ISSN: 2224-5227 (Print)
ISSN: 2518-1483 (Online)

ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE

№3
2025

ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

CENTRAL ASIAN ACADEMIC
RESEARCH CENTER

ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE

3 (355)

JULY – SEPTEMBER 2025

PUBLISHED SINCE JANUARY 1963
PUBLISHED 4 TIMES A YEAR

ALMATY, NAS RK

Academic Scientific Journal of Computer Science
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)
Owner: «Central Asian Academic Research Center» LLP (Almaty).
Certificate **№ KZ77VPY00121154** on the re-registration of the periodical printed and online publication of the information agency, issued on **05.06.2025** by the Republican State Institution «Information Committee» of the Ministry of Culture and Information of the Republic of Kazakhstan
Subject area*: information and communication technologies.*
Currently: *included in the list of journals recommended by the CCSES MSHE RK in the direction of «Information and communication technologies».*
Periodicity: *4 times a year.*
http://www.physico-mathematical.kz/index.php/en/

Academic Scientific Journal of Computer Science
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

Меншіктеуші: «Орталық Азия академиялық ғылыми орталығы» ЖШС (Алматы).

Ақпарат агенттігінің мерзімді баспасөз басылымын, ақпарат агенттігін және желілік басылымды қайта есепке қою туралы ҚР Мәдениет және Ақпарат министрлігі «Ақпарат комитеті» Республикалық мемлекеттік мекемесі **05.06.2025** ж. берген **№ KZ77VPY00121154** Куәлік.

Тақырыптық бағыты: *ақпараттық-коммуникациялық технологиялар*

Қазіргі уақытта: *«ақпараттық-коммуникациялық технологиялар» бағыты бойынша ҚР БҒМ БҒСБК ұсынған журналдар тізіміне енді.*

Мерзімділігі: *жылына 4 рет.*

http://www.physico-mathematical.kz/index.php/en/

Academic Scientific Journal of Computer Science
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

Собственник: *ТОО «Центрально-азиатский академический научный центр» (г. Алматы).*

Свидетельство о постановке на переучет периодического печатного издания, информационного агентства и сетевого издания **№ KZ77VPY00121154.** Дата выдачи **05.06.2025**

Тематическая направленность: *информационно-коммуникационные технологии.*

В настоящее время: *вошел в список журналов, рекомендованных КОКСНВО МНВО РК по направлению «информационно-коммуникационные технологии».*

Периодичность: *4 раза в год.*

http://www.physico-mathematical.kz/index.php/en/

CONTENTS

МАЗМҰНЫ

СОДЕРЖАНИЕ

ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE
ISSN 1991-346X
Volume 3. Namber 355 (2025). 271–285

https://doi.org/10.32014/2025.2518-1726.377

IRSTI 70.85.31
UDC 556.3.01:630.2(574)(282.255.5)

A.A. Tlepiyev[1*], A. Mukhamedgali[1], Y.T. Kaipbayev[2], A.N. Kalmashova[2], Y.G. Mukhanbet[2], 2025.
[1]Kazakh-British Technical University, Almaty, Kazakhstan;
[2]Kazakh National Agrarian Research University, Almaty, Kazakhstan.
E-mail: armantlepiev123@gmail.com

SURFACE WATER MONITORING IN KAZAKHSTAN USING NDWI AND RANDOM FOREST: A CASE STUDY OF LAKE AKKOL

Tlepiyev Arman Arystanovich — master student, Kazakh-British Technical University, Almaty, Kazakhstan,
E-mail: armantlepiev123@gmail.com;
Mukhamedgali Adil — PhD, Kazakh-British Technical University, Almaty, Kazakhstan,
E-mail: a.mukhamedgali@kbtu.kz, ORCID iD: https://orcid.org/0009-0003-4437-3757;
Kaipbayev Yerbolat Tolganbayevich — PhD, Associate Professor of the Water Resources and Melioration Department, Kazakh National Agrarian Research University, Almaty, Kazakhstan,
E-mail: yerbolat.kaipbayev@kaznaru.edu.kz, ORCID iD: https://orcid.org/0000-0002-7931-7881;
Kalmashova Ainur Nurlepesovna — PhD, Senior Lecturer of the Water Resources and Melioration Department, Kazakh National Agrarian Research University, Almaty, Kazakhstan,
e-mail: Ainur.Kalmashova@kaznaru.edu.kz, ORCID iD: https://orcid.org/0009-0007-7552-8271
Mukhanbet Yerlan Gabitovich — PhD student, Kazakh National Agrarian Research University, Almaty, Kazakhstan,
E-mail: yerlan.mukhanbet@kaznaru.edu.kz, ORCID iD: https://orcid.org/0009-0006-1365-2042.

Abstract. For nations like Kazakhstan, where dry and semi-arid climates together with human activity put increasing strain on lakes and rivers, monitoring water resources has become ever more crucial in recent years. Accurate and timely information on surface water dynamics is essential for effective water management, environmental protection, and adaptation to climate change. Advances in remote sensing technologies, particularly the use of indices like NDWI and machine learning algorithms such as Random Forest, have significantly enhanced the ability to detect and analyze surface water changes over time. These tools offer scalable, cost-effective solutions for continuous monitoring, especially in remote and vast landscapes typical of Central Asia. This work offers a useful method based on the Normalized Difference Water Index (NDWI) to detect water bodies. Each of the tools we used – QGIS, Python and Google Earth Engine (GEE) – had unique benefits for the work. We applied a supervised Random Forest technique using several spectral bands and indices to separate water covered from dry areas. Examining seasonal and long term fluctuations in water levels, our main case study was on Lake Akkol in the Zhambyl Region. To grasp their influence on local water dynamics, we also examined information from the Assy and Talas rivers. The consistent and dependable results across platforms underlined the great spatial and temporal heterogeneity of water distribution in the area and supported the need for continuous satellite based monitoring.

Key words: Normalized Difference Water Index, Google Earth Engine, QGIS, Python, water resources, monitoring

Acknowledgments

The author would like to thank the developers of QGIS, Python, and GEE for providing open source tools, and the Copernicus Data Space Ecosystem for access to Sentinel 2 data. Special thanks to colleagues and academic advisers for their support throughout the project.

А.А. Тлепиев[1], А. Мухамедгали [1], Е.Т. Кайпбаев [2], А.Н. Калмашова[2], Е.Ғ. Муханбет[2], 2025.

[1]Қазақстан-Британ техникалық университеті, Алматы, Қазақстан;
[2]Қазақ ұлттық аграрлық зерттеу университеті, Алматы, Қазақстан.
E-mail: armantlepiev123@gmail.com

ҚАЗАҚСТАНДАҒЫ БЕТКІ СУЛАРДЫ NDWI ЖӘНЕ RANDOM FOREST ӘДІСІ АРҚЫЛЫ МОНИТОРИНГІЛЕУ: АҚКӨЛ КӨЛІНІҢ МЫСАЛЫНДА

Тлепиев Арман Арыстанович — Қазақстан-Британ техникалық университетінің магистранты, Алматы, Қазақстан,
E-mail: armantlepiev123@gmail.com,
Мухамедгали Адиль — PhD, Қазақстан-Британ техникалық университеті, Алматы, Қазақстан,
E-mail: a.mukhamedgali@kbtu.kz; ORCID iD: https://orcid.org/0009-0003-4437-3757;
Кайпбаев Ерболат Толганбаевич — PhD, «Су ресурстары және мелиорация» кафедрасының қауымдастырылған профессоры, Қазақ ұлттық аграрлық зерттеу университеті, Алматы, Қазақстан,
E-mail: yerbolat.kaipbayev@kaznaru.edu.kz, ORCID iD: https://orcid.org/0000-0002-7931-7881;
Калмашова Айнур Нурлепесовна — PhD, Су ресурстары және мелиорация» кафедрасының аға оқытушысы, Қазақ ұлттық аграрлық зерттеу университеті, Алматы, Қазақстан,
E-mail: Ainur.Kalmashova@kaznaru.edu.kz, ORCID iD: https://orcid.org/0009-0007-7552-8271;
Муханбет Ерлан Габитович — Қазақ ұлттық аграрлық зерттеу университетінің докторанты, Алматы, Қазақстан,
E-mail: yerlan.mukhanbet@kaznaru.edu.kz, ORCID iD: https://orcid.org/0009-0006-1365-2042.

Аннотация. Қазақстан сияқты құрғақ және жартылай құрғақ климат белдеулерінде орналасқан елдер үшін, адамның шаруашылық әрекетімен қатар, көлдер мен өзендерге түсетін жүктеме жылдан-жылға артып келеді. Сондықтан су ресурстарын бақылау соңғы жылдары айрықша маңызды бола түсті. Беткі сулардың динамикасы туралы дәл әрі уақытылы ақпарат – су ресурстарын тиімді басқару, қоршаған ортаны қорғау және климаттың өзгеруіне бейімделу үшін аса қажет. Қашықтықтан зондтау технологияларының дамуымен, әсіресе NDWI сияқты индекстер мен Random Forest сияқты машиналық оқыту алгоритмдерінің қолданылуы, уақыт ішінде беткі судағы өзгерістерді анықтау мен талдау мүмкіндіктерін едәуір арттырды. Бұл құралдар, әсіресе Орталық Азияға тән кең және қашық аймақтарда, тұрақты бақылау жүргізу үшін тиімді және үнемді шешімдер ұсынады. Бұл жұмыста су айдындарын анықтау үшін нормаланған су индексіне (NDWI) негізделген тиімді әдіс ұсынылады. Қолданылған әрбір құралдың – QGIS, Python және Google Earth Engine (GEE) – өзіндік артықшылықтары болды. Біз су мен құрғақ жерлерді ажырату үшін бірнеше спектралды арналар мен индекстерді пайдалана отырып, Random Forest бақылаулы әдісін қолдандық. Маңызды зерттеу нысаны ретінде Жамбыл облысындағы Аққөл көлі алынып, су деңгейінің маусымдық және ұзақ мерзімді өзгерістері зерттелді. Сондай-ақ, жергілікті су динамикасына әсерін түсіну мақсатында Асы және Талас өзендері жөніндегі ақпарат та қарастырылды. Әртүрлі платформаларда алынған тұрақты әрі сенімді нәтижелер өңірдегі су таралуының кеңістіктік және уақыттық әркелкілігін көрсетті және спутниктік бақылаудың үздіксіз жүргізілу қажеттілігін дәлелдеді.

Түйін сөздер: нормаланған су индексі (NDWI), Google Earth Engine, QGIS, Python, су ресурстары, мониторинг

А.А. Тлепиев[1], А. Мухамедгали[1], Е.Т. Кайпбаев[2], А.Н. Калмашова[2], Е.Г. Муханбет[2], 2025.

[1]Казахстанско-Британский технический университет, Алматы, Казахстан;
[2]Казахский национальный аграрный исследовательский университет, Алматы, Казахстан.
E-mail: yerbolat.kaipbayev@kaznaru.edu.kz

МОНИТОРИНГ ПОВЕРХНОСТНЫХ ВОД В КАЗАХСТАНЕ С ИСПОЛЬЗОВАНИЕМ NDWI И МЕТОДА СЛУЧАЙНОГО ЛЕСА: НА ПРИМЕРЕ ОЗЕРА АККОЛЬ

Тлепиев Арман Арыстанович — магистрант Казахстанско-Британского технического университета, Алматы, Казахстан,
E-mail: armantlepiev123@gmail.com,
Мухамедгали Адиль — PhD, Казахстанско-Британский технический университет, Алматы, Казахстан,
E-mail: a.mukhamedgali@kbtu.kz, ORCID iD: https://orcid.org/0009-0003-4437-3757;
Кайпбаев Ерболат Толганбаевич — PhD, ассоциированный профессор кафедры «Водные ресурсы и мелиорация», Казахский национальный аграрный исследовательский университет, Алматы, Казахстан,
E-mail: yerbolat.kaipbayev@kaznaru.edu.kz, ORCID iD: https://orcid.org/0000-0002-7931-7881;

Калмашова Айнур Нурлепесовна — PhD, старший преподаватель кафедры «Водные ресурсы и мелиорация», Казахский национальный аграрный исследовательский университет, Алматы, Казахстан,
E-mail: ainur.kalmashova@kaznaru.edu.kz, ORCID iD: https://orcid.org/0009-0007-7552-8271;
Муханбет Ерлан Габитович — докторант Казахского национального аграрного исследовательского университета, Алматы, Казахстан,
E-mail: yerlan.mukhanbet@kaznaru.edu.kz, ORCID iD: https://orcid.org/0009-0006-1365-2042.

Аннотация. Для таких стран, как Казахстан, где засушливый и полузасушливый климат в сочетании с человеческой деятельностью усиливают нагрузку на озера и реки, мониторинг водных ресурсов в последние годы становится все более важным. Точная и своевременная информация о динамике поверхностных вод необходима для эффективного управления водными ресурсами, охраны окружающей среды и адаптации к изменениям климата. Развитие технологий дистанционного зондирования, особенно использование индексов, таких как NDWI, и алгоритмов машинного обучения, например, случайного леса (Random Forest), значительно повысило возможности выявления и анализа изменений поверхностных вод во времени. Эти инструменты обеспечивают масштабируемые и экономически эффективные решения для постоянного мониторинга, особенно в удалённых и обширных ландшафтах, характерных для Центральной Азии. В данной работе предлагается эффективный метод обнаружения водоёмов на основе нормализованного водного индекса (NDWI). Каждый из использованных нами инструментов – QGIS, Python и Google Earth Engine (GEE) – имел свои уникальные преимущества. Мы применили контролируемый метод случайного леса с использованием нескольких спектральных каналов и индексов для разделения водоёмов и сухих участков. Основным объектом исследования стало озеро Акколь в Жамбылской области, на котором мы изучали сезонные и долгосрочные колебания уровня воды. Также были проанализированы данные по рекам Асы и Талас для оценки их влияния на местную водную динамику. Последовательные и надёжные результаты, полученные на разных платформах, подчеркнули высокую пространственную и временную неоднородность распределения водных ресурсов в регионе и подтвердили необходимость постоянного спутникового мониторинга.

Ключевые слова: нормализованный водный индекс (NDWI), Google Earth Engine, QGIS, Python, водные ресурсы, мониторинг

Introduction. Particularly in dry and semi-arid areas like southern Kazakhstan, water is among the most vital natural resources for maintaining life, agriculture, and economic development. But climate change, upstream water diversion, poor irrigation techniques, and growing population pressure have made water availability in the region quite erratic. This fluctuation jeopardizes not only ecosystems but also the livelihoods of those depending on a constant water supply for home consumption and agriculture (Kozykeyeva, 2020).

Monitoring surface water dynamics across vast distances and over several times has proved to depend critically on remote sensing technologies. Among these, the Normalized Difference Water Index (NDWI) is among the most useful spectral indices for satellite image based open water body detection. Originally suggested by McFeeters in 1996 (Straube, 2013), NDWI highlights water features by enhancing the reflectance difference between the green and near infrared (NIR) bands, hence suppressing non water features including soil and plant. Sentinel 2 MSI Level 2A imagery was used in this work to calculate NDWI by means of atmospherically corrected surface reflectance data. NDWI values were computed espe- cially using Band 03 (Green, 560 nm) and Band 8A (Narrow NIR, 865 nm). Downloaded with a spatial resolution of 20 meters from the Copernicus Data Space Ecosystem platform (Gao, 1996), the imagery was focusing on the monitoring of Lake Akkol, a representative inland water body prone to seasonal and interannual fluctuations in the Zhambyl Region of Kazakhstan, this study aims. Important tributaries like the Assy and Talas rivers, which help to sustain water levels in surrounding lakes and marshes, have hydrologic impact across the area.

Three platforms – QGIS (Messager, 2016 and Breiman, 2001) and Google Earth Engine (GEE) – were used to compute NDWI and improve water classification and change detection. Moreover, a Random Forest machine learning model (U.S. Geological Survey,2024) was trained to use spectral indices and bands to distinguish non water from water locations. This combined approach offers a strong, flexible framework for evaluating the strengths and limits of any computational technique as well as for studying water dynamics.

The main goal of this work is to show how combining remote sensing indices with machine learning can improve the accuracy and efficiency of surface water monitoring in areas sensitive to water shortage.

Study Area and Data. The study focuses on Lake Akkol, in the semi arid continental environment of southern Kazakhstan's Zhambyl Region, with hot summers and cold winters. Lake Akkol is a small inland water feature whose water cover varies greatly both seasonally and yearly. Maintaining rural livelihoods, biodiversity, and local agriculture as well as helping to support.

The hydrology of the lake is affected by nearby tributary rivers, particularly the Talas River and Assy River, which are essential seasonal sources of inflow. Apart from surface runoff, these rivers affect the temporal dynamics of the water extent of the lake in seasons of snowmelt and spring rains. As its precipitation and evaporation rates vary as well, the region is a relevant test site for evaluating water resource changes under climatic stress.

For image processing and analysis, the area of interest (AOI) was selected to be a buffer zone around the lake and inflow channels.

This work takes use of freely accessible Sentinel 2 MSI Level 2A satellite photos including surface reflectance data modified for atmospheric impacts, obtained from the Ecosystem site of Copernicus Data Space (Gao, 1996).

Sentinel 2 offers high resolution multispectral data at 10–20 meter spatial resolution, which is well suited for detecting and monitoring surface water features. Two specific bands were used for water index analysis:

- Band 03 (Green, 560 nm)
- Band 8A (Narrow Near Infrared, 865 nm)

The work focused on just low cloud cover or cloud free models. Long term water dynamics and interannual comparisons were investigated with photos spanning many years: 2016–2024.

This work manually acquired Sentinel 2 photos from the Copernicus Data Space Ecosystem web portal (Gao, 1996). Visual and metadata based filtering was part of the chosen strategy to guarantee low cloud coverage or cloud free conditions over the Lake Akkol area. Scenes were selected depending on:

- **Cloud Coverage:** less than 10% total cloud cover.
- **Spatial Resolution:** 20 meter products (B03 and B8A bands).
- **Temporal Range:** July–August for representative seasonal water extent (2016 and 2024).
- **Level:** Sentinel 2 MSI Level 2A (surface reflectance, atmospherically corrected).

Images were previewed directly in the platform's browser interface using the quick look rendering option and downloaded as SAFE products. Bands were later extracted and preprocessed using Python and QGIS tools (Messager, 2016 and Breiman, 2001) for consistency across platforms.

№1 table – Summary statistics and metadata for Sentinel-2 Bands B8A and B03

Property	B8A	B03
Minimum Value	975	1038
Maximum Value	10077	11420
Mean Value	3536.03	2456.12
Standard Deviation	649.71	712.08
Projection Method	Universal Transverse Mercator (UTM)	
CRS	EPSG:32642 – WGS 84 / UTM zone 42N	

In addition to satellite imagery, the study incorporated vector data from the HydroLAKES database (European Space Agency, 2024), which provides global lake boundary polygons. These prevali-dated lake polygons were used to extract training samples for the Random Forest model, enabling more accurate and consistent labeling of water bodies across different image sources. Pixels located within HydroLAKES polygons were labeled as ”water,” while pixels outside were used to represent ”nonwater” classes.

№2 table – Classification metrics with definitions and formulas

Metric	Purpose	Formula
Accuracy	Overall correctness of predictions. May be misleading with imbalanced classes.	$\frac{TP+TN}{TP+TN+FP+FN}$

Precision	How many predicted positives (e.g. wa- ter) are actually correct. Important when false positives are costly.	$\frac{TP}{TP+FP}$
Recall	Measures how many actual positives are correctly detected. Crucial when missing water areas is critical.	$\frac{TP}{TP+FN}$
F1 score	Balances precision and recall. Reliable metric for imbalanced datasets.	$\frac{2 \cdot Precision \cdot Recall}{Precision+Recall}$

Where: TP – true positives, FP – false positives, TN – true negatives, FN – false negatives

This approach allowed for the creation of a balanced and geographically diverse training dataset, improving the robustness of the machine learning classification and supporting generalization across different types of water bodies.

The interpretation scale presented in Table 3 is grounded in widely accepted practices for evaluating classification models across domains such as remote sensing, medical diagnostics, and information retrieval. Although there is no universally standardized threshold, the defined value ranges (e.g., Excellent: 90 to 100%, Good: 75 to 84%) are commonly used in applied machine learning literature and supported by evidence from prior studies.

Maxwell et al. (2018) highlight that overall accuracy values above 90% are typically considered high performing in remote sensing classification tasks, especially when the dataset is well prepared and balanced. Similarly, Sokolova and Lapalme (2009) emphasize that metrics such as precision and F1 score provide more nuanced insights in imbalanced classification problems, which are common in environmental data analysis.

№3 table – Qualitative interpretation scale for classification metrics

Interpretation	Value Range (%)
Excellent	90 to 100
Very Good	85 to 89
Good	75 to 84
Moderate	60 to 74
Poor	Below 60

This qualitative scale helps provide intuitive, interpretable guidance when assessing model effectiveness, especially in interdisciplinary contexts where stakeholders may not be familiar with the technical significance of raw metric values. It also supports communication of results to non specialist audiences and aligns with performance reporting practices recommended in tools such as scikit learn (Gillies et al., 2022).

Methods and materials. This study combines remote sensing techniques with machine learning to detect and monitor surface water dynamics in the Lake Akkol region. The workflow consists of three main components: NDWI calculation using multiple platforms, training data preparation using HydroLAKES polygons (European Space Agency, 2024), and supervised classification using the Random Forest algorithm (Breiman, 2001).

The Normalized Difference Water Index (NDWI) (Straube et.al., 2013) was computed using the following formula:

$$NDWI = \frac{GREEN - NIR}{GREEN + NIR}$$

where:

Green is Band 03 (560 nm),

NIR is Band 8A (865 nm) from Sentinel 2 MSI Level 2A imagery.

To evaluate and compare processing flexibility, speed, and visual output quality, NDWI was calculated using three platforms:

QGIS: NDWI was computed using the Raster Calculator (Messager et.al., 2016). This approach was useful for localized analysis and manual quality control.

Python: A Python script utilizing rasterio, numpy, and matplotlib was developed to automate NDWI calculation and integrate with machine learning pipelines (U.S. Geological Survey, 2024).

Google Earth Engine (GEE): The NDWI was calculated and visualized using GEE's JavaScript API. This method allowed efficient processing of large datasets across multiple years.

To create labeled training data for the classification model, vector polygons from the HydroLAKES global database were used (European Space Agency, 2024). These polygons represent the outlines of known lakes and reservoirs.

Pixels within the HydroLAKES polygons were labeled as water.

Pixels from nearby nonwater areas (e.g., bare soil, vegetation) were labeled as non-water.

The dataset was balanced to reduce bias and ensure model generalization across vary- ing conditions and landscapes.

This semiautomated labeling method significantly improved the consistency and scal-ability of training data generation.

A Random Forest model was trained using the scikit learn Python library (Gillies et al., 2022). The classifier used a combination of spectral features and remote sensing indices to improve classification accuracy.

The following input features were included:

- **NDWI**: Highlights open water by contrasting green and NIR reflectance.
- **NDVI (Normalized Difference Vegetation Index)**, computed using:

$$NDWI = \frac{GREEN - NIR}{GREEN + NIR}$$

Where: NIR is Band 8A, and Red is Band 04 (665 nm) (McFeeters, 1996). NDVI was included to help differentiate water from vegetated areas.

- **Spectral Bands**:
- B03 (Green)
- B04 (Red)
- B08A (Narrow NIR)
- B11 (SWIR1, 1610 nm)
- B12 (SWIR2, 2190 nm)

These bands are particularly effective for distinguishing water from built up or dry areas, as SWIR reflectance is typically low over water but high over soil and manmade surfaces.

The model was trained on labeled pixels from the HydroLAKES dataset (European Space Agency, 2024) and validated using independent regions near Lake Akkol. Evaluation was based on:

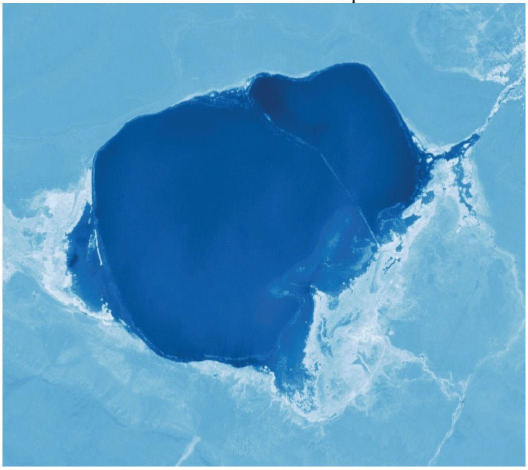
- Confusion matrix
- Overall classification accuracy
- Precision and recall for the water class

The Random Forest algorithm (Breiman, 2001) was chosen due to its ability to handle nonlinear relationships, resistance to overfitting, and strong performance in remote sensing applications.

Results and Discussion

NDWI maps generated using QGIS (Messager et al., 2016) provided a detailed view of water extent in the region. The NDWI raster was visualized using a **single band pseudocolor** rendering style, which enhances contrast between water and nonwater areas. A **blue to white color ramp** was applied, where darker blue tones represent higher NDWI values (closer to water), and white indicates drier or nonwater surfaces.

1- Figure. NDWI map of Lake Akkol generated in QGIS using single band pseudocolor with a blue to white color ramp



The computed NDWI values from the Sentinel 2 image in QGIS ranged from:

- **Minimum:**-0.63236
- **Maximum:** 0.330626

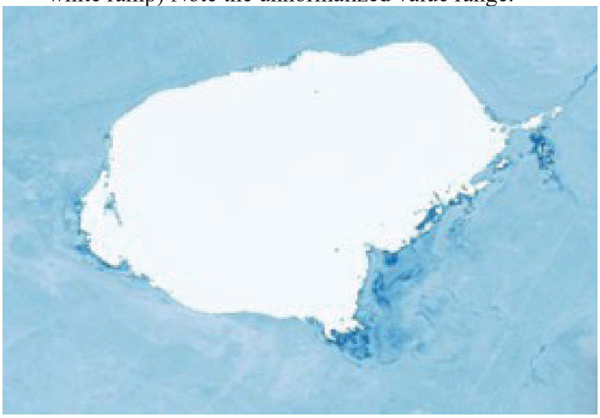
These values align with expected NDWI ranges (Straube et al., 2013) where water typically has NDWI and negative values correspond to dry soil, urban features, or vegetation. The contrast in the image helped clearly delineate Lake Akkol from surrounding land cover.

The NDWI was also calculated using a custom Python script developed and executed in the PyCharm environment. Sentinel 2 Level 2A images were processed using the rasterio library for reading raster bands and numpy for array based calculation of NDWI values (Breiman, 2001). The resulting raster was visualized using matplotlib with a **single band pseudocolor** rendering and a **blue to white color ramp**, similar to the one used in QGIS.

In this case, the raw band values were read as 16 bit integers without applying the standard reflectance scaling factor (typically 0.0001). As a result, the computed NDWI values were not normalized and fell within the range:

- **Minimum:** 0
- **Maximum:** 30.4243

2-Figure. NDWI map of Lake Akkol generated in Python using pseudocolor rendering (blue to white ramp) Note the unnormalized value range.



Although these values deviate from the expected NDWI range of −1 to +1, the structural correctness of the formula was preserved. This discrepancy primarily affects numerical interpretability rather than the visual pattern of water detection.

To obtain physically meaningful NDWI values, raw band values should be converted to surface reflectance by dividing each pixel value by 10,000. The normalized NDWI formula becomes:

$$NDWI = \frac{\frac{B3}{10000} - \frac{B8A}{10000}}{\frac{B3}{10000} + \frac{B8A}{10000}} = \frac{B3 - B8A}{B3 + B8A}$$

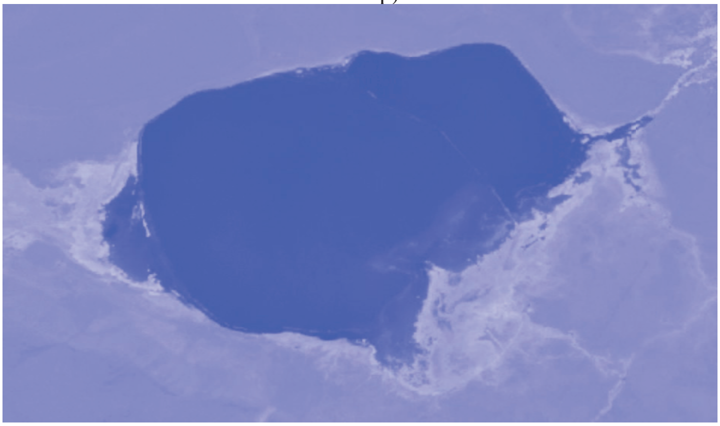
Mathematically, the scale factor cancels out, so the NDWI structure remains valid; however, normalization is crucial when comparing index thresholds or combining with other scaled indices like NDVI (McFeeters, 1996).

Despite the unnormalized range, the pseudocolor visualization remained effective. The blue to white colormap clearly emphasized water features with relatively higher NDWI values, enabling visual distinction of Lake Akkol from its surroundings.

This scenario highlights the importance of reflectance normalization when working with Sentinel 2 data in Python environments. Applying normalization ensures consistency with outputs from QGIS and Google Earth Engine, which automatically handle scaled reflectance.

Google Earth Engine (GEE) results: GEE allowed for direct access to atmospherically corrected Sentinel 2 imagery with automated scaling. NDWI was calculated using the standard formula and visualized using the same blue to white colormap. The results were clipped to the Akkol region.

3-Figure. NDWI map of Lake Akkol generated in GEE using pseudocolor rendering (blue to white ramp)



The NDWI values in GEE ranged from:

- **Minimum:** -0.6008
- **Maximum:** 0.3493

These values closely matched those observed in QGIS, confirming consistency between platforms that correctly handle reflectance.

№4 table – Classification Metrics and Their Formulas

Metric	Formula	Value
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	≈ 0.937
Precision	$\frac{TP}{TP+FP}$	≈ 0.854

Recall	$\frac{TP}{TP+FN}$	≈ 0.966
F1 score	$\frac{2 \cdot Precision \cdot Recall}{Precision+Recall}$	≈ 0.906

№5 table – Comparison of NDWI Value Ranges Across Methods

Method	Minimum NDWI	Maximum NDWI
QGIS	−0.63236	0.33063
Python (unnormalized)	0	30.4243
GEE	−0.60081	0.34934

Classification Results Using Random Forest

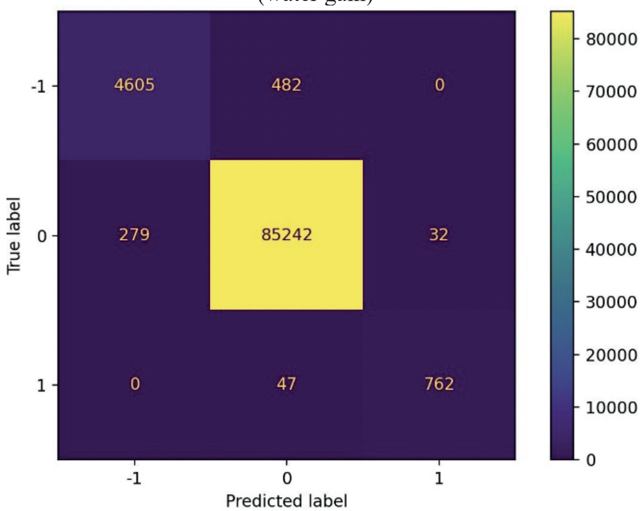
To evaluate surface water change detection in Lake Akkol, a supervised Random Forest classifier was trained using NDWI, NDVI, and spectral bands (B04, B11, B12). The model was trained using labeled water polygons from the HydroLAKES dataset (European Space Agency, 2024) focusing on three classes:

- -1 – water loss
- 0 – no change
- 1 – water gain

1. Multiclass Classification Performance

The confusion matrix in Figure 4 shows the classifier's performance on the full 3 class dataset. The matrix indicates a high number of correctly predicted stable water areas (class 0), as well as good performance in detecting both water appearance and disappearance.

Figure 4. Confusion matrix for 3 class Random Forest classification: -1 (water loss), 0 (stable), 1 (water gain)



Key values:

Class 0 (stable water): TP = 5,242

Class -1 (loss): TP = 4,605

Class 1 (gain): TP = 762

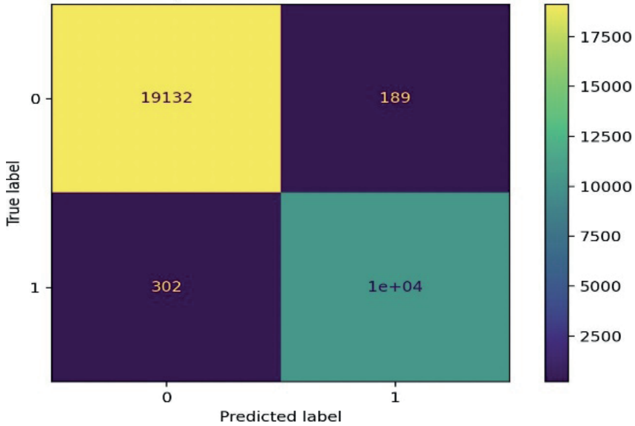
The matrix confirms that the classifier performs well across all classes, with low false positives and minimal confusion between gain and loss.

2. Binary Classification – Simplified Tasks

To compare performance under binary conditions, the model was evaluated in two simplified settings.

α) **Detecting only class 1 (water gain):**

Figure 5. Binary classification for class 1 (water gain) vs others



TP (gain correctly detected): 10,000

FP (false gain): 189

FN (missed gain): 302

β) **Detecting only class 1 (gain) with unbalanced dataset:**

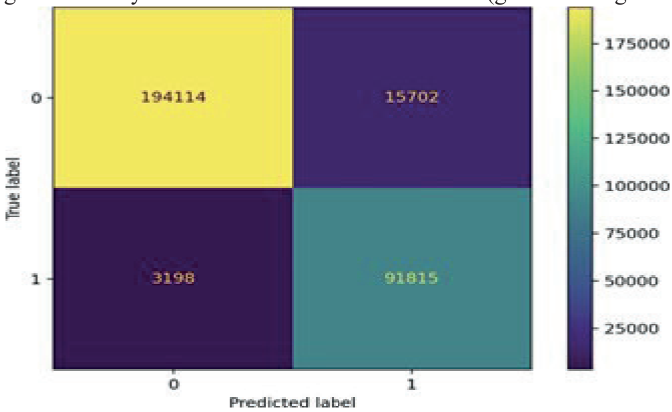
TP = 91,815

FP = 15,702

FN = 3,198

The model still maintains high recall, but precision decreases due to the high imbalance. This phenomenon is consistent with findings by Sokolova and Lapalme (Sokolova and Lapalme, 2009) who emphasize the importance of balancing precision and recall when evaluating classifiers under skewed distributions.

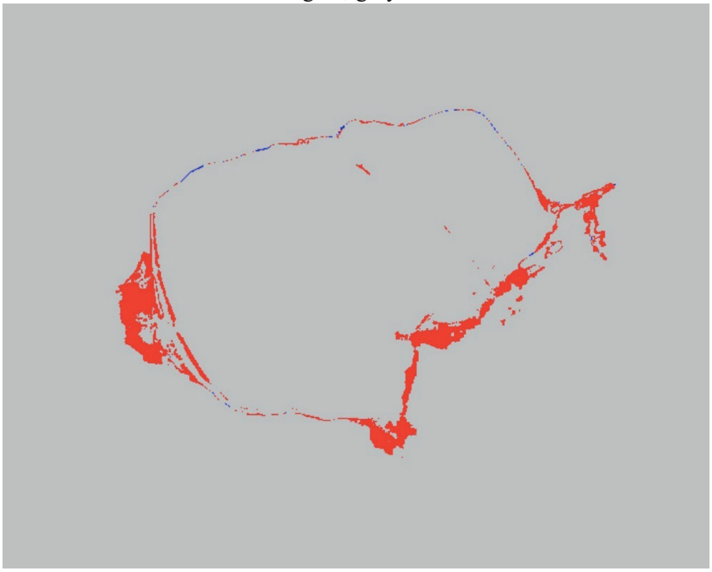
Figure 6: Binary classification with class imbalance (gain vs background)



3. Visualizing Water Change Classification

In addition to numeric metrics, the classified image was visualized as shown in Figure 7. Water changes around Lake Akkol are clearly captured, particularly shoreline expansion and contraction.

Figure 7. Classified water change map 2016–2024: red = water loss, blue = water gain, grey = stable areas



Conclusion

This paper shows how well NDWI combined with machine learning monitors water bodies. Using multiple platforms (QGIS, Python, and Google Earth Engine) offers flexibility, while the Random Forest classifier improves classification reliability. With their insights on regional water dynamics, Lake Akkol and its tributaries provide a valuable test ground for such studies. Future work may include precipitation integration, time series analysis, and extension to include wetland categorization and water quality indicators.

References

Breiman L. (2001) Random forests. Machine Learning, 45(1). — P.5–32. (in Eng.)

European Space Agency (ESA). (2024). Copernicus Data Space Ecosystem. Available: *https://dataspace.copernicus.eu (in Eng.)*

Gao B.C. (1996) NDWI – A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment, 58(3). — P. 257–266. (in Eng.)

Gorelick N., Hancher M., Dixon M., Ilyushchenko S., Thau D., Moore R. (2017) Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, 202. — P.18–27. Platform available:*https://earthengine.google.com (in Eng.)*

Gillies S. et al. (2022) Rasterio: geospatial raster I/O for Python programmers. Available: *https://rasterio.readthedocs.io (in Eng.)*

HydroLAKES database: https://www.hydrosheds.org/products/hydrolakes (in Eng.)

Kozykeyeva A.T., Mustafayev Z.S., Aldiyarova A.E., Arystanova A.B., Mosiej J. Methodical support of integrated management of water resources of the basin of transboundary rivers (2020) News of the National Academy of Sciences of the Republic of Kazakhstan, Series of Geology and Technical Sciences, 4 (442) — P.52 – 61. DOI: 10.32014/2020.2518-170X.84 (in Eng.)

McFeeters S.K. (1996) The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. International Journal of Remote Sensing, 17(7). — P.1425–1432. (in Eng.)

Messager M.L., Lehner B., Grill G., Nedeva I., Schmitt O. (2016) Estimating the volume and age of water stored in global lakes using a geo-statistical approach. Nature Communications, 7. — P.1-11. *https://doi.org/10.1038/ncomms13603* (in Eng.)

Maxwell A. E., Warner T. A., Fang F. (2018) Implementation of machine learning classification in remote sensing: An applied review. International Journal of Remote Sensing, 39(9). — P.2784–2817. (in Eng.)

OpenStreetMap contributors (2024) Planet Dump. Retrieved from *https://planet.openstreetmap.org* (in Eng.)

Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., ... Duchesnay E. (2011) Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12..— P.2825–2830. (in Eng.)

QGIS Development Team (2024) QGIS Geographic Information System. Open Source Geospatial Foundation Project. Available: *https://qgis.org (in Eng.)*

Straube F., Handfield R., Pfohl H.S. & Wieland A. (2013) Trends and strategies in logistic and supply chain management. Hamburg, Germany: Deutscher Verkehrs-erlag. Books. — P.1-16. (in Eng.)

Sokolova M., Lapalme G. (2009) A systematic analysis of performance measures for classification tasks. Information Processing & Management, 45(4). — P. 427–437. U.S.Geological Survey (USGS). (2024). EarthExplorer Data Portal. Available: *https://earthexplorer.usgs.gov (in Eng.)*

Publication Ethics and Publication Malpractice in the journals of the Central Asian Academic Research Center LLP

For information on Ethics in publishing and Ethical guidelines for journal publication see http://www.elsevier.com/publishingethics and http://www.elsevier.com/journal-authors/ethics.

Submission of an article to the journals of the Central Asian Academic Research Center LLP implies that the described work has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see http://www.elsevier.com/postingpolicy), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. In particular, translations into English of papers already published in another language are not accepted.

No other forms of scientific misconduct are allowed, such as plagiarism, falsification, fraudulent data, incorrect interpretation of other works, incorrect citations, etc. The Central Asian Academic Research Center LLP follows the Code of Conduct of the Committee on Publication Ethics (COPE), and follows the COPE Flowcharts for Resolving Cases of Suspected Misconduct (http://publicationethics.org/files/u2/New_Code.pdf). To verify originality, your article may be checked by the Cross Check originality detection service http://www.elsevier.com/editors/plagdetect.

The authors are obliged to participate in peer review process and be ready to provide corrections, clarifications, retractions and apologies when needed. All authors of a paper should have significantly contributed to the research.

The reviewers should provide objective judgments and should point out relevant published works which are not yet cited. Reviewed articles should be treated confidentially. The reviewers will be chosen in such a way that there is no conflict of interests with respect to the research, the authors and/or the research funders.

The editors have complete responsibility and authority to reject or accept a paper, and they will only accept a paper when reasonably certain. They will preserve anonymity of reviewers and promote publication of corrections, clarifications, retractions and apologies when needed. The acceptance of a paper automatically implies the copyright transfer to the Central Asian Academic Research Center LLP.

The Editorial Board of the Central Asian Academic Research Center LLP will monitor and safeguard publishing ethics.

Правила оформления статьи для публикации в журнале смотреть на сайтах:
www:nauka-nanrk.kz
http://physics-mathematics.kz/index.php/en/archive

ISSN2518-1726 (Online),
ISSN 1991-346X (Print)

Директор отдела издания научных журналов НАН РК *А. Ботанқызы*
Редакторы: *Д.С. Аленов, Ж.Ш.Әден*
Верстка на компьютере *Г.Д.Жадыранова*

Подписано в печать 25.09.2025.
Формат 60x881/8. Бумага офсетная.
Печать –ризограф. 20,0 п.л. Заказ 3.

Национальная академия наук РК
050000, Алматы, ул. Кунаева 132