ABSTRACT

For the dissertation titled «Assessment of soil degradation degree in Southern Kazakhstan based on remote sensing data and machine learning methods» submitted by Aisulu Ataniyazova for the degree of Doctor of Philosophy (PhD) in the educational program 8D06102 – «Computer Engineering»

Relevance. Southern Kazakhstan is one of the regions where the bulk of the country's agricultural potential is concentrated. However, in recent decades, soil degradation has been intensifying due to irrigated agriculture and extreme climatic conditions. Problems such as salinization, erosion, and organic matter depletion reduce agricultural productivity, disrupt ecosystem balance, and threaten food security. Therefore, rapid and accurate assessment of soil degradation is essential for effective natural resource management and land reclamation planning. Traditional methods of soil monitoring, including field surveys, geobotanical assessments, and laboratory analyses of soil samples, remain scientifically robust but are labor-intensive, time-consuming, and financially demanding. Their spatial coverage is limited, making it difficult to obtain a comprehensive regional picture of degradation dynamics. In large and heterogeneous territories such as southern Kazakhstan, reliance solely on ground-based observations does not allow for оперативное and systematic monitoring of degradation processes at multiple spatial and temporal scales.

Remote sensing technologies enable a wide range of systematic and comprehensive detection of changes in the physicochemical properties of soils. Multispectral satellite and radar data provide highly accurate monitoring of salinization hotspots, moisture dynamics, and vegetation conditions. Furthermore, the widespread use of machine learning algorithms enables the development of modern, reliable methods for predicting and mapping soil degradation, as well as identifying hidden patterns in large volumes of remote sensing data.

In the case of Kazakhstan, research in this area is limited and often limited to the regional level. Therefore, a quantitative assessment of soil degradation in the southern regions of Kazakhstan using remote sensing data and machine learning methods is of scientific and practical importance. Such research will contribute to improving the condition of irrigated lands, ensuring sustainable agricultural development, and refining national land use strategies.

Moreover, in his Address to the People's Republic of Kazakhstan by 2025, the President of the Republic of Kazakhstan emphasized the need to manage agricultural land using satellite monitoring and artificial intelligence technologies and instructed the creation of a unified digital map of land resources. This confirms that this research topic is a priority at the national level. In this regard, assessing the extent of land degradation in southern Kazakhstan using remote sensing data and advanced machine learning methods is a highly relevant study that will contribute to ensuring the efficient use of land resources, scientifically

based planning of measures to prevent degradation, and the digital transformation of the industry.

The aim of this research is to develop a method for assessing and mapping the salinity of agricultural soils in southern Kazakhstan by integrating field measurements with remote sensing data.

Research objectives:

1. Analysis of initial data and selection of study sites considering natural, climatic, and economic conditions.
2. Collection and pre-processing of remote sensing data: acquisition of Sentinel-1 radar satellite images (VH and VV polarization) and Sentinel-2 optical images for selected sites.
3. Development and training of machine learning models for soil salinity assessment.
4. Evaluation of model accuracy and comparative analysis of their effectiveness.
5. Compilation of soil salinity maps based on the analysis of spatiotemporal patterns using geographic information systems (GIS) to assess salinity degree, combining spectral indices and machine learning results.

Research methods:

Standard statistical methods for correlation analysis of remotely sensed parameters (NDSI, SI, etc.) and ground-based information, as well as methods for classification and regression of satellite data.

Research subject:

Textural features and spectral indices extracted from remote sensing data, machine learning algorithms for numerical and spatial modeling of soil degradation.

Research object:

Soil cover of the southern regions of Kazakhstan and its degradation processes.

Scientific novelty of the results obtained:

A method for assessing the degree of soil salinity based on geophysical measurements of the Earth's surface and remote sensing data of medium and high spatial resolution has been developed.

Thesis submitted for defense:

A method for numerically predicting the degree of soil salinity (electrical conductivity) with a resolution of ~10 m has been developed based on Sentinel-1 radar data, Sentinel-2 optical data, and machine learning algorithms calibrated using ground-based measurements. The combination of radar, optical, relief and temperature data increased the accuracy of mapping saline areas by approximately 15% compared to traditional spectral indices.

Practical significance of the study: The developed method is based on open satellite data and is suitable for direct use in agro-reclamation and environmental monitoring systems in Kazakhstan. Salinization maps obtained using the integrated satellite data method enable accurate assessment of the

melioration status of agricultural lands and the timely identification of salinized and degraded areas. This approach facilitates decision-making aimed at effectively planning irrigation and drainage systems, increasing soil productivity, and rationally using natural resources.

Scientific and technical significance of the study:

The study validated a method for integrating remote sensing data for predicting the degree of soil salinization with high spatial accuracy. Combining radar (Sentinel-1) and optical (Sentinel-2) data with terrain and temperature parameters increased the possibility of a comprehensive assessment of salinization processes. The informative value of radar microtextural characteristics (contrast, entropy) based on GLCM was demonstrated, and their reliable relationship with electrical conductivity indicators was determined. The comparative performance of linear and ensemble machine learning methods was investigated, and it was found that even simple models (Ridge) can perform well under conditions of limited data volume.

Reliability of scientific results:

The reliability and validity of the obtained results are supported by the fact that the study is based on a combination of field measurements and independent satellite data, as well as statistically reproducible model validation procedures. This dissertation describes the collection of 203 soil samples at the main test sites. Electrical conductivity was determined in the laboratory using a digital meter and standardized sample preparation (soil:water ratio of 1:5, settling for 24 hours).

The reliability of the methodological approach is enhanced by the openness of the satellite data processing chain: the use of Sentinel-1 SAR data (IW mode, GRD products), standard preprocessing steps (noise removal, radiometric calibration, elevation correction, mapping to a map projection), and the creation of a feature set including electrical conductivity, 16 GLCM textural features, temperature, and elevation parameters. Surface temperature data (MODIS LST) and a digital elevation model were used to expand the set of covariates.

The statistical reliability of the modeling results is confirmed by the multiple cross-validation procedure: the dissertation used a randomized ShuffleSplit cross-validation with an 80/20 split and 200 repetitions for each regressor.

The combined dataset (for all) shows R^2 values of up to 0.886 for Ridge and 0.859 for LightGBM, with correspondingly decreasing RMSE and MAPE. Similarly, for individual regions, Ridge/LightGBM show good values (e.g., for Kunaev, R^2 values of up to 0.894 for Ridge and 0.864 for LightGBM). These results confirm the validity of the proposed "heterogeneous data + regularization/gradient boosting" approach with a limited field observation base.

With regard to interpretation and validation results, the dissertation separately proposes and utilizes an approach to feature proportion estimation based on Shapley additive annotations (SHAP), which reduces the risk of "black-box" modeling and improves the reliability of feature selection and model architecture.

Relation of the topic to research programs:

This dissertation was completed at the Institute of Information and Computational Technologies as part of the 2021–2023 PTF program, state registration No. 0121PK00830, «Space monitoring and GIS for quantitative assessment of soil salinity and agricultural land degradation in Southern Kazakhstan».

Dissertation structure and contents:

The dissertation consists of an introduction, three main sections, a conclusion, and a bibliography. The total length of the work is 104 pages, including 43 figures, 23 tables, and 40 formulas. The bibliography includes 120 titles.

The first section is devoted to a review of domestic and international research on the use of radar and optical data in assessing soil salinity, as well as the application of machine learning based on this data.

The second section describes the assessment of soil salinity in irrigated areas of southern Kazakhstan using heterogeneous data, which is the main result of the dissertation.

The third section is devoted to the analysis of spatiotemporal patterns of salinization and vegetation status in the study areas based on optical data.

Publications:

In international publications indexed in the Scopus database:

1. Amirgaliyev Y., Mukhamediev R., Merembayev T., Kuchin Y., Ataniyazova A. Remote sensing and machine learning algorithms to predict soil salinity in southern Kazakhstan // Discover Sustainability. – 2024. – Vol. 5. – Article 363. – DOI: 10.1007/s43621-024-00594-8.

2. Omarova P., Amirgaliyev Y., Kozbakova A., Ataniyazova A. Application of physics-informed neural networks to river silting simulation // Applied Sciences. – 2023. – Vol. 13. – Article 11983. – DOI: 10.3390/app132111983.

3. Omarova P., Yang T., Ataniyazova A., Kozbakova A., Yedilkhan D. Assessment of air quality in the urban environment taking into account vegetation and building geometry // Journal of Ecological Engineering. – 2026. – Vol. 27, No. 2. – DOI: 10.12911/22998993/211688.

In publications recommended by the Committee for Quality Assurance in Science and Higher Education of the Ministry of Science and Higher Education of the Republic of Kazakhstan:

1. Amirgaliyev Y., Ataniyazova A., Merembayev T., Cherikbaeva L., Kassymova D. Mapping soil salinity in southern Kazakhstan based on Sentinel-2 satellite images // KazATC Bulletin. – 2024. – Vol. 132, № 3. – P. 194–205. – DOI: 10.52167/1609-1817-2024-132-3-194-205.

2. Amirgaliyev Y., Amirgaliyeva S., Ataniyazova A., Narynov S., Berdibayev R. Reinforcement learning as a technology enabling continuous monitoring of vast territories // KazATC Bulletin. – 2024. – Vol. 135, № 6. – P. 255–262. – DOI: 10.52167/1609-1817-2024-135-6-255-262.

3. Ataniyazova A., Merembayev T. Spatiotemporal assessment of soil salinity in irrigated agricultural lands of Kazakhstan using remote sensing // Journal of Problems in Computer Science and Information Technologies. – 2025. – Vol. 3, No. 3. – DOI: 10.26577/jpcsit2025332.

Also, based on the results of the research work, Certificate No. 60555 was registered on the entry of information into the State Register of Rights to Objects Protected by Copyright of the Republic of Kazakhstan, on the topic «Quantitative assessment of the degree of soil degradation using machine learning and remote sensing of the Earth», issued on July 4, 2025.