of neural network is described using equation (5).

$$\mu_i(x_i) = \exp\left(-\frac{(x_i - c_i)^2}{2\sigma_i^2}\right), \quad (5)$$

where $x = (x_1, x_2 \dots, x_n)$ input vector; and $\mu_i(x_i)$ is the membership function. The activation of rule j is described by equation (6). The output of the network is described by equation (7).

$$w_j = \prod_{i=1}^n \mu_i^j(x_i), \tag{6}$$

$$y = \frac{\sum_{j=1}^r w_j f_j(x)}{\sum_{j=1}^r w_j}, \text{ where } f_j(x) = a_j^T x + b_j. \tag{7}$$

A Takagi–Sugeno Fuzzy Neural Network (TS-FNN) extends the FNN by using linear functions in the consequents of fuzzy rules. Instead of constant outputs, each rule produces an output of the form:

$$y_j = a_1^j x_1 + a_2^j x_2 + \dots + a_n^j x_n + b^j. \tag{8}$$

The final output is a weighted sum of these rule outputs, normalized by the total rule strength. TS-FNNs offer higher precision and faster convergence compared to classical FNNs.

In the fuzzy neural network NEFCLASS (Neuro-Fuzzy CLASSification) classification is performed with automatic rule learning. Inputs are fuzzified by linguistic variables (low, medium, high) [19]:

$$\mu A_k(x) = \max\left(0.1 - \left| \frac{x - c_k}{\sigma_k} \right| \right). \tag{9}$$

The rules are presented in the form: *If x_1 is "High" and x_2 is "Low" then Class = Water*

The process of class estimation through fuzzy inference and max-aggregation is described in equation (10), [20].

$$y = \arg \max_c \left(\max_j \left(\min_i \mu_i^j(x_i) \right) \right). \tag{10}$$

The NEFCLASS-EM model enhances NEFCLASS with a metaheuristic optimizer, such as the Electromagnetic Algorithm (EM). NEFCLASS-EM is calculated in a similar way as in equations (9), (10), but with optimization of parameters (centers, widths) through metaheuristics (equation (11)).

$$\theta^{(t+1)} = \theta^{(t)} + F_{EM}(\theta^{(t)}), \tag{11}$$

where F_{EM} – is the vector of the force of attraction/repulsion between candidates. A schematic representation of the architecture of the four hybrid neural network models is shown in Fig. 2.

Fig. 2 illustrates the structural differences between four neuro-fuzzy models: FNN, TS-FNN, NEFCLASS, and NEFCLASS-EM. Each model includes key functional components such as input processing, fuzzification, rule evaluation, and output generation. The FNN and TS-FNN architectures are organized and layered, with TS-FNN producing linear outputs. NEFCLASS and NEFCLASS-EM focus on classification tasks, where NEFCLASS-EM integrates a metaheuristic optimizer (EM) to enhance the rule parameter optimization.

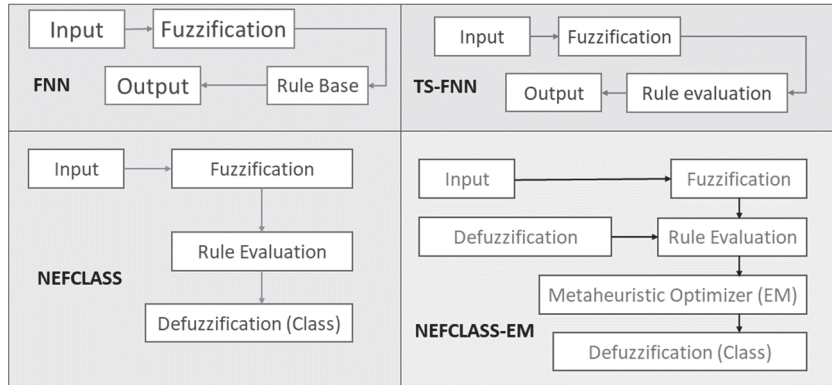


Fig. 2. Architectures of the fuzzy neural network models used in the study

In this study, fuzzy neural networks were applied to predict water quality indicators and groundwater potential. A fuzzy neural network model usually has four levels: an input level, a fuzzification level, a fuzzy rule calculation level, and an output level [21]. The input layers of the network model are connected through the vector x_i , so the number of nodes of the model network is consistent with the dimension of the input vector. The fuzzification layer uses the membership function of equation (3) to fuzziness the input values to obtain the membership value $u_{A_j^i}$. The value of ω at the fuzzy computing level is obtained by using equation (4) of successive phase multiplication, and then the output value of the output data level in this fuzzy model system is obtained by equation (8). The fuzzy neural network learning algorithm is as follows [21]:

Step 1. Calculation error:

$$e = \frac{1}{2} (y_d - y_c)^2, \quad (12)$$

where the expected output of the network is y_c and the error between the expected output and the actual output is e [21].

Step 2. Correction of the coefficient:

$$p_j^i(k) = p_j^i(k-1) - \alpha \frac{\partial e}{\partial p_j^i}, \quad (13)$$

$$\frac{\partial e}{\partial p_j^i} = \frac{(y_d - y_c) \omega^i}{\sum_{i=1}^m \omega^i \cdot x_j},$$

where p_j^i is the coefficient of the neural network; α is the learning rate of the network; x_j is the input parameters of the network; ω^i is the continuous product of the membership of the input parameters [21].

Step 3. Correction of parameters:

$$c_j^i(k) = c_j^i(k-1) - \beta \frac{\partial e}{\partial c_j^i}, \quad (14)$$

$$b_j^i(k) = b_j^i(k-1) - \beta \frac{\partial e}{\partial b_j^i},$$

where b_j^i and c_j^i are the width and center value of the representative membership function in the fuzzy rule [21].

Indicators of assessment of accuracy of models

We used four metrics to evaluate the accuracy of the models: overall accuracy (OA), misclassification error (ME), omission error (OE), and ROC-AUC value. OA is the sum of pixels correctly classified as water divided by the total number of water pixels represented by the confusion matrix. OE is the number of pixels that belong to water but are classified as other surface types that can be identified by the error matrix column. The ROC-AUC value [22] is the area under the curve of the ratio of sensitivity (equation 15) to specificity (equation 16). This value ranges from 0.50 to 1. The higher the value, the better the performance. If the value exceeds 0.70, the classification result is reliable [20].

$$\text{Sensitivity} = \frac{TPR}{TPR+FNR}, \quad (15)$$

$$\text{Specificity} = \frac{TNR}{TNR+FPR}, \quad (16)$$

where TPR (true positive) and FNR (false negative) are pixels correctly and incorrectly classified as water, and TNR (true negative) and FPR (false positive) are pixels correctly and incorrectly classified as non-water [23].

To assess water quality, a water quality model's performance can be measured using various metrics. These include the coefficient of determination (R^2 , Eq. 17), the mean absolute error (MAE, Eq. 18), the root means square error (RMSE, Eq. 19), the mean square error (MSE, Eq. 20), residual prediction deviation (RPD, Eq. 21), and confidence interval (CI, Eq. 22), [23].

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y'_i)^2}{\sum_{i=1}^n (y'_i - \bar{y})^2}, \quad (17)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y'_i|, \quad (18)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2}, \quad (19)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2, \quad (20)$$

$$RPD = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sum_{i=1}^n (y_i - y'_i)^2}}, \quad (21)$$

$$CI = \left[1 - \frac{\sum_{i=1}^n |y'_i - y_i|}{\sum_{i=1}^n (|y'_i - \bar{y}'| + |y_i - \bar{y}|)} \right] \times \left[1 - \frac{\sum_{i=1}^n (y'_i - y_i)^2}{\sum_{i=1}^n (y'_i - \bar{y}')^2} \right]. \quad (22)$$

R^2 values should not be too high, as excessively high values lead to overfitting and lack of model portability, since this metric is sensitive to outliers [23]. Therefore, in practice, R^2 is often used in combination with RMSE, RPD, CI, and other indicators to balance the fitting accuracy and computational complexity [23, 24].

RESULTS OF SIMULATION

Aerospace research and geoinformation modeling

Satellite technology has a significant advantage in its ability to capture light beyond the visible spectrum, which is undetectable to the human eye. Infrared light, a type of radiation, can be detected by satellites like Sentinel-2. This infrared radiation can provide valuable information about surface temperature, vegetation conditions, and atmospheric conditions. For instance, healthy vegetation reflects more infrared light than unhealthy vegetation or non-vegetated surfaces.

The raster image measures 1009 pixels in height and 1014 pixels in width, covering approximately 10 kilometers in both dimensions. It contains four spectral bands: blue, green, red, and near infrared. The satellite captures data across 11 different wavelengths, and we specifically selected the blue, green, red, and infrared spectra for analysis.

The near-infrared range is very helpful for analyzing vegetation. In this setup, healthy vegetation appears bright red, while non-vegetated surfaces appear in other colors. This method improves the visibility of different ground cover types and allows us to see details that are not visible in normal light.

We first rendered the image in different color spaces and then focused on the red and infrared ranges. To do this, we created a scatter diagram where the reflection coefficient of the red pixel is shown on the x-axis and the infrared pixel on the y-axis (Fig. 3). Next, we converted the images into tabular data. Each row in the resulting table represents one pixel for a specific date. For instance, the first row corresponds to the pixel (1, 1) on February 22, 2022. Each column represents the intensity of a spectral band (blue, green, red, infrared).

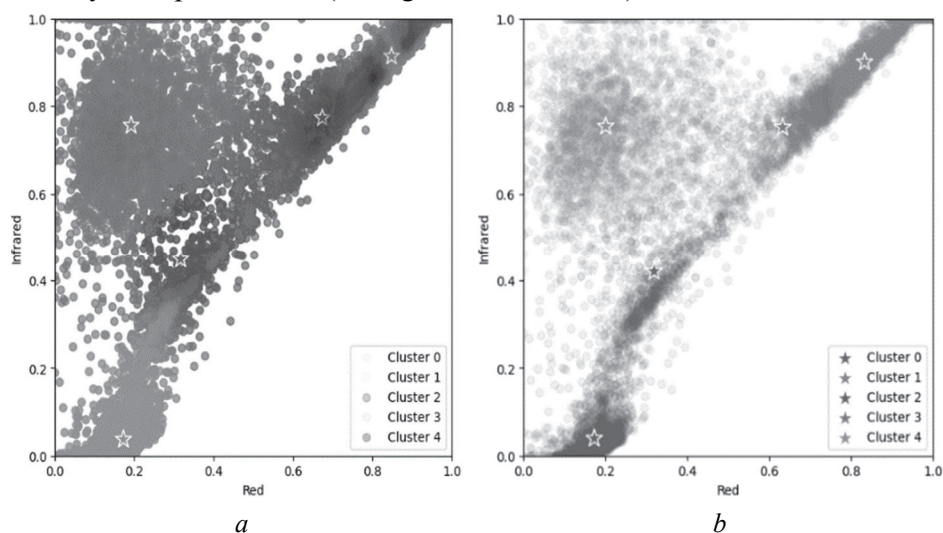


Fig. 3. Visualization of the results of clear and fuzzy clustering, where *a* – clear clustering; *b* – fuzzy clustering of pixels

To recognize clusters that correspond to water bodies, we applied the following intelligent algorithms: Fuzzy Neural Networks (FNN), Takagi–Sugeno Fuzzy Neural Networks (TS-FNN), NEFCLASS, and NEFCLASS-EM.

DEVELOPMENT OF COMPUTATIONAL INTELLIGENCE MODELS

Fuzzy neural network model

Fuzzy neural networks (FNNs) are hybrid models that combine fuzzy logic with the computational power of neural networks. They are designed to handle uncertainties and imprecise data in tasks such as classification, clustering, and regression. In the context of satellite imagery, FNNs are particularly useful because satellite data often contain noise, incomplete information, and inherent fuzziness. This is especially true when dealing with natural phenomena such as cloud cover, landforms, vegetation, and water bodies.

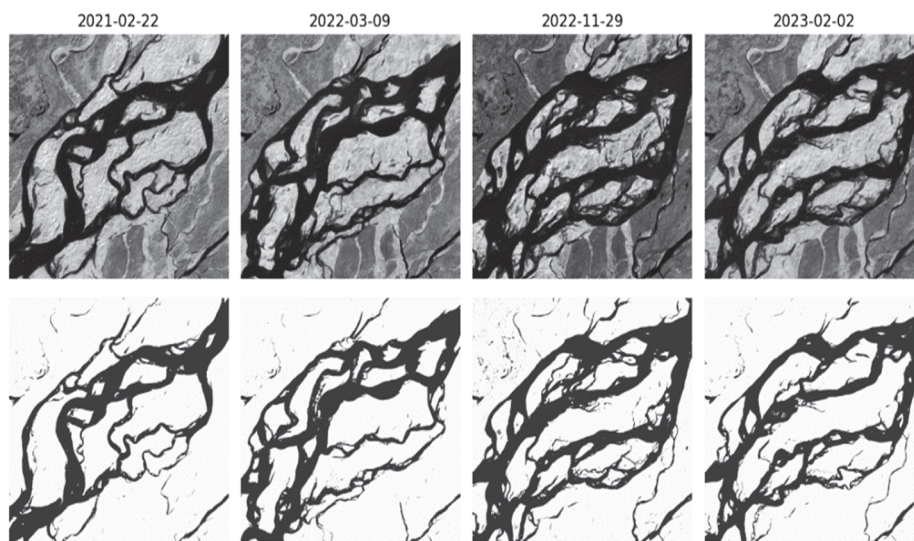


Fig. 4. Visualization of water recognition results on satellite images using a fuzzy neural network (FNN) is visualized

Fig. 4 shows the water extraction result, where we can see that the fuzzy neural network recognized water in satellite images very well. FNNs allow the use of fuzzy membership values (rather than binary solutions) to model these ambiguous or mixed pixels. Unlike traditional neural networks, FNNs assign membership degrees to different classes, so a single pixel can belong to multiple categories with different probabilities or degrees (e.g., 70 % forest, and 30 % water).

Takagi–Sugeno fuzzy neural network (TS-FNN) model

Takagi–Sugeno models are a type of fuzzy logic inference system. In these models, the output of fuzzy rules can be a linear combination of input variables or a constant. In TS-FNN, each fuzzy rule corresponds to a linear model or constant that is learned, and the output is calculated as the weighted average of the rule's output. The implementation of the TS-FNN model is depicted in Fig. 5. The model is trained to minimize the mean squared error (MSE) between the predicted outputs and the actual labels. Post-training, the network predicts cluster membership for each pixel in the satellite image, and the assigned clusters are then visualized.

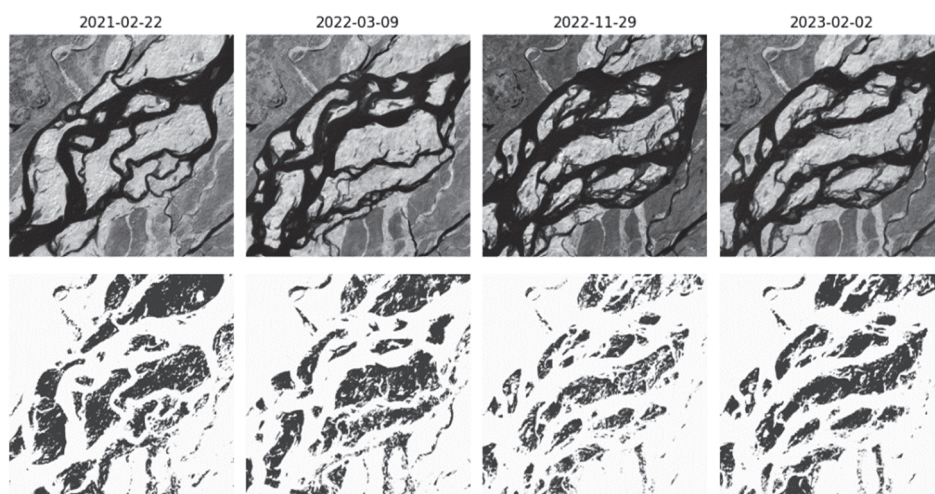


Fig. 5. Visualization of water recognition results on satellite images using Takagi-Sugeno fuzzy neural network (TS-FNN)

Fuzzy rules in TS-FNN are better interpreted because the output is a linear function of the input variables, making it easier to understand how the model makes its decisions.

NEFCLASS fuzzy neural network model

NEFCLASS is a neuro-fuzzy system specifically created for classification tasks. It integrates fuzzy logic with a feed-forward neural network framework. In this system, fuzzy rules are acquired from input data. The model generates fuzzy classification rules automatically, which are then refined using neural network training methods. The neural network adapts the fuzzy rules to minimize classification errors.

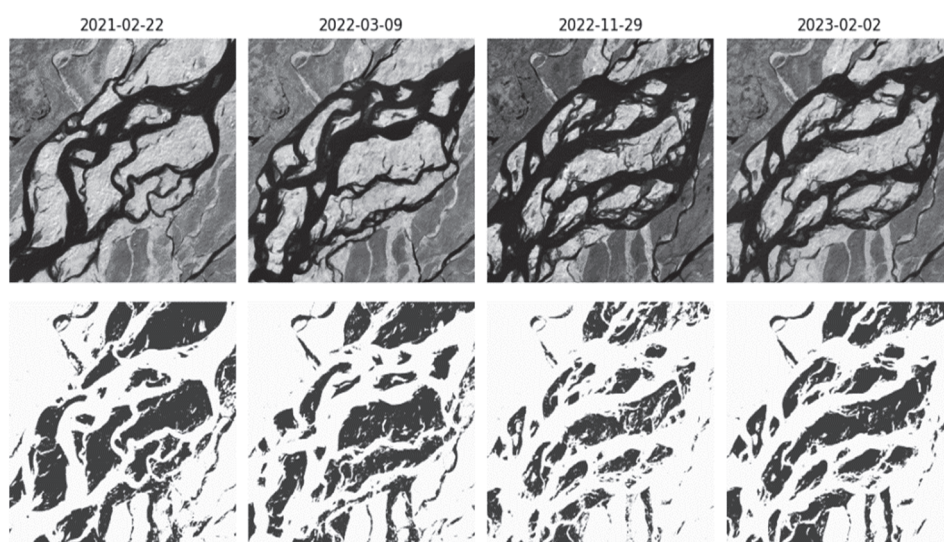


Fig. 6. Visualization of the results of water recognition on satellite images using the NEFCLASS fuzzy neural network

For simplicity, we implemented NEFCLASS-like functionality using a neural network to represent fuzzy rules, but we manually defined the fuzzification process. The result is shown in Fig. 6.

To implement this model, we utilized the same approach as in our previous models. This involved defining fuzzy sets for the red and infrared ranges, which were classified as low, medium, and high. The input layer represents the fuzzy inputs or fuzzy characteristics, and the subsequent layers model the combination of these fuzzy rules.

The main advantage of this model is its ability to learn fuzzy rules during training, making it suitable for handling complex decisions, such as those involved in satellite image classification. NEFCLASS can scale to larger data sets due to its neural network structure, which allows it to handle multidimensional data more efficiently.

Hybrid model NEFCLASS-EM

To achieve the highest clustering accuracy, we also tested the NEFCLASS network with the electromagnetic metaheuristic's algorithm. The result is displayed in Fig. 7.

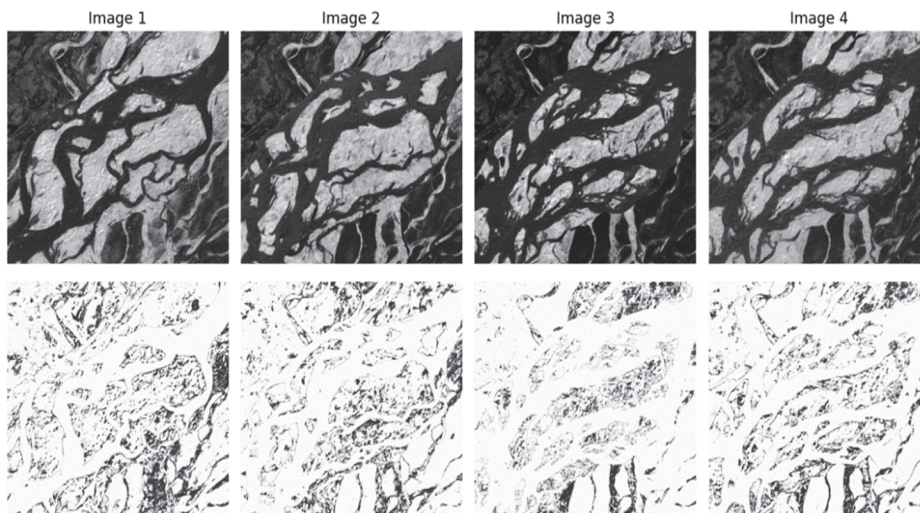


Fig. 7. Visualization of the results of water recognition on satellite images using the NEFCLASS-EM hybrid neural network

We combined a neural network with an electromagnetic metaheuristic (EM) algorithm to improve the network's performance. We also used the EM algorithm to optimize the hyperparameters of the NEFCLASS model, which enhanced its classification efficiency. The EM algorithm works by modeling the attraction and repulsion forces between different solutions (weight configurations) based on electromagnetic principles. After optimizing the weights, we further trained the NEFCLASS network using the Adam optimizer. This hybrid approach led to faster convergence and improved classification accuracy by leveraging the global search capabilities of the EM algorithm and the fine-tuning ability of gradient-based optimization.

Comparison of obtained results and assessment of the accuracy of models of fuzzy neural networks

The FNN fuzzy network shows a significant improvement in accuracy, quickly exceeding 91% by the second epoch and stabilizing around 96% by the 20th epoch. Losses also steadily decrease, indicating efficient learning and reduced errors. TS-FNN shows a rapid loss reduction, reaching 0.0114 by the 20th epoch. This rapid convergence indicates high accuracy. NEFCLASS shows a result like FNN but starts with slightly less accuracy. By the 20th epoch, it reaches over 95% accuracy. Loss reduction is slower compared to FNN, but still significant. NEFCLASS-EM starts with the lowest accuracy and highest losses but stabilizes quickly. By the end of training, the model achieves performance like NEFCLASS, with an accuracy of over 95% and a significant loss reduction. The results of model training accuracy are shown in Table 1.

So, from the obtained fuzzy neural network training results, we can see that FNN exhibits the best overall balance of rapid accuracy improvement and stable loss reduction, making it the most efficient model in terms of both learning speed and final performance. TS-FNN has the fastest loss reduction, indicating a very accurate model. NEFCLASS is comparable to FNN, but slightly slower in terms of increasing accuracy and reducing loss. NEFCLASS-EM starts with the lowest performance but catches up to achieve accuracy levels close to FNN and NEFCLASS by the last epoch.

The accuracy values reported in Table 1 were obtained from the training dataset over 20 epochs. To better assess the training dynamics, Figs. 8 and 9 show the accuracy and loss values over 20 epochs for each model.

Fig. 8 shows the “Accuracy vs Epoch” curve for four models: FNN, TS-FNN, NEFCLASS, and NEFCLASS-EM, based on the results presented in Table 1. FNN demonstrates the fastest accuracy growth and stable performance (~96–97 %).

TS-FNN starts slightly slower but quickly stabilizes at the same level. NEFCLASS and NEFCLASS-EM initially show lower accuracy but gradually catch up.

NEFCLASS-EM exhibits powerful dynamics after the 10th epoch. The plot shows that TS-FNN and FNN achieve high accuracy early in training, while NEFCLASS and NEFCLASS-EM improve gradually and stabilize after 10 epochs.

The “Loss vs Epoch” graph (Fig. 9) shows that TS-FNN achieves the fastest loss reduction, reaching approximately 0.0114 by the 20th epoch. FNN steadily reduces loss, although at a slightly slower pace. NEFCLASS and NEFCLASS-EM start with higher loss values but gradually decrease them to an acceptable level, demonstrating stable learning behavior.

TS-FNN shows the fastest reduction in loss, reaching approximately 0.0114 by epoch 20. FNN steadily reduces loss, while NEFCLASS and NEFCLASS-EM start from higher values but show gradual and stable convergence.

The results show that FNN and TS-FNN achieve the highest accuracy and the fastest convergence, while NEFCLASS and NEFCLASS-EM demonstrate stable learning with gradual improvement. The integration of fuzzy logic and neural networks enables accurate modeling of nonlinear processes in aquatic environments, offering promising opportunities for intelligent environmental monitoring.

Table 1. Comparison of the obtained results of training accuracy of fuzzy neural networks

No. Epoch	Fuzzy Neural Network (FNN)	Takagi–Sugeno Fuzzy Neural Network (TS-FNN)	NEFCLASS Fuzzy Neural Network	Hybrid NEFCLASS-EM
1	0.5781	0.5710	0.5630	0.4789
2	0.9162	0.9113	0.8616	0.8452
3	0.9488	0.9363	0.9102	0.9068
4	0.9578	0.9368	0.9177	0.9237
5	0.9594	0.9391	0.9227	0.9289
6	0.9582	0.9382	0.9279	0.9355
7	0.9584	0.9404	0.9330	0.9425
8	0.9647	0.9563	0.9398	0.9463
9	0.9630	0.9590	0.9369	0.9464
10	0.9662	0.9562	0.9466	0.9478
11	0.9642	0.9558	0.9437	0.9498
12	0.9639	0.9539	0.9445	0.9505
13	0.9639	0.9539	0.9513	0.9553
14	0.9640	0.9540	0.9444	0.9531
15	0.9626	0.9526	0.9506	0.9547
16	0.9643	0.9543	0.9488	0.9587
17	0.9643	0.9543	0.9485	0.9569
18	0.9664	0.9564	0.9452	0.9544
19	0.9619	0.9519	0.9543	0.9534
20	0.9679	0.9579	0.9512	0.9560

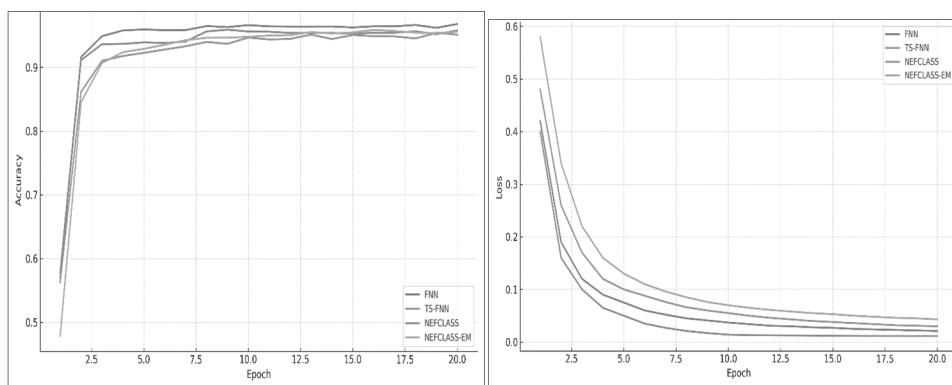


Fig. 8. Accuracy vs Epoch curves for all models Fig. 9. Loss vs Epoch curves for all models

Assessment of water quality and groundwater potential

Based on our research, we have developed a methodology for evaluating water quality by utilizing satellite images and computational intelligence techniques, specifically fuzzy neural networks. This method integrates remote sensing and geoinformation modeling with sophisticated machine learning and artificial intelligence

(AI) models. This approach enables the automated analysis of water quality parameters on a large scale, including turbidity, chlorophyll-a concentration, and total suspended solids (TSS).

To reflect water quality using this method, we have identified the following four spatial resolution bands:

1. Blue (450–500 nm): sensitive to chlorophyll-a and water clarity.
2. Green (500–600 nm): reflects organic matter and suspended particles.
3. Red (600–700 nm): Helps detect deposits.
4. Near infrared (700-1100 nm): useful for turbidity and TSS determination.

In the next phase, we extracted spectral and spatial features from satellite images that are important for assessing water quality. These features serve as input data for computational intelligence models. We calculated the chlorophyll estimate using the normalized difference chlorophyll index (NDCI) with formula 16.

$$NDCI = \frac{Red\ edge1 - Red}{Red\ edge1 + Red}. \quad (23)$$

To determine the turbidity, we utilized the turbidity index (NTU) [22] derived from the red and near-infrared spectrums [23]. The combination of red and green wavelengths assisted in estimating total suspended solids (TSS). To enhance our analysis and capture spatial variations in water bodies, we employed spatial texture functions, which involve a matrix of adjacent gray levels. Additionally, to gain a better understanding of the water body characteristics, we calculated statistical measures such as the mean value, variance, and entropy of pixel values.

Data from ground-based measurements of water quality parameters such as chlorophyll, turbidity, and suspended solids were utilized to calibrate and validate computational intelligence models.

The idea of using fuzzy neural networks produced better results than those based on traditional neural networks and traditional machine learning methods. Fuzzy logic systems are ideal for handling uncertainty in water quality assessment, especially when the boundaries between quality classes are unclear. Fuzzy neural networks enable the determination of fuzzy membership functions for water quality parameters based on satellite features (for example, a pixel can belong to both “clean water” and “polluted water” with different degrees of membership).

Once the model is trained and tested, it can be used to predict water quality parameters based on new satellite images and to create maps of water quality. By applying the models to satellite images over different periods, temporal changes in water quality can be monitored. This is particularly useful for detecting trends such as algal blooms, pollution, or seasonal variations in water quality. Continuously monitoring water quality using computational intelligence models can help detect early signs of pollution or algal blooms and trigger warnings.

Obtaining the correct characterization from satellite imagery is critical to assessing groundwater potential. These features often serve as input to computational intelligence models. For example, the Normalized Difference Vegetation Index (NDVI) helps assess the health and density of vegetation, which is related to the availability of groundwater

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (24)$$

Digital relief models (Fig. 10, *a*) can help us understand the topography of an area. Low-lying areas and valleys have a higher potential for groundwater recharge. The slope affects water runoff, while the height (slope direction) affects moisture retention. Analyzing river patterns and drainage patterns using satellite data can help identify areas with high infiltration potential.

Surface soil moisture (SSM): obtained from satellites such as Sentinel-1 (Fig. 10, *b*) shows how much water the soil holds. High soil moisture provides good conditions for groundwater replenishment. The water content of vegetation can indirectly indicate the level of soil moisture and the potential presence of groundwater.

Satellite radar and optical images can help identify geologic lineaments (faults, cracks) that act as conduits for groundwater. Radar data from satellites such as Sentinel-1 can be useful for this purpose. Determining different types of rock or formations (e.g., porous rock, fractured aquifers) from satellite data is critical because certain geological formations are more favorable for groundwater storage. Mapping lakes, rivers, and wetlands from satellite imagery provides insight into groundwater recharge zones (Fig. 11), as surface water bodies are often associated with aquifers.

To train and test the computational intelligence models, we collected ground-based data such as groundwater levels from boreholes, soil moisture profiles, geological and hydrogeological surveys, and basic climate data (precipitation, and evaporation rates) to determine recharge potential.

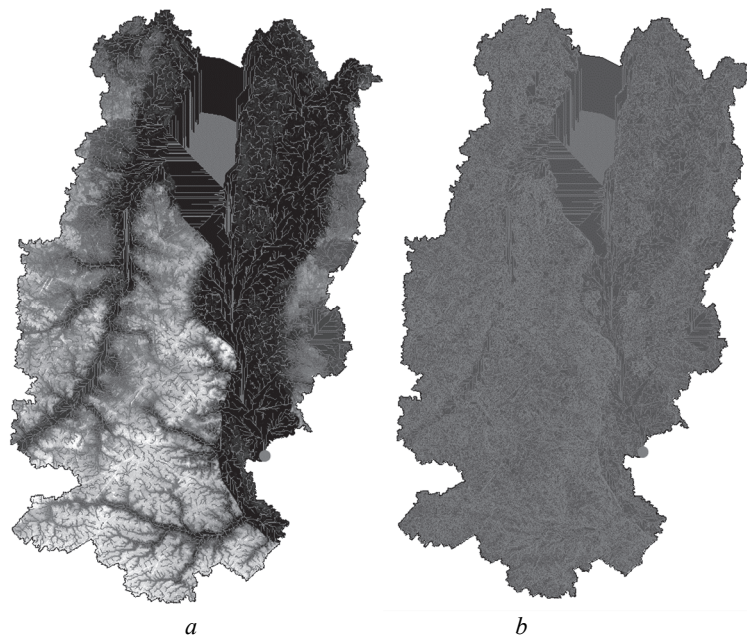


Fig. 10. Visualization of the results of obtaining a digital model of the relief – *a*; and soil moisture – *b*

After extracting features from satellite imagery, we applied computational analysis techniques to model the relationships between these features and groundwater potential. Characteristics such as NDVI index, DEM, soil moisture, land use,

slope, and drainage density were used as input variables. As a result, we received a model capable of predicting areas with high or low groundwater potential based on input characteristics.

Fuzzy logic is particularly useful in groundwater assessment because it can handle uncertainty and variability in data (e.g., variable soil moisture). The spatial and temporal resolution of satellite data can affect the accuracy of groundwater estimation. Higher resolution data increases accuracy but may not be available in all regions.

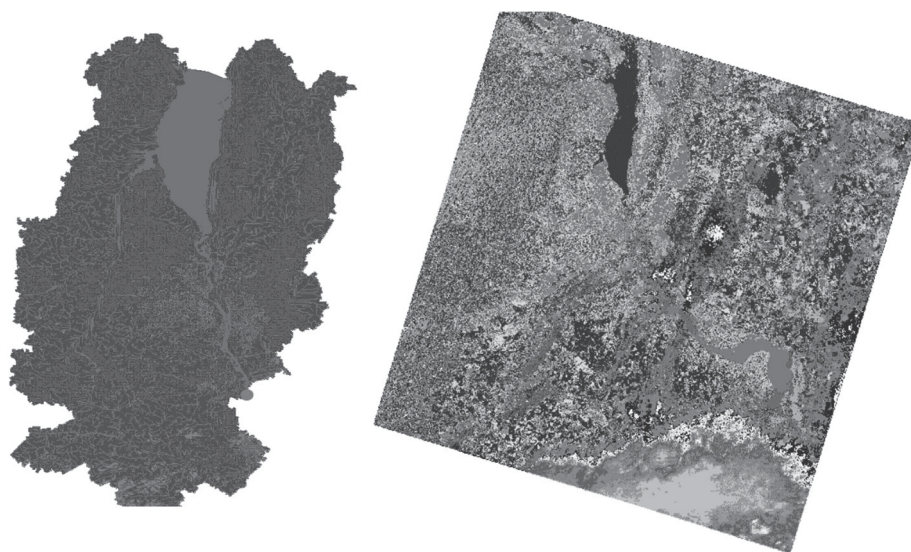


Fig. 11. The result of mapping lakes, rivers, and wetlands from satellite images

Groundwater potential assessment using satellite imagery and computational analysis techniques combines the strengths of remote sensing, machine learning, and data science. By combining key characteristics such as soil moisture, vegetation indices, topography, and geological structures, these methods provide a powerful, cost-effective approach to groundwater investigation and monitoring.

CONCLUSION

Water quality assessment from satellite images using computational intelligence methods is a powerful approach for large-scale automated monitoring of water bodies. Combining remote sensing data with techniques such as artificial neural networks, fuzzy logic, decision trees, and support vector machines enables accurate prediction of key water quality parameters such as chlorophyll-a, turbidity, and total suspended solids. By continuously monitoring water quality using satellite images, computational intelligence techniques can help detect environmental changes, manage water resources, and prevent water pollution.

As a result of the conducted research, the following key points and advantages of the methods proposed in the article can be noted:

- Remote sensing allows continuous coverage of a wide area of water bodies and terrestrial landscapes, enabling large-scale monitoring that would be impractic-

cal using traditional terrestrial methods. Satellite data provide important information about water quality and groundwater potential over vast regions, including remote and hard-to-reach areas.

- Computational intelligence techniques such as fuzzy neural networks can automate the process of analyzing complex satellite data, greatly reducing the time and manpower required for water resources assessment. These models can quickly identify patterns, anomalies, and trends, providing real-time or near-real-time information critical to timely decision-making.

- Fuzzy neural networks excel at dealing with the uncertainty and imprecision inherent in the natural environment. Water quality parameters and groundwater potential often exhibit complex, non-linear relationships influenced by multiple factors (e.g., land use, vegetation, climate). FNNs effectively model these complexities and provide soft classifications that allow for more flexible predictions.

- The use of satellite data and AI-based models minimizes the need for extensive field studies, reducing the costs associated with traditional water and groundwater quality assessments. Thanks to the use of freely available satellite platforms (e.g. Sentinel, Landsat) and powerful computing tools, environmental monitoring is becoming more accessible and scalable.

- Geo-information modeling allows the integration of spatial and temporal data, allowing water quality trends and groundwater potential to be tracked over time. It helps identify seasonal patterns, long-term environmental changes, and the effects of human activities such as agriculture, urbanization, or pollution.

Therefore, the combination of fuzzy neural networks, remote sensing of the Earth and geoinformation modeling offers a reliable, dynamic, and effective framework for the management and protection of water resources, ensuring their sustainability in the face of growing environmental problems.

ACKNOWLEDGMENTS

The ideas for this study are based on the knowledge of water quality assessment deepened and expanded during the course “Operation and Maintenance of Urban Water Supply System (Water Quality and Purification) (B)” in Japan, with the financial support of the Japan International Cooperation Agency (JICA).

REFERENCES

1. UN General Assembly. Transforming Our World: The 2030 Agenda for Sustainable Development. 21 October 2015. Available: <https://www.refworld.org/legal/resolution/unga/2015/en/111816>
2. *Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy*; European Parliament: Bruxelles, Belgium, 2003. Available: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32000L0060>

3. F. Faur, M. Lazăr, I.-M. Apostu, O. Pinchuk, S. Klimov, "Monitoring the water quality of Jiu River in Dolj County," *E3S Web Conf.*, vol. 280, article no. 10002, 2021. doi: <https://doi.org/10.1051/e3sconf/202128010002>
4. J. He, Y. Chen, J. Wu, D.A. Stow, G. Christakos, "Space-Time Chlorophyll-a Retrieval in Optically Complex Waters that Accounts for Remote Sensing and Modeling Uncertainties and Improves Remote Estimation Accuracy," *Water Research*, vol. 171, 115403, 2019. doi: <https://doi.org/10.1016/j.watres.2019.115403>
5. B. Nas, S. Ekercin, H. Karabörk, A. Berktaş, D.J. Mulla, "An Application of Landsat-5TM Image Data for Water Quality Mapping in Lake Beyşehir, Turkey," *Water, Air, & Soil Pollution*, vol. 212, pp. 183–197, 2010. doi: <https://doi.org/10.1007/s11270-010-0331-2>
6. M. Govedarica, G. Jakovljevic, "Monitoring spatial and temporal variation of water quality parameters using time series of open multispectral data," in *Seventh International Conference on Remote Sensing and Geoinformation of the Environment, Paphos, Cyprus, 18–21 March 2019*, vol. 11174. doi: <https://doi.org/10.1117/12.2533708>
7. C. Wu et al., "Empirical estimation of total phosphorus concentration in the mainstream of the Qiantang River in China using Landsat TM data," *Int. J. Remote Sens.*, vol. 31, issue 9, pp. 2309–2324, 2010. doi: <https://doi.org/10.1080/01431160902973873>
8. Gordana Jakovljevic, Flor Álvarez-Taboada, Miro Govedarica, "Long-Term Monitoring of Inland Water Quality Parameters Using Landsat Time-Series and Back-Propagated ANN: Assessment and Usability in a Real-Case Scenario," *Remote Sensing*, vol. 16, issue 1, 68, 2024. doi: <https://doi.org/10.3390/rs16010068>
9. N. Ha, K. Koike, M. Nhuan, "Improved Accuracy of Chlorophyll-a Concentration Estimates from MODIS Imagery Using a Two-Band Ratio Algorithm and Geostatistics: As Applied to the Monitoring of Eutrophication Processes over Tien Yen Bay (Norther Vietnam)," *Remote Sensing*, vol. 6, issue 1, pp. 421–442, 2013. doi: <https://doi.org/10.3390/rs6010421>
10. B. Nechad, K. Ruddick, Y. Park, "Calibration and validation of a generic multisensor algorithm for mapping of total suspended matter in turbid waters," *Remote Sensing of Environment*, vol. 114, issue 4, pp. 854–866, 2010. doi: <https://doi.org/10.1016/j.rse.2009.11.022>
11. K.T. Peterson, V. Sagan, J.J. Sloan, "Deep learning-based water quality estimation and anomaly detection using Land-sat-8/Sentinel-2 virtual constellation and cloud computing," *GIScience & Remote Sensing*, vol. 57, issue 4, pp. 510–525, 2020. doi: <https://doi.org/10.1080/15481603.2020.1738061>
12. S. Hafeez et al., "Comparison of Machine Learning Algorithms for Retrieval of Water Quality Indicators in Case-II Waters: A Case Study of Hong Kong," *Remote Sensing*, vol. 11, issue 6, 617, 2019. doi: <https://doi.org/10.3390/rs11060617>
13. L.F. Arias-Rodriguez, U.F. Tüzün, Z. Duan, J. Huang, Y. Tuo, M. Disse, "Global Water Quality of Inland Waters with Harmonized Landsat-8 and Sentinel-2 Using Cloud-Computed Machine Learning," *Remote Sensing*, vol. 15, issue 5, 1390, 2023. doi: <https://doi.org/10.3390/rs15051390>
14. D. Gómez, P. Salvador, J. Sanz, J.L. Casanova, "A new approach to monitor water quality in the Menor sea (Spain) using satellite data and machine learning methods," *Environmental Pollution*, vol. 286, 117489, 2021. doi: <https://doi.org/10.1016/j.envpol.2021.117489>

15. G. Jakovljevic, M. Govedarica, F. Alvarez-Taboada, "Water body mapping: A comparison of remotely sensed and GIS open data sources," *International Journal of Remote Sensing*, vol. 40, issue 8, pp. 2936–2964, 2018. doi: <https://doi.org/10.1080/01431161.2018.1538584>
16. *DHI Education Resource*. Available: https://raw.githubusercontent.com/DHI/Intro_ML_course/main/module_6/Brahmaputra_images.zip
17. Z. Zhou, *The Application of Fuzzy Neural Network Based on T-S Model in Water Quality Evaluation. Master's Thesis*. East China University of Political Science and Law, Nanjing, China, 2007.
18. X. Wang, *MATLAB Neural Networks 43 Case Studies*; 1st ed. Beijing, China: Beijing University of Aeronautics and Astronautics Press, 2013, pp. 288–289.
19. O.I. Chumachenko, "Deep learning classifier based on nefclass neural network," *Electronics and Control Systems*, vol. 3, no. 49, 2016. doi: <https://doi.org/10.18372/1990-5548.49.11242>
20. Jamileh Yousefi, "A modified NEFCLASS classifier with enhanced accuracy-interpretability trade-off for datasets with skewed feature values," *Fuzzy Sets and Systems*, vol. 413, pp. 99–113, 15 June 2021. doi: <https://doi.org/10.1016/j.fss.2020.07.011>
21. H.Y. Hong, A. Jaafari, E.K. Zenner, "Predicting spatial patterns of wildfire susceptibility in the Huichang County, China: An integrated model to analysis of landscape indicators," *Ecological Indicators*, vol. 101, pp. 878–891, 2019. doi: <https://doi.org/10.1016/j.ecolind.2019.01.056>
22. Jinyue Chen et al., "Remote Sensing Big Data for Water Environment Monitoring: Current Status, Challenges, and Future Prospects," *Earth's Future*, vol. 10, issue 2, 2022. doi: <https://doi.org/10.1029/2021EF002289>
23. L.L. Tang, S. Zhang, J.H. Zhang, Y. Liu, Y. Bai, "Estimating evapotranspiration based on the satellite-retrieved near-infrared reflectance of vegetation (NIRv) over croplands," *GIScience and Remote Sensing*, vol. 58, issue 6, pp. 889–913, 2021. doi: <https://doi.org/10.1080/15481603.2021.1947622>
24. *Normalized difference turbidity index*. Available: <https://developers.arcgis.com/python/latest/samples/river-turbidity-estimation-using-sentinel2-data/>
25. *Turbidity Index (NTU)*. Available: <https://www.summerland.ca/docs/default-source/works-and-utilities/water/turbidity-index.pdf?sfvrsn=2>
26. Nazarij Buławka, Hector A. Orengo, Iban Berganzo-Besga, "Deep learning-based detection of qanat underground water distribution systems using HEXAGON spy satellite imagery," *Journal of Archaeological Science*, vol. 171, 106053, 2024. doi: <https://doi.org/10.1016/j.jas.2024.106053>
27. Bingxue Zhao, Lei Wang, "Surface water monitoring from 1984 to 2021 based on Landsat time-series images and Google Earth Engine," *Heliyon*, vol. 10, issue 17, e36660, 2024. doi: <https://doi.org/10.1016/j.heliyon.2024.e36660>

Received 28.11.2024

INFORMATION ON THE ARTICLE

Serhii V. Klimov, ORCID: 0000-0002-5993-847X, National University of Water and Environmental Engineering, Ukraine, e-mail: s.v.klimov@nuwm.edu.ua

Tetiana V. Starovoit, ORCID: 0009-0008-6335-7679, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: starovoyt.tania@iill.kpi.ua

ДОВГОСТРОКОВИЙ МОНІТОРИНГ ЯКОСТІ ПОВЕРХНЕВИХ ВОД ТА ПОТЕНЦІАЛУ ПІДЗЕМНИХ ВОД ІЗ ВИКОРИСТАННЯМ ОБЧИСЛЮВАЛЬНОГО ІНТЕЛЕКТУ, GIS-ТЕХНОЛОГІЙ ТА ДИСТАНЦІЙНОГО ЗОНДУВАННЯ / С.В. Клімов, Т.В. Старовойт

Анотація. Дефіцит води і зниження її якості через зростання населення, урбанізацію, індустріалізацію й зміну клімату підкреслюють важливість ефективного керування водними ресурсами. Досягнення в дистанційному зондуванні, хмарних обчисленнях та обчислювальному інтелекті підкреслюють необхідність використання сучасних технологій для моніторингу якості поверхневих вод. Це дослідження містить розроблення гібридних інтелектуальних моделей із використанням зображень Landsat та Sentinel-2 і даних WISE із гібридними мережами глибокого навчання для оцінювання якості поверхневих вод та потенціалу підземних вод. Кореляційний аналіз виявив сильні зв'язки між даними дистанційного зондування та параметрами якості води (такими як хлорофіл-а, розчинений кисень, азот та фосфор). Гібридні моделі перевершили традиційні методи машинного навчання, продемонструвавши свою ефективність у реальному керуванні водними ресурсами.

Ключові слова: обчислювальний інтелект, нечітка логіка, дистанційне зондування, супутникові знімки, моніторинг якості поверхневих вод, оцінювання потенціалу підземних вод, гібридні нейронні мережі, NEFCLASS-EM, TS-FNN, Fuzzy C-Means, K-Means.