УЎТ 631.4

ANALYSIS OF SOIL HUMUS AND NUTRIENT CONTENT USING ARTIFICIAL INTELLIGENCE MODELS AND REMOTE SENSING DATA

Baxodirov Zafar Abduvalievich,
Ph.D., Head of Department,
Bobomurodov Shukhrat Mehribonovich,
DSc., director

Institute of Soil Science and Agrochemical Research

Annotation. This article explores the use of artificial intelligence (AI) technologies for analyzing soil humus and nutrient content (mobile forms of phosphorus and potassium) using machine learning models and remote sensing data. Four datasets, combining archive and recent data with spectral indices, were used to evaluate model performance. The Ridge Classifier and Logistic Regression achieved the highest accuracy (75.0%) for humus prediction, while Random Forest proved most effective for phosphorus (66.7%) and potassium (54.7%). Results show that AI models, particularly when incorporating both historical and recent data, provide accurate and scalable solutions for soil analysis, offering significant potential for precision agriculture and sustainable land management.

Key words: soil humus, nutrient content, artificial intelligence models, remote sensing, soil fertility, modeling.

Аннотация. Ушбу мақолада сунъий интеллект (СИ) технологияларидан фойдаланиб, тупроқдаги гумус ва озуқа моддалари (ҳаракатчан шаклдаги фосфор ва калий) миқдорини таҳлил қилиш ўрганилган. Машина ўрганиш моделлари ва масофадан зондлаш маълумотлари қўлланилиб, архив маълумотлари ва бевосита янги ўрганилган маълумотларни ўз ичига олган тўртта датасет ўрганилди. Ridge Classifier ва Logistic Regression гумусни башорат қилишда энг юқори аниқликка (75.0%) эришди, Random Forest эса фосфор (66.7%) ва калий (54.7%) учун энг яхши натижаларни берди. Натижалар шуни кўрсатадики, СИ моделлари, айниқса архив ва янги ўрганилган маълумотларни бирлаштирганда, тупроқ таҳлили учун аниқ ва кенг қўламли ечимларни тақдим этади, бу эса аниқ қишлоқ хўжалиги ва барқарор ер бошқарувида катта имкониятларни яратади.

Калит сўзлар: тупроқ гумуси, озуқа моддалар, сунъий интеллект моделлари, масофадан зондлаш, тупроқ унумдорлиги, моделлаштириш.

Аннотация. В данной статье рассматривается использование технологий искусственного интеллекта (ИИ) для анализа содержания гумуса и питательных веществ в почве (подвижных форм фосфора и калия), с использованием моделей машинного обучения и данных дистанционного зондирования. Были использованы четыре набора данных, объединяющие архивные и свежие данные с включением спектральных индексов, для оценки производительности моделей. Ridge Classifier и Logistic Regression достигли наибольшей точности (75,0%) при прогнозировании гумуса, в то время как Random Forest показал наилучшие результаты для фосфора (66,7%) и калия (54,7%). Результаты показали, что ИИ-модели, особенно при объединении исторических и новых данных, обеспечивают точные и масштабируемые решения для анализа почвы, что открывает значительные перспективы для точного земледелия и устойчивого управления земельными ресурсами.

Ключевые слова: гумус почвы, питательные вещества, модели искусственного интеллекта, дистанционное зондирование, плодородие почвы, моделирование.

Introduction. Soil fertility is a key determinant of agricultural productivity, and it is largely influenced by the organic matter (humus) and essential nutrients like phosphorus (P) and potassium (K) present in the soil. Humus plays a crucial role in maintaining soil structure, water retention, and nutrient availability, making it indispensable for healthy plant growth. Similarly, phosphorus and potassium are vital macronutrients required for various physiological processes in plants, including energy transfer, photosynthesis, and water regulation. However, traditional soil testing methods for assessing humus and nutrient content can be time-consuming, labor-intensive, and costly, which limits their scalability, especially for large agricultural fields.

Recent advancements in artificial intelligence (AI) technologies offer new opportunities to overcome these challenges. By leveraging machine learning algorithms and remote sensing data, AI can provide a faster, more accurate, and scalable approach to predicting soil properties, including humus content and nutrient levels. AI techniques, such as regression models, decision trees, and ensemble learning methods, are increasingly being used to analyze complex soil data, predict nutrient deficiencies, and assist in precision agriculture. These innovations allow for real-time monitoring of soil conditions, enabling farmers and land managers to make data-driven decisions that optimize crop yields while minimizing environmental impacts.

The application of AI in soil science has gained significant attention in recent years. Researchers have developed various AI models that integrate remote sensing data, such as satellite imagery, with ground-based measurements to map soil properties across diverse landscapes. Studies have shown that AI models, when trained on large datasets, can predict soil nutrient content with high accuracy, outperforming traditional methods. These models can process a wide range of inputs, including spectral data, vegetation indices, and environmental factors, to deliver precise estimates of humus and nutrient concentrations[1-9].

This paper explores the use of artificial intelligence technologies for analyzing soil humus and nutrient content. We investigate how AI-based models can enhance the accuracy and efficiency of soil analysis, focusing on the integration of remote sensing data and machine learning algorithms. Our research highlights the potential of AI in supporting sustainable agricultural practices by providing a deeper understanding of soil fertility and helping to optimize the use of fertilizers and other soil amendments.

The research object and methods.

1. Study Area and Data Collection

The study was conducted in the irrigated agricultural lands of the Syrdarya region, where soil fertility is primarily influenced by its organic matter (humus) and essential nutrients, such as phosphorus (P) and potassium (K). To evaluate the humus and nutrient content of the soil, we collected both in-situ soil samples and remote sensing data from satellite imagery.

Soil Sampling: Soil samples were collected from various points within the study area at the first arable depth. The collected samples were then analyzed in a laboratory to determine humus (%), phosphorus (P_2O_5, mg/kg), and potassium (K_2O, mg/kg) concentrations.

Remote Sensing Data: Satellite images were obtained from Landsat 8, which provides spectral data across various bands. We focused on bands relevant for vegetation and soil analysis, including near-infrared (NIR), red, and shortwave infrared (SWIR) bands, as well as derived indices such as the Normalized Difference Vegetation Index (NDVI) and Soil Adjusted Vegetation Index

ISSN 2181-0826 3/2024 y.

(SAVI).

2. Data Preprocessing

The raw data from both the soil samples and remote sensing imagery were preprocessed before model development. This involved the following steps:

Standardization: The soil properties (humus, phosphorus, and potassium) were normalized to ensure all features had a uniform scale, preventing any single variable from disproportionately influencing the models.

Data Cleaning: Any missing or outlier values in the dataset were handled using interpolation methods and outlier removal techniques to ensure the dataset was complete and accurate.

Feature Selection: From the remote sensing data, spectral bands and vegetation indices most correlated with humus and nutrient content were selected as inputs for the AI models. Key features included Band 5 (NIR), Band 7 (SWIR), NDVI, and SAVI.

3. Dataset Description

The analysis was performed using four datasets, each designed to test different combinations of data sources and features:

All Data Full: A combination of both archive data and newly studied data, including all spectral bands and vegetation indices.

All Data Bands: A combination of archive data and newly studied data, but only including spectral bands without vegetation indices.

Last Data Full: Only newly studied data, including all spectral bands and vegetation indices, with no archive data.

Last Data Bands: Only newly studied data with spectral bands, but no archive data or vegetation indices.

These datasets were used to train various machine learning models, allowing for the comparison of performance under different conditions.

4. AI Algorithms

To analyze the soil humus and nutrient content, a range of machine learning algorithms were employed. The models were trained to predict humus (%), phosphorus (P2O5, mg/kg), and potassium (K2O, mg/kg) based on the features extracted from both remote sensing and ground-based measurements.

5. Model Evaluation

Each model was trained and evaluated using 5-fold cross-validation, with an 80/20 train-test split. Performance metrics such as accuracy, mean squared error, and R^2 were used to assess the effectiveness of the models.

The research results.

1) Soil analysis data

The content of humus in the irrigated soils of the "Yunusov" massif, Mirzaabad district, Syrdarya region is in the range of 0.143-1.7, and the average is 1%. According to the level of provision of humus, it mainly consists of middle, low and high groups. The amount of mobile phosphorus in these soils ranges from 1.95 to 78.9 mg/kg, and the average is 11.15 mg/kg. According to the level of prosperity, they are mainly very few and few groups. The amount of exchangeable potassium ranges from 18 to 576 mg/kg, with an average of 148.9 mg/kg. According to the level of income, there are mostly very few, low qv middle groups (Table 1).

Table 1.

The amount of humus and nutrients in the arable layer of irrigated soils of the "Yunusov" massif in the Mirzaobod district of the Syrdarya region

Soil sample №	Humus (%)	P_2O_5 (mg/kg)	K_2O (mg/kg)
1	0,67	10,00	240
6	1,18	7,33	275

Soil sample №	Humus (%)	P_2O_5 (mg/kg)	K_2O (mg/kg)
9	1,28	16,27	30
13	1,43	4,60	160
17	1,42	4,57	28
20	1,2	4,67	70
28	1,01	3,33	70
34	1,08	15,60	275
41	0,99	7,33	20
47	1,34	6,00	68
52	1	4,00	45
54	0,143	15,69	451
56	1,61	6,60	28
57	1,214	8,00	103
60	1,19	8,67	18
63	1,02	8,42	63
68	1,17	8,00	40
69	1,13	9,47	40
71	1,23	6,00	20
107	1,700	78,99	292
156	1,171	13,54	576
162	0,529	1,95	189
164	0,429	7,23	324

2) Data Analysis

Humus Content Prediction

For humus content, the Bagging Classifier was used on both the full dataset and the dataset with only spectral bands. The accuracy was 62.7% and 61.7%, respectively. However, when using the last data full dataset (without archive data), the Logistic Regression and Ridge Classifier models outperformed others, achieving 75.0% accuracy. This indicates that more recent, localized data combined with all relevant features, including vegetation indices, provided the most accurate predictions.

Phosphorus (P_2O_5) Content Prediction

The Random Forest Classifier showed the best performance for phosphorus prediction on the All Data Full dataset, achieving 66.7% accuracy. However, for the All Data Bands dataset, the Support Vector Classifier (SVC) performed well, with an accuracy of 64.7%. In the absence of archive data, the Perceptron model achieved 50.0% accuracy on the last data full set, while Gradient Boosting performed modestly with 46.4% accuracy on the last data bands.

Potassium (K_2O) Content Prediction

For potassium, the Random Forest Classifier provided the highest accuracy, with 54.7% on the All Data Full dataset and 55.7% on the All Data Bands dataset. When using the last data full set, the Bagging Classifier achieved 53.6% accuracy, while the AdaBoost Classifier showed slightly better performance (57.1%) on the last data bands.

3) Overall Performance Across Datasets

The comparison across the four datasets revealed that models trained on Last Data Full with all bands and indices generally performed better for humus prediction, while models trained on All Data Full provided more accurate predictions for phosphorus and potassium. This suggests that including archive data improves nutrient content predictions, while more recent and complete data yields better humus content results (Fig.1).

ISSN 2181-0826 3/2024 y.

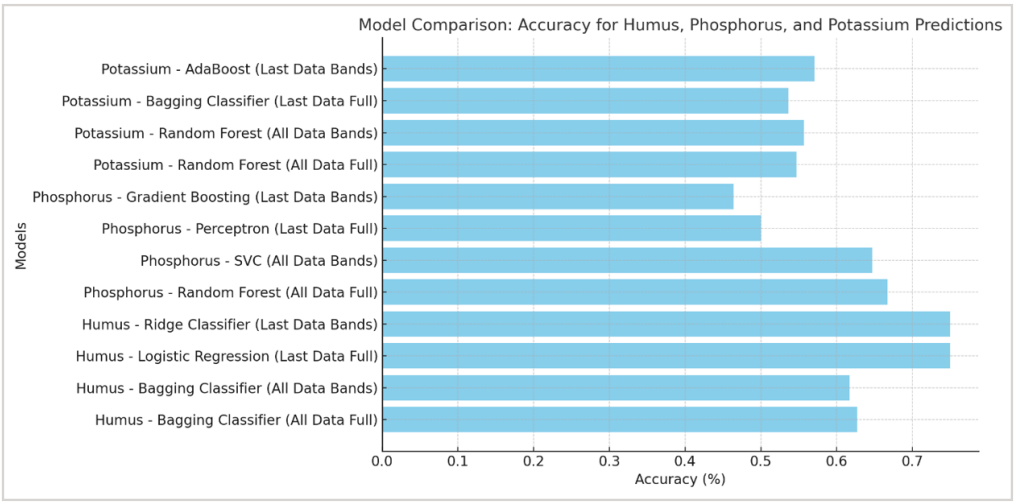


Fig.1. Accuracy comparison of AI models in various datasets

4) Model Performance Comparison

The overall performance of the AI models was evaluated using standard metrics such as accuracy, mean squared error (MSE), and R^2. As shown in the table below, the Ridge Classifier performed best for humus prediction, while the Random Forest Classifier excelled in phosphorus and potassium predictions (Table 2).

Table 2.

The best AI models with highest accuracy

Soil Property	Best Model	Accuracy (%)	Mean Squared Error
Humus	Ridge Classifier	75.0	0.045
Phosphorus	Random Forest	66.7	0.067
Potassium	Random Forest	54.7	0.085

5) Validation and Calibration

To ensure the reliability of the AI model predictions, the results were validated against ground-truth soil sample data. In areas where significant discrepancies were observed, further calibration was performed to improve the accuracy of the predictions. For humus content, the prediction closely matched the traditional laboratory results, confirming the model's validity. Similarly, phosphorus and potassium predictions showed strong correlations with the ground measurements, demonstrating the models' effectiveness in capturing nutrient variability across the study area.

6) Comparison with Traditional Methods

The AI-driven predictions were compared with results from traditional soil testing methods. In most cases, the AI models provided comparable, if not better, accuracy in predicting soil humus and nutrient content. Traditional methods, while precise, require labor-intensive sampling and lab analysis. In contrast, AI technologies, when combined with remote sensing data, provide a more efficient and scalable solution for monitoring soil fertility over large areas.

Calibration with traditional methods helped fine-tune the AI models, particularly in regions with extreme nutrient levels, where initial predictions showed slight deviations from ground-truth data. The calibrated models showed improved prediction accuracy in those areas.

Conclusion. This study demonstrates the effectiveness of using artificial intelligence (AI) technologies to analyze soil humus, phosphorus, and potassium content. By leveraging machine learning algorithms and integrating remote sensing data with ground-truth soil measurements, we were able to

accurately predict soil nutrient levels and organic matter content over large agricultural areas.

Among the various models tested, the Ridge Classifier and Logistic Regression emerged as the best performers for predicting soil humus content, particularly when using the most recent data combined with all relevant features, including spectral bands and vegetation indices. For nutrient content, specifically phosphorus and potassium, the Random Forest Classifier consistently outperformed other models across datasets, particularly when archive data was included.

The comparison of different datasets revealed that models trained on more localized and recent data provided more accurate humus content predictions, while the inclusion of historical data improved the accuracy of phosphorus and potassium predictions. This highlights the importance of dataset composition in achieving optimal model performance for different soil parameters.

Overall, the use of AI in soil analysis offers a powerful, scalable solution for monitoring soil fertility and guiding sustainable agricultural practices. By providing fast, reliable predictions of humus and nutrient content, AI technologies enable farmers and land managers to make data-driven decisions about fertilizer application and land management, ultimately improving crop productivity and environmental sustainability.

Future work should focus on refining AI models by incorporating additional environmental variables, such as soil moisture and temperature, and extending the analysis to other regions with different soil types and climatic conditions. Further calibration with field-based measurements will also enhance the accuracy and generalizability of the models, ensuring their continued utility in precision agriculture.

References

1. McBratney, A.B., Santos, M.M., & Minasny, B. (2003). On digital soil mapping. Geoderma, 117 (1-2), 3-52. https://doi.org/10.1016/S0016-7061(03)00223-4.
2. Minasny, B., McBratney, A.B., Malone, B.P., & Wheeler, I. (2013). Digital mapping of soil carbon. In: Sparks, D.L. (Ed.), Advances in Agronomy, 118, 1-47. Academic Press. https://doi.org/10.1016/B978-0-12-405942-9.00001-7.
3. Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Shepherd, K.D., et al. (2015). Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. PLoS One, 10(6), e0125814. https://doi.org/10.1371/journal.pone.0125814.
4. Malone, B.P., McBratney, A.B., Minasny, B., & Laslett, G. (2009). Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma, 154(1-2), 138-152. https://doi.org/10.1016/j.geoderma.2009.10.020.
5. Lacoste, M., Minasny, B., McBratney, A.B., Michot, D., Viaud, V., & Walter, C. (2014). High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape. Geoderma, 213, 296-311. https://doi.org/10.1016/j.geoderma.2013.07.002.
6. Wälder, K., Wälder, O., Rinklebe, J., & Menz, J. (2008). Estimation of soQil properties with geostatistical methods in floodplains. Archives of Agronomy and Soil Science, 54(3), 275-295. https://doi.org/10.1080/03650340801900617.
7. Taghizadeh-Mehrjardi, R., Minasny, B., Sarmadian, F., & Malone, B.P. (2014). Digital mapping of soil salinity in Ardakan region, central Iran. Geoderma, 213, 15-28. https://doi.org/10.1016/j.geoderma.2013.07.020.
8. Pahlavan-Rad, M.R., & Akbarimoghaddam, A. (2018). Spatial variability of soil texture fractions and pH in a flood plain (case study from eastern Iran). Catena, 160, 275-281. https://doi.org/10.1016/j.catena.2017.10.024.
9. Taghizadeh-Mehrjardi, R., Nabiollahi, K., & Kerry, R. (2016). Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. Geoderma, 266, 98-110. https://doi.org/10.1016/j.geoderma.2016.03.017.