Secure Systems Design and Technology Development. Veliko Tarnovo, Bulgaria
Proceedings of the International Scientific Conference, 2026. Volume III, 83-89

Evaluation of the Performance of Classification Algorithms in the QGIS System

Peter Grabusts
Engineering Centre
Rezekne Academy of Riga Technical University
Rezekne, Latvia
peteris.grabusts@rtu.lv

Jurijs Musatovs
Engineering Centre
Rezekne Academy of Riga Technical University
Rezekne, Latvia
jurijs.musatovs@rtu.lv

***Abstract*—The use of lake reeds as a source of green energy is becoming increasingly significant. Accurate estimation of reed-covered areas is essential, as it enables forecasting of potential biomass yields and supports carbon footprint calculations. However, in Latvia, there is currently a lack of comprehensive research on the availability and spatial distribution of reeds. The main objective of this research is to develop a method for modeling the distribution of lake reeds in order to predict their future availability. The research focuses on satellite imagery of lakes, with spatial analysis and modeling performed using QGIS. A critical stage in the methodology is classification, which allows reeds to be distinguished from surrounding land cover types and enables estimation of their extent within lake environments. For this purpose, several semi-automatic classification algorithms were applied, including Minimum Distance and Random Forest. The performance of these algorithms was evaluated experimentally, and the most suitable approach was selected based on its agreement with available historical data.**

Keywords—reeds, classification algorithms, modeling, satellite data, QGIS

I. Introduction

Alongside economic development, demand for renewable energy is increasing due to its sustainability and lower environmental impact compared to fossil fuels. Among renewable energy sources, biomass remains one of the most widely utilized. Current trends emphasize the use of non-traditional biomass sources, including cultivated energy crops, wood residues, and aquatic vegetation such as reeds.

With rising fossil fuel prices and growing concern over environmental pollution, interest in locally available biomass resources for energy production is steadily increasing. As a result, the rational and efficient use of reed resources has become an important research topic.

Reed (*Phragmites australis*) [1] is a widespread plant species in Latvia that has traditionally been regarded as a nuisance due to its rapid expansion in water bodies. Reed-covered areas continue to increase, and mowing is often carried out to control overgrowth in lakes and other aquatic environments. At the same time, these ecosystems contain substantial amounts of reed biomass, which remains an underutilized renewable energy resource. Despite its high availability and potential, reed is currently used only to a limited extent in Latvia, primarily in construction, such as for roofing materials.

Existing literature reports a wide range of values for reed biomass yield. For instance, the amount of biomass obtained from winter harvesting varies significantly, ranging from 3 to 30 t/ha [2]. This variability highlights the need for more precise and location-specific assessments.

To evaluate the potential use of reeds as a biomass resource, it is essential to accurately determine reed-covered areas within individual lakes and to monitor their changes over time. Such analysis enables the prediction of trends in reed expansion or decline. In this context, Geographic Information Systems (GIS), such as QGIS, provide powerful tools for analyzing satellite imagery across different time periods, calculating reed area within defined reference intervals, and modeling future spatial dynamics [3] – [6].

The paper [7] examines a significant ecological monitoring problem – changes in the degradation and distribution of aquatic reeds in European lakes. Given that traditional monitoring methods are expensive and time-consuming, the study shows that Green Li-DAR technology can significantly improve the efficiency and accuracy of monitoring. This is especially important for smaller lakes, where the resolution of satellite images is insufficient.

The paper [8] analyses the biomass potential of common reed in the wetlands of the Syr Darya Delta region of Kazakhstan using remote sensing data and Random Forest machine learning models. The authors combine field measurements with high-resolution Sentinel satellite images to create a spatially detailed map of reed biomass and estimate the resource volume in arid climate conditions

Online ISSN 3044-7771
https://doi.org/10.68302/std2026.vol3.196

near the former Aral Sea. The main advantage of the study is the methodological approach - the integration of precise field data with satellite images and improved regression models allows for reliable biomass estimates. The performance indicators of the models are particularly convincing, indicating high forecast accuracy. The authors find that the largest biomass is concentrated along river branches and lake shores, where water availability is stable.

This research aimed to develop a methodology for identifying reed-covered areas in lakes. Satellite imagery was obtained from the Copernicus Sentinel database and processed using the QGIS environment. Known land cover classes – such as water, forest, soil, grass, and reeds – were identified and used for classification. The accuracy of the results depends heavily on how precisely these classes are defined.

To determine the most suitable approach, several classification algorithms were tested, including Minimum Distance, Multilayer Perceptron, Random Forest, and Support Vector Machine. A significant amount of time was devoted to selecting appropriate satellite images, as cloud cover is frequent in Latvia even during summer, making image quality a critical factor.

The results demonstrated that the estimated proportion of reed-covered areas in Lake Rāzna is consistent with findings reported in other scientific studies, indicating that the proposed methodology has strong potential for further development. This work builds upon previous research presented in article [9].

II. Materials and Methods

A. Research Object

The object of this research is Lake Rāzna (see Fig. 1), which is the second-largest lake in Latvia [10]. The lake's key characteristics are summarized in Table 1.

Fig. 1. Lake Rāzna.

TABLE 1 LAKE CHARACTERISTIC

Characteristic	Parameter
Coordinates	56°19′N 27°27′E
Area of water surface	5,756.4 hectare
Mean depth	7.0 metre(s)
Maximum depth	17.0 metre(s)
volume	405.000 million solid metres
Length of lakeshore line	43.5 kilometre
Maximum length	12.10 kilometre
Maximum width	7.00 kilometre
Average altitude (above sea level)	163.40 metre(s)
Maximum altitude (above sea level)	163.50 metre(s)

Sentinel-2 images of Lake Rāzna look very impressive, because it is a big water square and the complex shoreline is very visible even with 10 m resolution:

- Lake appears dark blue or almost black because deep water absorbs light.
- Coastal in the shallows, water is brighter, sometimes with a green pasture tone.
- Coastline is very twisted, with many folds and peninsulas.
- Surrounding fields and forests form a contrasting mosaic.

B. Satellite Data and QGIS

The study utilised data from the Copernicus Programme Sentinel satellites, one of the most valuable open-access Earth observation resources worldwide. The Sentinel system consists of a series of satellites developed specifically for the Copernicus programme [11], [12].

Among these, Sentinel-2 plays a particularly important role in land monitoring. Its primary application areas include agriculture, vegetation analysis, land cover classification, and fire monitoring. Sentinel-2 provides high spatial resolution imagery (10–20 m) and frequent revisit times (approximately every 5 days), making it one of the most powerful tools currently available for open-access environmental mapping.

Satellite data from Copernicus support a wide range of users, including service providers, governmental institutions, and research organisations, enabling efficient acquisition and analysis of environmental information. These data and derived information services are freely available to users, significantly enhancing accessibility.

In this research, Sentinel-2 satellite imagery was used for the detection of lake reed areas, and the data were subsequently processed using the QGIS environment [13].

QGIS is a free and open-source Geographic Information System (GIS) that enables users to view, edit, and analyse spatial data, create maps, process satellite imagery, and perform spatial analysis and modelling. It is a powerful tool for cartographic production and geospatial data processing, offering extensive functionality through a wide range of plugins. QGIS supports both raster and vector data formats and integrates numerous open-source libraries to ensure compatibility with diverse geospatial datasets.

QGIS is one of the most widely used GIS platforms globally due to its lack of licensing costs, cross-platform compatibility, support for multiple data formats, and a large ecosystem of community-developed plugins.

Among these plugins, the Semi-Automatic Classification Plugin (SCP) was used in this study. SCP is one of the most powerful and widely used tools within QGIS for processing and analysing Sentinel-2 satellite imagery. In this study, it was applied for image classification to distinguish reed-covered areas from other land cover classes.

C. Classification Algorithms

To detect reeds in lakes using GIS, it is necessary to combine several methods. First, satellite data is used to identify various objects in the surrounding field. Second, these objects are classified to distinguish reeds.

We chose the following classification of objects and their designations (see Fig.2).

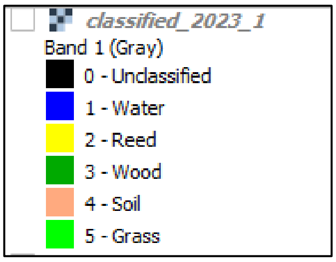

Fig.2. Classification of objects (legends).

The following discusses four popular classification algorithms that will be used in reed recognition: Minimum Distance, Multilayer Perceptron, Random Forest, and Support Vector Machine.

<u>Minimum Distance.</u>

Minimum Distance Estimation (MDE) [14] is a statistical parameter estimation approach that seeks model parameters that minimize the "distance" between the theoretical (model) distribution and the empirical (observed data) distribution.

The essence of the algorithm is as follows: for a given point x, the distances to all candidates (e.g., classes or nodes) are computed, and the one with the smallest distance is selected.

The most commonly used distance measure is the Euclidean distance:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}, \qquad (1)$$

where x and y are vectors in n-dimensional space, and x_i, y_i - their components.

Algorithm steps:

1. Input:
   - Dataset (points or class representatives).
   - New point x that needs to be classified or analyzed.
2. Distance calculation:
   - Calculates the distance between x and each element of the data set.
3. Finding the minimum:
   - Finds the element with the smallest distance.
4. Result:
   - In case of classification: assigns a class to the closest element.
   - In case of graphs: chooses the shortest path.

Advantages: simple implementation, intuitive, and no complex training required.

Disadvantages: slow for large data sets, sensitive to noise and scale, and depends on the chosen distance measure.

<u>Multilayer Perceptron.</u>

A multilayer perceptron (MLP) [15] is an artificial neural network widely used in machine learning and deep learning tasks. It consists of multiple layers of neurons and is widely used in classification and regression tasks.

MLP consists of:

Input layer – receives data;
Hidden layers – performs calculations and feature transformation;
Output layer – produces a result.
Each neuron calculates a weighted sum and applies an activation function:

$$y = f\left(\sum_i w_i x_i + b\right), \qquad (2)$$

where w_i – weights, b – bias, f – activation function (relu, sigmoid, or others).

Algorithm steps:

1. Initialization:
Randomly initialize weights and biases.
2. Forward propagation:
Data is fed through the network from the input to the output. Each layer computes its own output.
3. Error calculation:
Compares the predicted result with the true value.
Uses a loss function (e.g., MSE or cross-entropy).
4. Backpropagation:
Computes the error gradients with respect to the weights, based on gradient descent.
5. Weight update:
Weights are adjusted to reduce the error.
6. Iteration:
The process is repeated for several epochs until the model converges.

Advantages: able to model nonlinear relationships, universal approximator, flexible architecture.

Disadvantages: requires a lot of data, can overfit, difficult to interpret ("black box"), sensitive to hyperparameters.

Random Forest.

Random Forest [16] is an ensemble machine learning method that relies on combining many decision trees to obtain a more accurate and stable prediction.

The idea of Random Forest is based on the fact that several "weak" models (decision trees) can together form a strong model.

Each tree is trained on a randomly selected subset of data. A random subset of features is used in each splitting step. The final result in classification is obtained by voting (majority vote).

Algorithm steps:

1. Data sampling (Bootstrap):
Randomly select subsets from the original data set (with repetition).
2. Tree building:
For each tree, a random subset of features is selected at each node, and the best distribution is found according to a criterion (e.g., Gini index).
3. Model merging: in classification - voting
4. Prediction for new data: each tree gives a result, and the final result is the combined result.

Advantages: high accuracy, reduces overtraining compared to a single tree, and works well with large and complex data.

Disadvantages: more difficult to interpret than a single decision tree, requires more computing resources, and is slower in prediction.

Support Vector Machine.

Support Vector Machine (SVM) [17] is a supervised machine learning algorithm used for classification and regression. Its basic idea is to find a boundary (hyperplane) that best separates data into different classes.

SVM attempts to find a hyperplane that separates the data into two classes and maximizes the margin to the nearest points:

$$w^T x + b = 0 \tag{3}$$

Support vectors are the points closest to the boundary; they determine the position of the hyperplane, and only these points affect the model.

The SVM solves the equation:

$$\min_{w,b} \frac{1}{2}\|w\|^2 \tag{4}$$

with the condition

$$y_i(w^T x_i + b) \geq 1 \tag{5}$$

This means that the weight norm is minimized and correct classification is ensured.

If the data is not linearly separable, SVM uses kernel functions to transform the data into a higher dimension. The most popular kernels are linear, polynomial, and RBF (Gaussian).

Algorithm steps:

1. Choose a kernel function.
2. Formulate the optimization problem.
3. Solve the quadratic programming problem.
4. Identify support vectors.
5. Construct a classification function.

Advantages: efficient on high-dimensional data, good with small amounts of data, and strong theoretical basis.

Disadvantages: slow for large data sets, sensitive to parameter selection (kernel), and more difficult to interpret.

III. Results and Discussion

In the authors' previous work [9], a study was conducted with satellite data of Lake Rāzna for different years to classify objects with the Random Forest algorithm using QGIS and separate reeds from other objects around the lake.

The research methodology is based on the following sequence of actions:

- a suitable lake map with 5% cloudiness is found using Copernicus;
- suitable objects are identified for classification purposes in the QGIS environment (according to the legends in Figure 2);
- a classification algorithm is selected from SCP options;
- classification is performed, and the obtained results are analyzed.

The classification results are shown in Figure 3 and Table 2.

Table 2 Classification Results (year 2024)

Raster value	Pixel Sum	Percent age %	Area (m^2)	Area (hectare)
1-Water	529140	6.03	52914000	5291.40
2-Reed	266918	3.04	26691800	2669.18
3-Wood	4599805	52.39	459980500	45998.05
4-Soil	412493	4.70	41249300	4124.93
5-Grass	2972380	33.85	297238000	29723.80

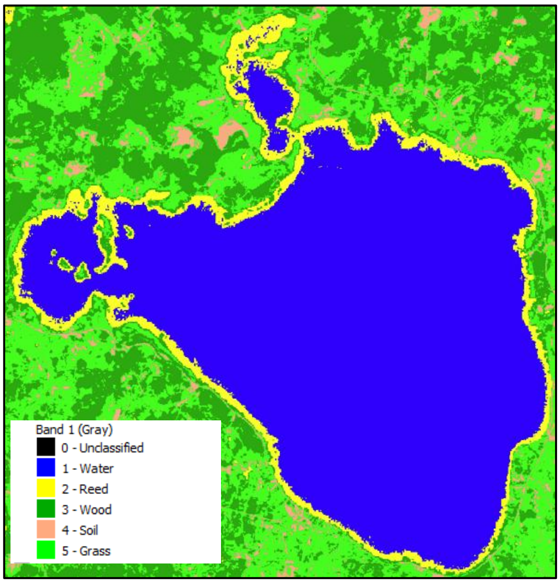

Fig.3. Reeds (yellow) in Lake Rāzna (year 2024).

From now on, we will use only the percentage of reeds in the water surface area of Lake Rāzna in 2024 to evaluate the performance of various classification algorithms.

We will have the opportunity to compare the experimental results with data from [2], which reported that in 2008 the area of reeds constituted approximately 4 percent of the water surface area.

Five experiments were performed for each algorithm. Of course, the initial settings and parameters in classification algorithms are different, but our task is to determine the suitability of the algorithms for classifying reeds.

Minimum Distance.

For the Minimum Distance algorithm in the SCP, the threshold value represents a minimum probability percentage (0–100).

Experiment 1: Very Strict: Only pixels with very high probability (95%+ confidence) are classified. Best for homogeneous areas where you need high confidence, but will leave many pixels unclassified.

Experiment 2: Strict: A common starting point for high-quality results. Filters out most ambiguous pixels while retaining well-defined features.

Experiment 3: Moderate: A balanced threshold. Allows more pixels to be classified while still rejecting many uncertain ones. Useful for heterogeneous landscapes.

Experiment 4: Relaxed: Includes a wider range of probabilities. Suitable for initial exploratory runs or when you want to minimize unclassified areas at the cost of potential noise.

Experiment 5: No Threshold: Classifies every pixel. All pixels are assigned to the class with the highest probability, regardless of how low that probability is. This is the default behavior when the threshold option is unchecked.

The results for the Minimum Distance algorithm are shown in Table 3.

TABLE 3 MINIMUM DISTANCE RESULTS

N	Probability percentage	Reeds (%)
1	95	3.88
2	90	4.09
3	80	4.35
4	70	4.48
5	0	3.45

Score range: 3.45 - 4.48. Average value: 4.05.

Multilayer Perceptron.

Unlike the Minimum Distance algorithm, the MLP is a neural network with many tunable parameters that control its architecture and learning behaviour.

Experiment 1: Baseline (Default Configuration). One hidden layer with 100 neurons, activation – relu, learning rate – 0.001.

Experiment 2: Simple Architecture (Reduced Complexity). Uses a much smaller network with only 10 neurons in a single hidden layer. This creates a simpler model that is less likely to overfit on small training datasets. Activation – relu, learning rate – 0.001.

Experiment 3: Deep Architecture (Increased Complexity). Uses three hidden layers with 50 neurons each. This creates a deeper network capable of learning more complex patterns, at the risk of overfitting. Activation – relu, learning rate – 0.001.

Experiment 4: Strong Regularization (Overfitting Prevention). Increases the alpha regularization parameter from 0.0001 to 0.01. A higher alpha adds more penalty for large weights, producing a simpler, more generalized model. Activation – relu, learning rate – 0.001.

Experiment 5: Slow Learning with SGD. Uses the SGD (Stochastic Gradient Descent) solver, with a higher learning rate (0.01) and tanh activation function. This configuration learns more slowly and may generalize differently.

The results for the MLP algorithm are shown in Table 4.

TABLE 4 MLP RESULTS

N	Model	Reeds (%)
1	Baseline	7.54
2	Simple architecture	4.84
3	Deep architecture	5.87
4	Strong regularization	5.52
5	Slow learning	-

Score range: 4.84 - 7.54. Average value: 5.94.

Random Forest.

For the Random Forest algorithm in the SCP, the main parameter is the number of trees.

Experiment 1: Baseline (Fast Preview). This configuration prioritizes speed and is ideal for initial previews before committing to a full classification.

Experiment 2: Standard Forest. This configuration matches parameters used in published remote sensing studies with SCP. A study on mangrove forest mapping achieved 99.12% accuracy using these settings.

Experiment 3: Large Forest. The SCP documentation explicitly recommends that "Number of trees should be at least 500 for good results" in production classifications.

Experiment 4: Very Large Forest (Maximum Accuracy). For final production maps, where accuracy is critical and processing time is not a limiting factor. The typical range for estimators is 100 to 1000.

Experiment 5: Conservative Splitting (Overfitting Prevention). Requires more samples per internal node – min samples to fit – 5. This is useful when you have a small training dataset and want to avoid overly complex trees that memorize noise rather than generalizing patterns.

The results for the Random Forest algorithm are shown in Table 5.

TABLE 5 RANDOM FORESTS RESULTS

N	Number of trees	Reeds (%)
1	10	5.80
2	100	4.67
3	500	5.06
4	1000	4.63
5	100	5.15

Score range: 4.63 - 5.80. Average value: 5.06.

<u>Support Vector Machine.</u>

For the SVM algorithm in the SCP, the key adjustable parameters are Kernel type, Gamma, and Cost (C). Gamma controls the influence of a single training sample, while Cost (C) controls the trade-off between minimizing error on the training data and maintaining a simple decision boundary. The kernel type traditionally is Radial Basis Function.

Experiment 1: Baseline (Default Configuration). This experiment uses the standard defaults for SVM classification. It serves as a control group for comparing other variations.

Experiment 2: Low Gamma (Smooth Decision Boundary). This experiment uses a smaller Gamma value, which means each training sample has a wider influence radius. The resulting decision boundary will be smoother and more generalized.

Experiment 3: High Gamma (Detailed Decision Boundary). This experiment uses a larger Gamma value, which means each training sample has a narrow influence radius. The decision boundary will be more detailed and can capture complex patterns. A high gamma creates a more detailed, potentially "wiggly" decision boundary. This can capture fine spatial patterns, but risks overfitting if your training data has outliers or noise

Experiment 4: Low Regularization (Higher Penalty for Errors). This experiment increases the Cost parameter. A higher C value tells the algorithm to work harder to classify every training sample correctly, even if that means creating a more complex boundary. With higher C, the classifier tries harder to fit the training data exactly. This can improve accuracy but increases the risk of overfitting to noise in data samples.

Experiment 5: High Regularization (Softer Penalty for Errors). This experiment decreases the Cost parameter. A lower C value allows more training samples to be misclassified if it leads to a simpler, more generalized decision boundary. Lower C creates a softer, more generalized boundary that may ignore some training samples.

The results for the SVM algorithm are shown in Table 6.

TABLE 6 SVM RESULTS

N	Gamma	Cost	Reeds (%)
1	0.1	1	4.71
2	0.01	1	5.36
3	1	1	6.70
4	0.1	10	4.73
5	0.1	0.1	4.77

Score range: 4.71 - 6.70. Average value: 5.25.

The model trained with the classification data from Lake Rāzna produced the following results: Training Accuracy: 98.39%, Validation Accuracy: 98.74%.

The most important limiting factors of the research are the following:

- Geographic limitations.

The research focuses on Latvian lake ecosystems, which may limit the applicability of results to regions with different climatic, hydrological, or ecological characteristics. Limitations related to satellite data and resolution.

- Classification model limitations.

The neural network model trained using SCP was validated on a limited number of lake areas, which may reduce generalizability to other lakes with different hydrological or ecological conditions. Potential misclassification of vegetation may influence the estimated reed-covered area and, consequently, biomass calculations.

IV. CONCLUSIONS

This research evaluated the effectiveness of several classification algorithms for identifying reed-covered areas in Lake Rāzna using Sentinel-2 satellite imagery and the QGIS environment. Four algorithms were experimentally tested — the Minimum Distance algorithm, Multilayer Perceptron, Random Forest, and Support Vector Machine — each through five experiments with different parameter settings.

The Minimum Distance algorithm demonstrated the closest agreement with available historical data: the average calculated reed coverage was 4.05%, which closely matches the estimate reported in a 2008 study [2] (approximately 4% of the lake surface area). The Random Forest algorithm produced an average value of 5.06%, while the Support Vector Machine yielded 5.25%. The Multilayer Perceptron produced the highest average value (5.94%) and also showed the greatest variability between experiments, indicating higher sensitivity to parameter selection.

Overall, the research confirms that the developed methodology — based on Sentinel-2 satellite image processing in the QGIS environment and established classification techniques — is suitable for determining reed distribution and monitoring their spatial changes over time. The Minimum Distance and Random Forest algorithms appear particularly promising: the former provides a simple and reliable baseline estimate, while the latter offers greater suitability for more complex landscape structures and heterogeneous environments.

Future research will focus on improving the assessment of reed biomass potential in lakes and evaluating the carbon footprint associated with industrial reed processing.

V. Acknowledgments

This research was funded by grant number RTU-PA-2024/1-0077, "Towards a sustainable bioeconomy: assessing reed biomass potential and applications (ReedREvolution)".

References

[1] "Plant family," [Online] Available: https://www.britannica.com/plant/Poaceae [Accessed: Feb. 20, 2026].

[2] E.Čubars and G. Noviks, Influence of Different Factors on Productivity of Reed Growths, Environment. Technology. Resources, Proceedings of the 9th International Scientific and Practical Conference. Volume 1, Rezekne, Latvia, 2013. https://doi.org/10.17770/etr2013vol1.801

[3] E.Tiškus, D.Vaičiūtė, M.Bučas, and J. Gintauskas," Evaluation of common reed (Phragmites australis) bed changes in the context of management using earth observation and automatic threshold," European Journal of Remote Sensing, 2022. https://doi.org/10.1080/22797254.2022.2161070

[4] L. Sciuto, F. Licciardello, A.C. Barbera, and G. Cirelli, "A GIS‐based multicriteria decision analysis to reduce riparian vegetation hydrogeological risk and to quantify harvested biomass (Giant reed) for energetic retrieval," Ecological Indicators, Volume 144, November 2022, https://doi.org/10.1016/j.ecolind.2022.109548

[5] S. Baier, N.C. Meneses, J. Geist, and T.Schneider, "Assessment of Aquatic Reed Stands from Airborne Photogrammetric 3K Data," Remote Sens. 14(2), 337, 2022, https://doi.org/10.3390/rs14020337

[6] H. Runkui and L.Yuqin, "Application of remote sensing technique to the investigation on reed resources of Bosten Lake," Chinese Geographical Science, Volume 2, pages 266–276, 1992, https://doi.org/10.1007/BF02664604

[7] S. Baier, N.C. Meneses, J. Geist and T. Schneider, "Assessment of Aquatic Reed Stands from Airborne Photogrammetric 3K Data. Remote Sens," 14(2), 337, 2022. https://doi.org/10.3390/rs14020337

[8] A. Baibagyssov, A. Magiera, N. Thevs and R. Waldhardt, "Resource Characteristics of Common Reed (Phragmites australis) in the Syr Darya Delta, Kazakhstan, by Means of Remote Sensing and Random Forest," Plants, 2025. https://www.mdpi.com/2223-7747/14/6/933

[9] P. Grabusts, J. Musatovs and Ē. Teirumnieka, Modeling the Distribution of Lake Reeds Using GIS, In Proceedings of the 16th International Scientific and Practical Conference Environment. Technology. Resources, Rezekne Latvia, June 2025. https://doi.org/10.17770/etr2025vol2.8619

[10] "Latvian lakes database." [Online] Available: https://www.ezeri.lv/database/2196/ [Accessed: Feb. 20, 2026].

[11] "Copernicus project," [Online] Available: https://www.copernicus.eu/en [Accessed: Feb. 20, 2026].

[12] "SentiWiki Home Page," [Online] Available: https://sentiwiki.copernicus.eu/web/sentiwiki [Accessed: Feb. 20, 2026].

[13] "QGIS project," [Online] Available: https://qgis.org/project/overview/ [Accessed: Feb. 20, 2026].

[14] R.O. Duda, P.E. Hart and D.G. Stork, Pattern Classification, 2nd edition, 2001.

[15] S.Haykin, Neural Networks and Learning Machines, 3rd edition, Pearson, 2009.

[16] L. Breiman, Random Forests. Machine Learning, 45(1), 5–32, 2001. https://doi.org/10.1023/A:1010933404324

[17] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.