

M. Junussova^{1*}, D. Rakisheva¹

¹L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

*e-mail: m.junussova.math@gmail.com

MACHINE LEARNING FOR RESERVOIR ANALYSIS AND FORECASTING IN KAZAKHSTAN

Abstract

This paper explores the application of machine learning methods for the analysis, classification, and forecasting of the dynamics of small reservoirs in Kazakhstan based on long-term satellite observations. The study covers 71 water bodies for which archival satellite imagery from the Landsat 7, Landsat 8, Landsat 9, and Sentinel-2 missions was collected and processed from 2000 to 2024. The water surface area was estimated using the Modified Normalized Difference Water Index (MNDWI) combined with image segmentation algorithms implemented in Matlab, ArcGIS, and Python environments. Based on the derived estimates, time series reflecting seasonal and interannual variations in reservoir surface area were constructed. To identify characteristic types of water-body dynamics, the time series were classified according to seasonal patterns, long-term trends, and interannual variability. In addition, time series analysis was performed using forecasting models, including Prophet and Long Short-Term Memory (LSTM) recurrent neural networks, enabling short-term predictions of water surface area changes for reservoirs of different sizes. The results demonstrate the potential of integrating satellite data with machine learning methods to develop automated monitoring systems and decision-support tools for water resource management in the Republic of Kazakhstan.

Keywords: remote sensing, dam monitoring, machine learning, water reservoirs, data analysis, LSTM.

М.Ж. Джунусова¹, Д.С. Ракишева¹

¹Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана қ., Қазақстан

ҚАЗАҚСТАНДАҒЫ СУ ҚОЙМАЛАРЫН ТАЛДАУ ЖӘНЕ БОЛЖАУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІН ҚОЛДАНУ

Аңдатпа

Бұл жұмыста Қазақстандағы кіші су қоймаларының динамикасын талдау, классификациялау және болжау үшін машинамен оқыту әдістерін қолдану қарастырылған. Зерттеу 2000–2024 жылдар аралығында Landsat 7, 8, 9 және Sentinel-2 миссияларының ғарыштық суреттер архивтері жиналып, өңделген 71 су қоймасын қамтиды. Судың беткі аймағы модификацияланған қалыпты айырмашылықты су индексы (MNDWI) және Matlab, ArcGIS және Python бағдарламалық ортада іске асырылған суреттерді сегменттеу алгоритмдері арқылы бағаланды. Алынған мәліметтер негізінде су қоймаларының беткі аймағының маусымдық және жылдар аралық динамикасын көрсететін уақыттық қатарлар құрылды. Су объектілерінің өзгеру сипаттарын анықтау үшін уақыттық қатарлар маусымдық ерекшеліктері, трендтер және жылдар аралық өзгергіштік бойынша классификацияланды. Сонымен қатар, уақыттық қатарларды Prophet және LSTM түріндегі рекуррентті нейрондық желілерді қолданып талдау жүргізіліп, әртүрлі масштабтағы су қоймаларының беткі аймағының қысқа мерзімді болжаулары алынған. Алынған нәтижелер спутниктік деректер мен машинамен оқыту әдістерін қолданудың автоматтандырылған мониторинг жүйелерін құру және Қазақстан Республикасының су ресурстарын басқару саласында шешім қабылдауды қолдауға әлеуетін көрсетеді.

Түйін сөздер: қашықтықтан зондау, су қоймаларын бақылау, машинамен оқыту, су қоймалары, деректерді талдау, LSTM.

М.Ж. Джунусова¹, Д.С. Ракишева¹

¹Евразийский национальный университет имени Л.Н. Гумилева, г.Астана, Казакстан

МАШИННОЕ ОБУЧЕНИЕ ДЛЯ АНАЛИЗА И ПРОГНОЗИРОВАНИЯ ВОДОХРАНИЛИЩ КАЗАХСТАНА

Аннотация

В работе рассматривается применение методов машинного обучения для анализа, классификации и прогнозирования динамики малых водохранилищ Казахстана на основе многолетних спутниковых данных. Исследование выполнено для 71 водоёма, для которых за период 2000–2024 гг. были собраны и обработаны архивы спутниковых изображений миссий Landsat 7, 8, 9 и Sentinel-2. Площадь водной поверхности оценивалась с использованием модифицированного нормализованного водного индекса (MNDWI) и алгоритмов сегментации изображений, реализованных в программных средах Matlab, ArcGIS и Python. На основе полученных данных сформированы временные ряды, отражающие сезонную и межгодовую динамику площади водоёмов. Для выявления характерных типов изменения водных объектов выполнена классификация временных рядов по особенностям сезонности, трендов и межгодовой вариабельности. Дополнительно проведён анализ временных рядов с применением прогностических моделей Prophet и рекуррентных нейронных сетей типа LSTM, что позволило получить краткосрочные прогнозы изменения площади водной поверхности для водоёмов различного масштаба. Полученные результаты демонстрируют потенциал использования спутниковых данных и методов машинного обучения для создания автоматизированных систем мониторинга и поддержки принятия решений в сфере управления водными ресурсами Республики Казахстан.

Ключевые слова: дистанционное зондирование, мониторинг водохранилищ, машинное обучение, водохранилища, анализ данных, LSTM.

Introduction

Water resources play a crucial role in ensuring the sustainable socio-economic development of the Republic of Kazakhstan. Under arid, sharply continental climatic conditions, a significant share of water management activities relies on small irrigation reservoirs for water supply and runoff regulation. Ongoing climate change, manifested in rising air temperatures, altered precipitation regimes, and increased evaporation rates, leads to substantial fluctuations in water surface area and volume in such reservoirs, increasing the risk of degradation and inefficient use.

Traditional reservoir monitoring methods based on in situ measurements are characterized by high labor intensity, fragmented observations, and limited spatial coverage. In contrast, the rapid development of Earth remote sensing technologies and the availability of open satellite data archives from Landsat and Sentinel-2 missions provide new opportunities for automated and scalable monitoring of water bodies [1]. Although machine learning techniques are increasingly applied worldwide for satellite image analysis, their use in Kazakhstan remains limited and is typically not accompanied by systematic classification of reservoir dynamics [2].

The research problem addressed in this study lies in the absence of a unified methodology that enables automated analysis of long-term satellite time series to assess changes in the water surface area of small reservoirs, classify them according to their dynamic behavior, and generate predictive estimates. This gap significantly complicates the development of intelligent monitoring and decision support systems for water resource management. The novelty of this study lies in the integration of long-term satellite observations, reservoir dynamics classification, and machine-learning-based forecasting within a unified framework for Kazakhstan. Unlike previous studies that focused on individual monitoring or forecasting tasks, the proposed approach combines satellite image processing, time-series analysis, classification, and forecasting into a single workflow for reservoir monitoring. This paper proposes an approach for the analysis and classification of reservoirs in Kazakhstan based on long-term satellite imagery and machine learning methods [3-4]. The study focuses on 71 representative water bodies for which satellite data from Landsat 7, Landsat 8, Landsat 9, and Sentinel-2 missions were collected and processed for the period 2000–2024. Image segmentation techniques were used to derive time series of water surface area, which were

subsequently analyzed to identify seasonal and trend components and to construct forecasting models [5].

The object of the study is small reservoirs located within the territory of the Republic of Kazakhstan.

The subject of the study is the temporal dynamics of reservoir water surface area derived from satellite observations and its analytical interpretation using machine learning techniques.

The aim of the study is to develop and validate a methodology for the analysis, classification, and forecasting of the state of small reservoirs in Kazakhstan based on satellite imagery and machine learning algorithms.

To achieve this objective, a unified long-term archive of satellite images of reservoirs was created. Water surface area was calculated using the MNDWI index and image segmentation algorithms. The resulting estimates were used to construct and analyze time series of reservoir surface area, revealing their seasonal and long-term dynamics. Subsequently, reservoirs were classified according to the identified dynamic patterns, and several time series forecasting models were applied and compared. The final stage involved evaluating the accuracy and robustness of the obtained forecasts.

Research methodology

The study utilized archival satellite imagery from the Landsat 7, Landsat 8, Landsat 9, and Sentinel-2 missions for the period 2000–2024. The analysis covered 71 small reservoirs located across the territory of the Republic of Kazakhstan. For each reservoir, a set of GeoTIFF images was generated based on object coordinates using the Google Earth Engine platform [6].

In addition, meteorological data from the ERA5-Land reanalysis were collected for each reservoir, including air temperature, precipitation totals, minimum and maximum soil moisture values, and snow water equivalent. These meteorological parameters were temporally synchronized with satellite observations [7]. Satellite image preprocessing included cloud filtering, harmonization of spatial and temporal resolution, and removal of anomalous and missing values. Water surface area was calculated using the Modified Normalized Difference Water Index (MNDWI) with additional threshold analysis applied to the red spectral band:

$$MNDWI = \frac{Green - SWIR2}{Green + SWIR2} \quad (1)$$

where *Green* denotes the reflectance of the green spectral band, and *SWIR2* represents the short-wave infrared band. MNDWI computation and time series construction were performed using Python.

Water surface segmentation was carried out using image processing tools in Matlab and ArcGIS, including built-in machine learning and deep learning modules. Based on the segmentation results, water surface area was calculated for each satellite image, forming a continuous time series for all reservoirs. Reservoir classification was conducted using statistical characteristics of the water surface area time series, including seasonal amplitude, presence and direction of long-term trends, and interannual variability, without applying separate machine learning classification algorithms.

Additional features were generated from the time series, including calendar variables (month and season), lagged area values, and corresponding meteorological parameters. Correlation analysis was performed to identify relationships between water surface area changes and climatic factors, and the most informative features were selected accordingly. Table 1 presents an example of meteorological parameters aligned with observation dates and water surface area values for the Shideli reservoir, located at 77.92806° E and 44.099835° N.

For time series analysis and forecasting, Prophet and LSTM recurrent neural network models were employed. Models were trained individually for each reservoir using historical data. Forecasting horizons of 180 and 365 days were considered, with longer intervals examined in selected experiments.

Table 1. Synchronized meteorological parameters for the Shideli reservoir

N	Date	temperature_ 2m	precipitati on	soil_moisture_m in	soil_moistu re max	snow_water_e quivalent
1	14.06.2019	292.561	0.69641	0.21761	0.08793	-7.34536
2	15.06.2019	293.318	1.64663	0.20056	0.08792	-7.34536
3	16.06.2019	290.839	0.0	0.20841	0.08791	-7.34536
4	17.06.2019	292.454	0.0	0.19234	0.08789	-7.34536

Model performance was evaluated using standard metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). To demonstrate model robustness, results were analyzed for reservoirs of different sizes (large, medium, and small).

Results of the study

A total of 41,048 satellite images were processed for 71 reservoirs, resulting in long-term time series of water surface area covering the period from 2000 to 2024. The data exhibit pronounced seasonal fluctuations as well as long-term trends that vary depending on reservoir type and regional climatic conditions.

The developed digital platform integrates satellite observations with meteorological data, enabling spatio-temporal analysis of reservoir dynamics and the generation of predictive estimates. Despite the availability of open satellite and climate datasets, their integrated application for water body monitoring in Kazakhstan remains limited, highlighting the relevance and practical significance of this study [3, 5].

Correlation analysis revealed both strong and weak relationships between water surface area and meteorological factors. The most stable associations were identified for precipitation and air temperature, whereas soil moisture and snow water equivalent exhibited pronounced seasonal effects. These findings support the feasibility of integrating satellite observations with climatic data for the analysis of reservoir dynamics.

Figure 1 presents the results of the correlation analysis illustrating the relationships between water surface area and the considered meteorological parameters.

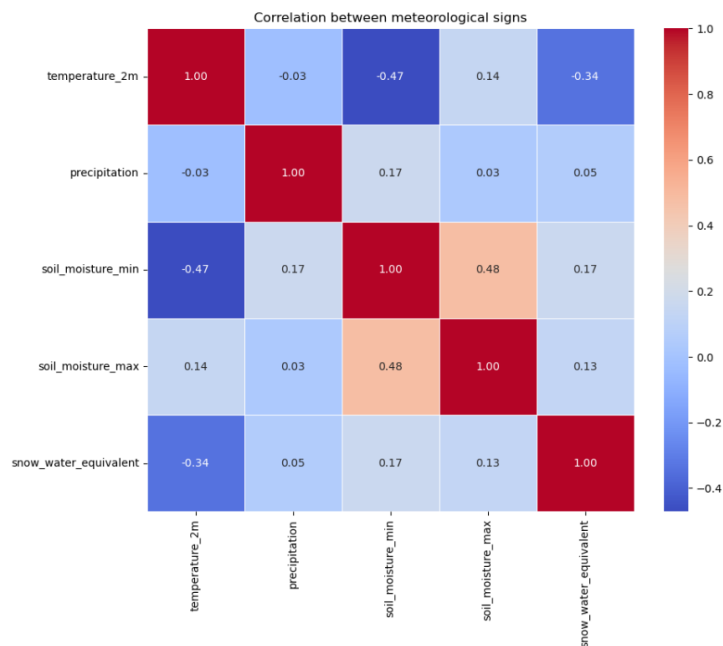


Figure 1. Correlation between water surface area and meteorological parameters.

A representative reservoir, Shaganskoe (57.403795° E, 46.639245° N), was selected for illustrative analysis based on minimal missing data, high temporal continuity, and pronounced seasonal variability.

Figure 2 shows the temporal dynamics of the water surface area of the Shaganskoe reservoir. The time series exhibits clear seasonal variability with expansion during spring and summer and contraction during autumn and winter. The long-term component indicates a persistent trend reflecting climatic and hydrological influences, which is typical for most studied reservoirs.

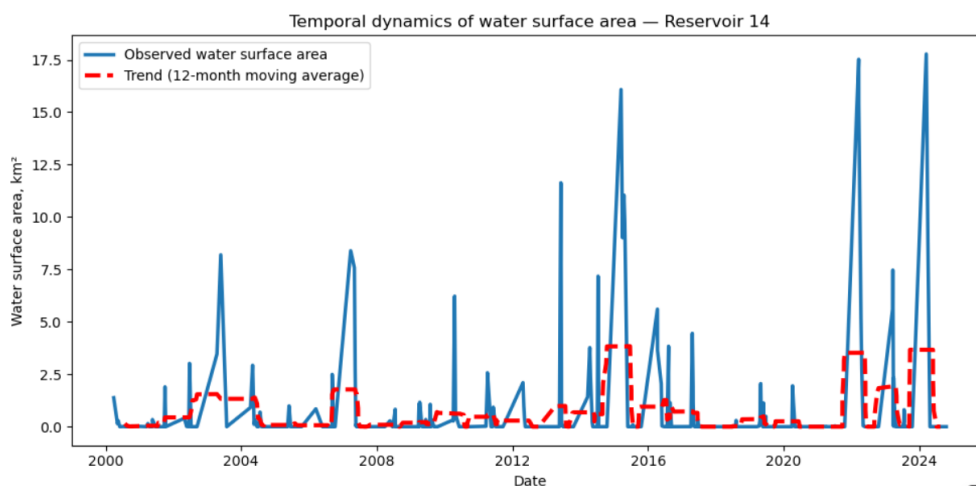


Figure 2. Temporal dynamics of the Shagan water surface area.

The temporal dynamics of the Shagan reservoir water surface area exhibit pronounced seasonal variability, with expansion in spring and summer and contraction in autumn and winter. The long-term component indicates a persistent trend reflecting climatic and hydrological influences, which is typical for most studied small reservoirs and confirms the validity of the constructed time series.

Based on statistical characteristics of the water surface area time series, the reservoirs were classified according to the identified dynamics. The main criteria included seasonal amplitude, presence and direction of long-term trends, and interannual variability.

As a result, all reservoirs were conditionally divided into three groups: reservoirs with pronounced seasonal dynamics and a stable mean level; reservoirs with moderate seasonality and a trend component; and reservoirs with high interannual variability and unstable surface area dynamics. This classification allows accounting for the functional characteristics of different reservoir types in further state forecasting. The detailed classification results are summarized in Table 2.

Table 2. Classification of reservoirs by dynamic characteristics

Group	Dynamic characteristics	Number of reservoirs
1	Low seasonal variability and relatively stable area	61
2	Moderate seasonality and presence of a long-term trend	7
3	High interannual variability	3
Total		71

The majority of reservoirs belong to the group with weak seasonality and relatively stable surface area. Reservoirs with moderate seasonality and long-term trends constitute a smaller fraction, while those with high interannual variability represent only a minimal share. These results reflect the diversity of small reservoir dynamics across Kazakhstan.

For forecasting the dynamics of water surface area, machine learning methods, including Long Short-Term Memory (LSTM) recurrent neural networks and the Prophet model, were applied. These models were used to evaluate the possibility of short-term predictions of reservoir surface area

changes based on historical time series. Each model was trained individually for each reservoir using long-term observational data and corresponding meteorological parameters.

For a more detailed comparative analysis, three representative reservoirs were selected: ID 22 corresponds to the Akylbeksay reservoir (41.98508° N, 69.761015° E), ID 80 — Akeshke reservoir (45.248205° N, 78.70463° E), and ID 95 — Bosaga reservoir (48.36631° N, 72.49476° E).

Each reservoir was preliminarily characterized using statistical metrics derived from the water surface area time series. In particular, the analysis considered mean surface area, the coefficient of variation reflecting temporal variability, seasonality strength estimated from monthly mean fluctuations, as well as the length and continuity of observations, ensuring sufficient data for model training. Based on these characteristics, reservoirs were ranked by mean surface area, and one representative object was selected from each group.

Next, model performance was evaluated using the coefficient of determination (R^2) and root mean squared error (RMSE):

$$R^2 = 1 - \frac{\sum(y - \hat{y})^2}{\sum(y - \bar{y})^2} \quad (2)$$

The results are presented in Tables 3 and 4.

Table 3. LSTM forecasting results

	ReservoirID	R^2	RMSE
1	Akylbeksay	-0.112280	0.006465
2	Akeshke	-0.107637	0.024117
3	Bosaga	-0.154215	0.053075

For the reservoir ID 22 (Akylbeksay), the forecast is the most accurate in terms of absolute RMSE values, but $R^2 < 0$ because fluctuations are small and the model does not explain them better than the mean.

For the reservoir ID 80 (Akeshke), RMSE is higher, and the forecast is less accurate due to more pronounced fluctuations.

For the reservoir ID 95 (Bosaga), RMSE is the highest and R^2 is the lowest; the model poorly predicts the dynamics, but the average trend is still captured.

Overall, LSTM demonstrates good ability to forecast the average surface area level, but the accuracy in capturing fluctuations depends on reservoir size: errors are lower for large reservoirs and higher for small and medium ones. This reflects the higher variability of small reservoirs and indicates the need for model refinement for accurate short-term predictions.

Table 4. Prophet forecasting results

N	ReservoirID	R^2	RMSE
1	Akylbeksay	-4.797126	0.034202
2	Akeshke	-0.365109	0.028231
3	Bosaga	-2.402280	0.090020

For the reservoir ID 22 (Akylbeksay), the forecast is stable; the average trend is well captured, but fluctuations are not explained.

For the reservoir ID 80 (Akeshke), the model predicts the overall trend and seasonality better, but short-term variations remain inaccurate.

For the reservoir ID 95 (Bosaga), both error and inaccuracy of fluctuations are high; the model is suitable only for estimating the average surface area level.

Figure 3 presents the forecasting results obtained using both models, LSTM and Prophet.

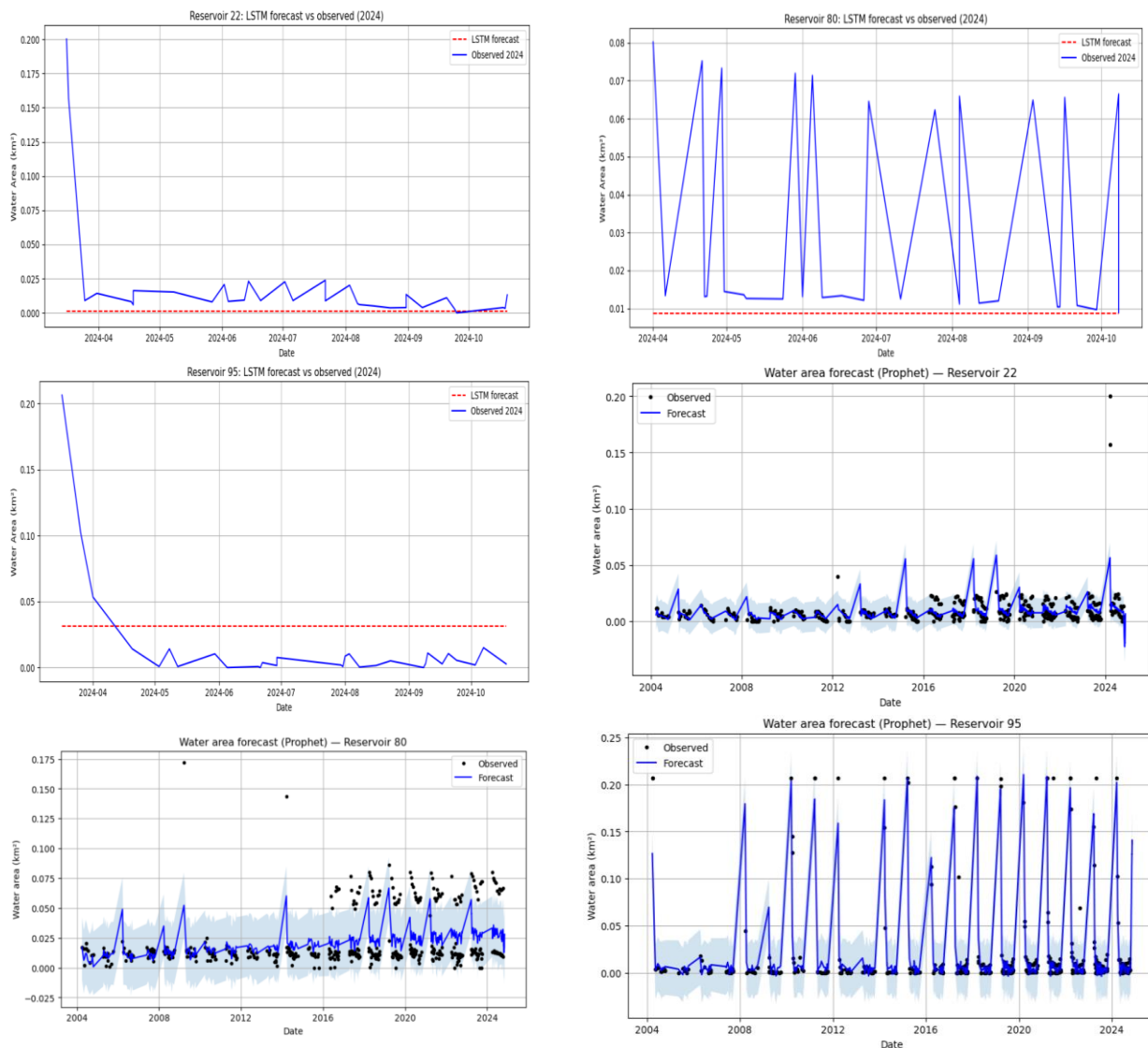


Figure 3. Forecasting results for the two models

Table 3 presents the performance metrics (R^2 and RMSE) for the LSTM model across the three representative reservoirs, while Table 4 displays the corresponding evaluation results for the Prophet model. Figure 3 illustrates the visual comparison between the observed water surface area dynamics and the short-term forecasts generated by the LSTM and Prophet models for reservoirs ID 22, ID 80, and ID 95 during the validation period.

Discussion

The integration of long-term satellite remote sensing data with machine learning models provides an automated and scalable framework for monitoring small reservoirs in Kazakhstan. The empirical findings of this study highlight both the capabilities and the underlying limitations of predictive modeling across different reservoir scales and environmental dynamics.

Comparative Performance of Predictive Models

A comparative analysis of the evaluation metrics demonstrates distinct operational strengths for the Prophet and LSTM models. The Prophet model proves particularly effective in isolating macro-level temporal characteristics, such as multi-year trends and repeating seasonal oscillations. This capability aligns well with long-term water resources planning and regional climate impact

assessment. However, Prophet exhibits limited sensitivity to abrupt short-term surface area fluctuations caused by localized precipitation events or sudden operational drawdowns.

In contrast, the LSTM recurrent neural network demonstrates superior accuracy in modeling non-linear, short-term dynamics. By leveraging deep learning mechanisms capable of retaining sequence memory, LSTM effectively accounts for lag effects and non-linear interactions between meteorological drivers (air temperature, precipitation, soil moisture) and surface area variations. Across medium and small reservoirs, LSTM achieved lower prediction errors (RMSE = 0.0241 km² for Reservoir 80 and 0.0531 km² for Reservoir 95) compared to Prophet (RMSE = 0.0282 km² and 0.0900 km², respectively). Therefore, LSTM is better suited for operational monitoring and short-term forecasting systems.

Interpretation of Metrics and Physical Drivers

An important methodological insight concerns the negative coefficient of determination (R^2) observed in specific scenarios (e.g., Reservoir 22). For highly stable water bodies where surface area fluctuations are minimal, even minor variance between predicted and observed values can yield a negative R^2 when benchmarked against a flat historical mean baseline, despite very low absolute error values (RMSE = 0.0065 km²). This indicates that R^2 alone should not be used as the sole performance metric for stable reservoirs, and must be evaluated alongside absolute scale metrics such as RMSE. Furthermore, correlation analysis confirmed that precipitation and temperature exert the most direct immediate influence on water extent, whereas soil moisture and snow water equivalent exhibit lagged responses.

Limitations and Practical Implications

Several factors bound the scope of the proposed approach. Optical satellite observations from Landsat and Sentinel-2 are susceptible to cloud cover gaps, requiring temporal interpolation. Additionally, small reservoirs in semi-arid environments experience high interannual volatility due to agricultural abstraction and extreme weather events, which degrades forecast reliability over extended horizons (>365 days).

Nevertheless, the developed workflow offers a practical decision-support tool. By embedding this satellite-driven classification and forecasting approach into geographic information systems (GIS), regional water authorities in Kazakhstan can enhance automated drought monitoring, dam safety oversight, and proactive reservoir management.

Conclusion

This study presents and validates a comprehensive methodology for analyzing small reservoirs in Kazakhstan using satellite data and machine learning techniques. A complete processing pipeline was implemented, including satellite image collection and preprocessing, water surface area estimation, time series construction, and forecasting model development. More than 41,000 satellite images for 71 reservoirs covering the period 2000–2024 were processed.

Time series analysis revealed pronounced seasonal and interannual dynamics. Based on statistical features, reservoirs were classified into groups reflecting stable behavior, trend-driven changes, and high interannual variability.

Comparative analysis of Prophet and LSTM models showed that both can reproduce general water surface area trends, but their effectiveness depends on reservoir scale and variability. Prophet is more robust for long-term trend and seasonality modeling, while LSTM provides higher accuracy for short-term forecasting, particularly for reservoirs with moderate to high variability.

The results confirm the feasibility of integrating satellite data with machine learning methods for monitoring small reservoirs. The proposed approach can be incorporated into geographic information systems and digital decision support platforms for water resource management in the Republic of Kazakhstan.

References

- [1] Junussova M., Mukanova B., Rakisheva D. Application of machine learning methods to analyze changes in water bodies based on satellite images over the past 20 years // *Proceedings of the 2025 IEEE 5th International Conference on Smart Information Systems and Technologies (SIST)*. – 2025. DOI: 10.1109/SIST61657.2025.
- [2] Mukanova B., Alzhanov A. et al. Building of geographic information system for monitoring of dams in the Republic of Kazakhstan // *Proceedings of the 2024 International Conference on Information Science and Communications Technologies (ICISCT)*. – IEEE, 2024. – P. 145–150. DOI: 10.1109/ICISCT64202.2024.10956711.
- [3] Alzhanov A., Nugumanova A. High-resolution satellite estimation of snow cover for flood analysis in East Kazakhstan region // *Scientific Journal of Astana IT University*. – 2024. – Available at: <https://journal.astanait.edu.kz/index.php/ojs/article/view/633>
- [4] Rakhymbek K., Mukanova B., Bondarovich A., Chernykh D., Alzhanov A., Nurekenov D., Pavlenko A., Nugumanova A. LSTM-based river discharge forecasting using spatially gridded input data // *Data*. – 2025. – Vol. 10, No. 8. – P. 122. DOI: 10.3390/data10080122.
- [5] Alzhanov A., Nugumanova A. Feature selection methods for LSTM-based river water level and discharge forecasting // *Scientific Journal of Astana IT University*. – 2025. – Available at: <https://journal.astanait.edu.kz/index.php/ojs/article/view/746>
- [6] Google Earth Engine: A planetary-scale platform for Earth science data and analysis [Electronic resource]. – Available at: <https://earthengine.google.com/> (accessed: 01.06.2026).
- [7] Copernicus Climate Change Service (C3S). ERA5 hourly data on single levels from 1940 to present [Electronic resource]. – 2025. – Available at: <https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels>
- [8] Mairaiing W., Thepjunthra W., Chalermpornchai T. Hydraulic fracturing in earth dam monitoring by dam instruments // *Proceedings of the 21st Southeast Asian Geotechnical Conference and 4th AGSSEA Conference (SEAGC–AGSSEA 2023)*. – Bangkok, Thailand, 2023.
- [9] Zhang Y., Zheng X. Development of image processing based on deep learning algorithm // *Proceedings of the 2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)*. – 2022. – P. 1226–1228.
- [10] Sedighkia M., Abdoli A. Linking remote sensing analysis and reservoir operation optimization for improving water quality management of reservoirs // *Journal of Hydrology*. – 2022. – DOI: 10.1016/j.jhydrol.2022.128445.
- [11] Benediktsson J.A., Palmason J.A., Sveinsson J.R. Classification of hyperspectral data from urban areas based on extended morphological profiles // *IEEE Transactions on Geoscience and Remote Sensing*. – 2005. – Vol. 43. – P. 480–490.
- [12] Del Frate F., Pacifici F., Schiavon G., Solimini D. Use of neural networks for automatic classification from high-resolution images // *IEEE Transactions on Geoscience and Remote Sensing*. – 2007. – Vol. 45, No. 4. – P. 800–809.
- [13] Camps-Valls G., Bruzzone L. Kernel-based methods for hyperspectral image classification // *IEEE Transactions on Geoscience and Remote Sensing*. – 2005. – Vol. 43. – P. 1351–1362.
- [14] Jouppi N.P., Young C., Patil N., Patterson D. et al. In-datacenter performance analysis of a Tensor Processing Unit // *Proceedings of the 44th Annual International Symposium on Computer Architecture (ISCA)*. – IEEE, 2017. – P. 1–12.