Информация об авторах

Даниял Каратаев* – магистрант кафедры информационной безопасности; Евразийский национальный университет им. Л.Н. Гумилева, Астана, Казахстан; e-mail: dkarataev2000@gmail.com. ORCID: https://orcid.org/0009-0000-9109-533X.

Баян Абдураимова – научный руководитель, кандидат технических наук, доцент, Евразийский национальный университет им. Л.Н. Гумилева, Казахстан, Астана.

Авторлар туралы ақпарат

Даниял Қаратаев*– ақпараттық қауіпсіздік кафедрасының магистранты; Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан; e-mail: dkarataev2000@gmail.com. ORCID: https://orcid.org/0009-0000-9109-533X.

Баян Абдурайымова – ғылыми жетекші, техника ғылымдарының кандидаты, доцент, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Қазақстан, Астана.

Received 30.10.2024
Revised 05.11.2024
Accepted 07.11.2024

https://doi.org/10.53360/2788-7995-2024-4(16)-10

IRSTI: 36.16.35; 34.03.47; 34.23.22

D.V. Son
Astana IT University,
010000, Republic of Kazakhstan, Astana, Mangilik El Avenue, C1
e-mail: qwerty.01.qwerty0op@gmail.com

SOIL YIELD FORECASTING

Abstract: *This research project serves as a comprehensive meta-analysis in the field of agricultural science, specifically focusing on the prediction of crop yields. This endeavor involves collating and synthesizing findings from a variety of studies and articles that have explored different methodologies and models for forecasting agricultural outputs. The objective of this comprehensive review is to identify trends, methodologies, and key factors that consistently influence crop yield predictions across different studies.*

It synthesizes methodologies from various studies, emphasizing machine learning (ML) techniques like Support Vector Machines (SVM), Random Forest (RF), and Convolutional Neural Networks (CNN). These studies integrate high-resolution satellite imagery with environmental indices such as NDVI, EVI, and LAI. Soil chemical properties (pH, nutrients) and satellite-derived data were used to enhance the prediction of crop yields for diverse crops. The findings highlight the comparative effectiveness of different models in handling the spatial and temporal variability of both above-ground and below-ground data, improving prediction accuracy under varying environmental and soil conditions.

Through this theoretical analysis, the research underscores the potential of advanced analytical models to transform agricultural monitoring and prediction, providing critical insights that can aid in the optimization of agricultural policies and resource management.

Key words: *Crop Yield Prediction, Satellite Imagery, Machine Learning, Convolutional Neural Networks (CNN), Vegetation Indices, Soil Chemical Properties.*

Introduction

Accurate crop yield prediction is essential for enhancing agricultural productivity and ensuring food security. Advances in technology have significantly improved these predictions, helping manage environmental risks and optimize resources.

Traditional methods of yield prediction, relying on historical data and simple empirical models, often struggle to capture the complexity of modern agricultural ecosystems. Factors like unpredictable weather, soil variability, and crop management practices add challenges that these models cannot address. The integration of high-resolution satellite imagery and vegetation indices like NDVI and EVI has transformed crop monitoring, offering detailed insights into plant health and biomass [1, 2].

Additionally, machine learning techniques – such as Random Forests, Support Vector Machines, and deep learning models like Convolutional Neural Networks (CNNs) – have significantly enhanced prediction accuracy by analyzing large, complex datasets [3-6]. These models leverage both spatial and temporal data, revealing patterns that traditional approaches often miss.

While remote sensing and machine learning have improved above-ground crop monitoring, the integration of soil chemical properties – such as pH, nutrient content, and organic matter – into predictive models has proven essential for better accuracy. Soil health directly affects plant growth, and models that combine satellite data with detailed soil profiles offer more precise yield predictions [7]. This comprehensive approach not only enhances short-term accuracy but also aids in the sustainable management of soil resources for long-term agricultural productivity.

The fusion of advanced computational techniques and detailed environmental data marks a pivotal shift in agricultural forecasting, paving the way for more reliable and sustainable crop yield predictions.

Literature Review

In recent years, machine learning methods have become essential tools in agriculture, especially for predicting crop yields. Various studies have introduced approaches that leverage satellite imagery, soil chemical properties, and deep learning models, but each method faces limitations related to data types and computational resources.

One of the most commonly used approaches involves satellite imagery and deep learning, such as Convolutional Neural Networks (CNN). For instance, in a study utilizing CNN, Landsat 8 satellite data was employed to predict yields for cabbage and radish [6]. This approach performs well with spatial data, allowing the model to incorporate vegetation indices like NDVI and GDVI. However, a reliance on surface data alone leads to the omission of critical below-ground factors influencing crop yields. Soil chemical properties, such as nutrient levels and acidity, are essential in yield prediction but were not included in the model, causing inaccuracies, particularly in situations where soil composition plays a significant role [6].

To address these gaps, some researchers incorporated soil chemical properties into their models. For example, one study used a OLS regression model to analyze the impact of soil properties on rice yield [1]. However, while OLS regression can effectively reveal relationships between soil properties and yield, its limitation lies in capturing the nonlinear dependencies and interactions that often exist among environmental variables. This restricts the model’s applicability to more complex, dynamic environments where soil, weather, and plant responses interact in nonlinear ways.

To enhance these predictive models, recent studies have turned to more advanced machine learning algorithms, such as Random Forest (RF) and Support Vector Machines (SVM), which handle nonlinear relationships better than OLS regression. Furthermore, machine learning models like RF, ERT, DL, SVM and CNN require substantial computational power and high-quality data, which can be restrictive in resource-constrained settings. Moreover, there is a problem with overfitting and also certain problems with data dependence.

Each approach thus has unique strengths and limitations: CNNs effectively process satellite images but require significant resources and overlook underground factors; OLS regression yield accurate results but face scalability limitations; and RF, ERT, DL, SVM achieves high clustering accuracy but demands labor-intensive parameter tuning and avoiding overfitting. This research aims to integrate the strengths of each method to create a hybrid model, addressing their respective limitations and enabling a more comprehensive model that considers essential factors and overcomes data and resource constraints.

Data and Method

Methods

Data Collection and Preparation

The data used in this study encompassed satellite imagery, soil chemical properties, and additional climate and topsoil information. Satellite images were sourced from high-resolution Landsat 8, focusing on vegetation indices such as NDVI (Normalized Difference Vegetation Index), EVI (Enhanced Vegetation Index), and LAI (Leaf Area Index), which indicate plant health, density, and overall crop condition. Soil chemical data, including organic matter, phosphorus (P), potassium (K), calcium (Ca), magnesium (Mg), and silicon dioxide (SiO_2), were gathered to assess the influence of soil composition on yield predictions over several years.

Preprocessing involved standardizing satellite images to maintain consistency across spatial and temporal dimensions. Soil data were cleaned, scaled, and interpolated for compatibility with satellite data resolution, ensuring a cohesive dataset for model integration. For statistical analyses, explanatory and response variables were normalized using z-scores, facilitating direct comparison across variables of different scales.

Statistical Modeling

The Ordinary Least Squares (OLS) regression model served as a benchmark for predicting crop yields, structured to minimize residuals in the regression equation:

$y=\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_i x_i+\epsilon$ Here, y denotes the response variable, x_1 to x_i are the explanatory variables, β_0 is the intercept, β_1 to β_i represent the slopes correlating the response variable (y) with the explanatory variables (x_1 to x_i), and ϵ is the error term.

In one analysis, corn and soybean yields were modeled using fourteen variables, including vegetation indices (NDVI, EVI, LAI), climate factors (precipitation, Tmin, Tmax, Tmean), topsoil and subsoil properties (e.g., clay fraction, bulk density, pH, calcium carbonate, exchangeable sodium, and electrical conductivity), and nitrogen fertilizer inputs [1].

Correlation coefficients were calculated to assess the relationships between the variables and the response, guiding the selection of relevant explanatory variables. Redundancies were further identified using the Variation Inflation Factor (VIF) to avoid multicollinearity.

$$VIF(x_j)=\frac{1}{1-R_j^2}$$

(Picture 1 – VIF formula)

Where Rj2 is the R-squared value from regressing xj against the other explanatory variables. A VIF exceeding 10 indicates significant multicollinearity, leading to the exclusion of variables such as EVI, Tmax, Tmean, and CaCO3. The final model included ten variables: NDVI, LAI, precipitation, Tmin, clay content, bulk density, pH, exchangeable sodium percentage (ESP), electrical conductivity (ECE), and nitrogen fertilizer (NTRG).

To standardize comparisons across variables, z-scores were used to normalize both the response and explanatory variables:

$$Z=\frac{(v-\mu)}{\sigma}$$

(Picture 2 – z-scores formula)

Machine Learning Models

The process of setting up models involves a structured series of steps aimed at capturing the complexities of agricultural environments. First, it is essential to define the prediction objective, specifying the crop type, geographical scope, and temporal range of the analysis. For example, studies targeting corn yield prediction in Iowa structured their models around the entire growing season, May to September, while research on cabbage and radish in Gangwon-do, South Korea, focused on data from June to September to align with local agricultural cycles. Clearly defining these parameters helps ensure that the data inputs and model configurations align with the biological growth stages of the crops under study [2, 6].

The data preprocessing stage is critical for preparing the satellite and climate data to ensure consistency, cleanliness, and structure compatible with machine learning models. The first task involves data filtering to isolate relevant cropland or specific crop type. This filtering process may use land cover classification maps to extract only the areas designated for the crops of interest, such as corn fields or vegetable plots, thus minimizing noise from irrelevant regions. Next, temporal grouping segments the data into meaningful periods that correspond with the crops' growth stages, enhancing the model's sensitivity to phenological phases. Grouping data by growth stages or specific seasonal windows, such as monthly or bi-monthly intervals, allows the model to capture variations in crop development during critical periods. For instance, the data may be organized as the entire growing season (e.g., May to September) or as discrete monthly intervals, depending on crop sensitivity to environmental conditions. Additionally, data normalization and scaling are applied to ensure consistency across different data sources, which reduces bias and improves compatibility between vegetation indices and climate variables. Normalization techniques like Z-score standardization or min-max scaling are especially important for deep learning architectures that require well-scaled input for effective learning.

Model selection and configuration depend on the characteristics of the data and the specific objectives of the prediction task. Random Forest (RF) models are frequently used for structured agricultural data, as they leverage ensembles of decision trees to capture complex interactions within the dataset. Support Vector Machine (SVM) models are beneficial for datasets with clear class separability, though they require careful tuning of kernel functions, such as linear or Gaussian, to fit the data's structure. Deep learning approaches, particularly Convolutional Neural Networks (CNNs), are increasingly applied in agriculture due to their ability to process spatially structured data from satellite images.

Support Vector Machine (SVM)

SVM is widely used for its accuracy in classification and predictive tasks. The model identifies an optimal hyperplane that maximizes the margin between support vectors from different classes, reducing errors. A Gaussian RBF kernel was used in this study to capture nonlinear patterns. The optimization process involved tuning kernel parameters and regularization constants to balance accuracy with computational efficiency, as SVM is sensitive to overfitting when parameters are not optimized [10]:

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

subject to $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1$ for $i = 1, \ldots, n$, where:

(Picture 3 – SVM optimization equation)

Random Forest (RF)

RF, based on the CART algorithm, leverages ensemble learning by combining multiple decision trees using bootstrap aggregation. This method builds decision trees from random subsets of the training data, with final predictions determined by majority voting (classification) or averaging (regression). In this study, RF was configured with 500 trees, splitting nodes based on a third of the total input variables (n/3). The model's performance was evaluated using out-of-bag error as a metric, providing an unbiased estimation of prediction accuracy [2]:

$$Y = \frac{1}{L}\sum_{l=1}^{L} Y_l$$

(Picture 4 – Averaging equation)

This experimental configuration included 500 trees, with the splitting variables set to a third of the total input variables (n/3). In the research also utilized the out-of-bag error as a performance metric [2].

Extremely Randomized Trees (ERT)

ERT is a variant of RF that builds trees without bootstrap resampling and selects split points randomly, which enhances diversity among trees. This approach reduces model bias and enhances generalization but increases variance. The settings for ERT, such as the number of trees and node splitting variables, were aligned with RF for direct performance comparison [2]. The ERT model demonstrated greater flexibility in handling noisy data, which is common in large agricultural datasets.:

$$\hat{y} = \frac{1}{M}\sum_{m=1}^{M} T_m(x)$$

(Picture 5 – Ensemble prediction formula)

Deep Learning (DL)

DL expands traditional artificial neural networks (ANN) with multi-layer architectures, effectively managing large, complex datasets. The training process involves unsupervised pre-training to refine representations, followed by supervised fine-tuning to optimize classification performance. In this study, a deep learning model with a 200×200 configuration was utilized, undergoing pre-training and fine-tuning to capture high-dimensional patterns within the agricultural data [3].

Convolutional Neural Network (CNN)

CNNs are particularly well-suited for analyzing visual data such as satellite imagery, making them ideal for capturing spatial and temporal patterns related to crop health. The CNN model in this study analyzed time-series satellite data, automatically extracting features without manual engineering. ReLU activation and Adam optimization were used, and a dropout rate of 0.5 was applied across all layers to prevent overfitting. The model input was organized as histograms, with 11 variables represented in 32-bin formats for each district, allowing the CNN to learn subtle spatial correlations between satellite imagery and crop yields [5].

Result and Discussion

In the study, various machine learning models, including SVM, Random Forest (RF), Extremely Randomized Trees (ERT), and Deep Learning (DL), were used to predict corn yields, benchmarked against USDA statistics. The dataset, spanning the MJJAS growing season, was validated using the leave-one-year-out cross-validation method, generating 11 sets of results (2004-2014). The DL model achieved the highest correlation (0.776) with an RMSE of 0.844 ton/ha, while RF and ERT demonstrated comparable accuracy (correlation coefficients of 0.651 and 0.654, respectively). SVM had the lowest performance with a correlation of 0.590 and an RMSE of 0.959 ton/ha.

RF and ERT performed well across different seasonal periods, capturing the seasonal sensitivities of corn yields. DL outperformed across all periods (MJJAS=0.776), while RF and ERT also showed robust performance. SVM was more prone to overfitting, a common issue in complex models, where minor dataset fluctuations impacted performance.

Table 1 – Validation statistics for the period groups MJJAS

	Mean bias (ton/ha)	MAE (ton/ha)	RMSE (ton/ha)	MAPE (%)	r
SVM	0.112	0.730	0.959	8.1	0.590
RF	0.063	0.666	0.879	7.3	0.651
ERT	0.091	0.674	0.891	7.4	0.654
DL	-0.031	0.657	0.844	6.9	0.776

In a separate analysis using OLS regression, climate variables and soil properties were assessed for their influence on corn and soybean yields. NDVI had the most significant influence on yields for both crops, with precipitation (PPT) and minimum temperature (Tmin) playing important roles for corn and soybeans, respectively. The study highlighted how soil texture and nitrogen fertilizer (NTRG) also contributed to crop performance, emphasizing the need for further analysis on how irrigation practices interact with soil properties.

Validation results for 2011-2012 showed reasonable error metrics (MAE of 0.726 ton/ha for corn in 2011 and 1.046 ton/ha in 2012). Prediction errors increased in 2012 due to drought conditions, impacting model performance. However, correlation coefficients remained high (0.909 for corn in 2011 and 0.903 for soybeans).

The proposed CNN model demonstrated strong predictive capabilities for cabbage and radish yields in temperate, seasonally variable climates, achieving high correlation values (up to 0.7046 for radish and 0.6350 for cabbage) and relatively low RMSE (1,358 to 1,553), thereby highlighting its ability to capture complex, spatiotemporal data patterns. Despite these strengths, the model exhibited limitations during early prediction stages and under extreme climate conditions, such as the 2018 heatwave, where it overestimated yields, suggesting a need for architectural adaptations or additional weather-related variables to enhance robustness.

Table 2 – Validation results of the regression models (OLS) for corn and soybean yields across different years, including metrics such as mean bias, MAE, RMSE, MAPE

	No. of Counties (ton/ha)	Min Obs. (ton/ha)	Max Obs. (ton/ha)	Mean Obs. (ton/ha)	Mean Bias (ton/ha)	MAE (ton/ha)	RMSE (ton/ha)	MAPE (%)	R
Corn (2011)	180	4.871	12.334	9.456	-0.350	0.726	0.861	8.144	0.909
Corn (2012)	182	1.883	12.208	8.463	-0.675	1.046	1.240	12.681	0.854
Soybeans (2011)	181	1.715	4.304	2.871	-0.093	0.252	0.313	8.812	0.903
Soybeans (2012)	177	0.780	3.766	2.762	-0.347	0.392	0.442	14.616	0.877

The study emphasizes the importance of selecting stable regions to minimize the impact of unpredictable factors like natural disasters on yield predictions. Machine learning models such as RF, ERT, and DL demonstrate strong potential for improving agricultural productivity and food security through precision agriculture. Comprehensive evaluations, including metrics like MAE, RMSE, and R-squared, along with visualizations like scatter plots and residual analysis, provide a thorough understanding of model performance.

Through a comparative evaluation of models, including CNNs, RF, ERT, SVM, and Ordinary Least Squares (OLS) regression, this review identifies specific advantages under varying agricultural conditions. CNNs perform effectively in large-scale spatial analysis using satellite imagery, while neural networks based on soil properties offer superior accuracy in areas where soil chemistry significantly impacts yield.

Table 3 – 8:2 validation results for radish and cabbage using the proposed CNN model, reference CNN model (Mu et al., 2019), and Random Forest

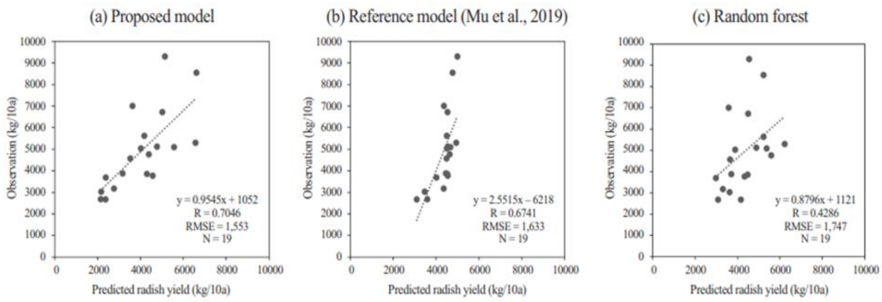

Fig. 4. 8:2 validation results for radish: (a) proposed CNN model result, (b) reference CNN model (Mu et al., 2019) result, and (c) RF model result.

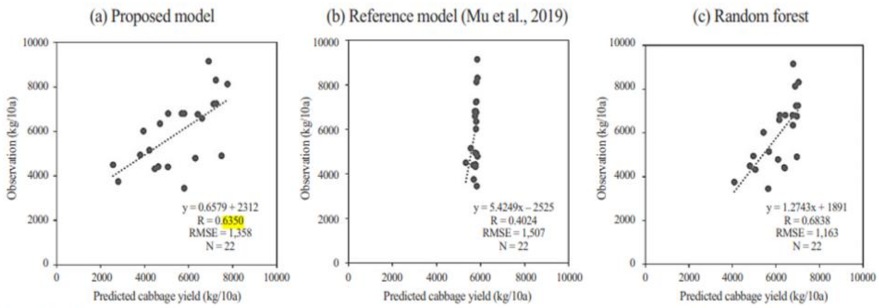

Fig. 5. 8:2 validation results for cabbage: (a) proposed CNN model result, (b) reference CNN model (Mu et al., 2019) result, and (c) RF model result.

Random Forest and ERT models provide robustness in heterogeneous environments with diverse soil types and climates, making them ideal for multi-crop or variable conditions. In homogeneous or monoculture systems, OLS regression and SVM offer simplicity and effectiveness, particularly under stable climatic patterns. The integration of multiple data sources – satellite imagery, soil properties, and climate metrics – further enhances model accuracy, enabling tailored adaptations across crops and regions.

Table 4 – Summarized results in the table

Model	**Crop**	**Climate Zones**	**Key Metrics**	**r**	**MAE**	**RMSE**	**MAPE**	**Strengths**
OLS Regression	Corn, Soybeans	Temperate and Stable	Moderate accuracy, higher error margin for complex patterns	0.854 - 0.909	0.252 - 1.046	0.313 - 1.240	8.144% - 14.616%	Simple and interpretable, suitable for stable environments
Support Vector Machine (SVM)	Corn	Temperate	Lower accuracy in extreme conditions; performs better in moderately stable climates	0.575 - 0.606	0.650 - 0.730	0.852 - 0.959	7.3% - 8.1%	Effective for simple patterns; struggles in highly variable climate zones
Random Forest (RF)	Corn	Temperate with Seasonal Variation	High correlation in optimal month combinations, effective for seasonal sensitivity	0.651 - 0.774	0.616 - 0.666	0.767 - 0.879	6.3% - 7.3%	Handles non-linear patterns well, reduced overfitting compared to simpler models
Extremely Randomized Trees (ERT)	Corn	Temperate with Seasonal Variation	Similar performance to RF with slight variance reduction	0.654 - 0.785	0.568 - 0.674	0.756 - 0.891	6.1% - 7.4%	Highly efficient, suitable for large datasets with complex patterns
Deep Learning (DL)	Corn	Temperate and Drought-Prone	High accuracy, especially in drought-resistant predictions, handles overfitting effectively	0.776 - 0.800	0.608 - 0.709	0.787 - 0.901	6.5% - 7.5%	Best for complex, non-linear relationships, adaptable to both stable and extreme conditions
CNN	Cabbage, Radish	Temperate, Seasonal	High prediction accuracy, performs well under seasonal variation	0.635 - 0.7046	N/A	1,358 - 1,553	N/A	Effective for handling spatiotemporal data, but less adaptable to seasonal changes

Conclusion

This comprehensive meta-analysis underscores the transformative impact of integrating advanced machine learning techniques with high-resolution satellite imagery and soil chemical properties for crop yield prediction. By synthesizing methodologies across various studies, this research highlights how models like Convolutional Neural Networks (CNNs), Random Forests (RF), Extremely Randomized Trees (ERT), Support Vector Machines (SVM), and Ordinary Least Squares (OLS) regression each offer unique advantages in handling the complexities of modern agricultural ecosystems.

CNNs demonstrated superior predictive accuracy for crops like radish under stable environmental conditions, leveraging their ability to automatically extract features from high-dimensional data. Random Forests and ERT models exhibited robustness in heterogeneous environments with diverse soil types and climatic variables, making them well-suited for multi-crop systems and regions with variable conditions. These ensemble methods effectively handled nonlinear relationships and reduced overfitting risks, contributing to reliable yield predictions.

In contrast, SVMs and OLS regression models offered simplicity and computational efficiency, performing effectively in homogeneous or monoculture systems with stable climatic patterns.

A critical insight from this research is the significant enhancement of prediction accuracy when integrating soil chemical properties – such as pH, nutrient content, and organic matter—with satellite-derived vegetation indices. This comprehensive approach acknowledges the essential role of below-ground factors in crop development, providing a more holistic understanding of the factors influencing yields. The fusion of above-ground and below-ground data allows models to account for the spatial and temporal variability inherent in agricultural environments.

The study also emphasizes the importance of selecting stable regions for yield prediction to minimize the impact of unpredictable factors like natural disasters. By focusing on areas with consistent environmental conditions, models can achieve higher accuracy, which is crucial for informing agricultural policies and resource management strategies.

Through the comparative evaluation of different models, the research illustrates that no single method universally outperforms others across all conditions. Instead, the choice of model should be tailored to specific agricultural contexts, considering factors such as data availability, computational resources, crop types, and environmental variability.

References

1. Statistical Estimation of Crop Yields for the Midwestern United States Using Satellite Images, Climate Datasets, and Soil Property Maps / N. Kim et al // Korean Journal of Remote Sensing. – 2016. – № 32(4). – P. 383-401.
2. Kim N. Machine Learning Approaches to Corn Yield Estimation Using Satellite Images and Climate Data: A Case of Iowa State / N. Kim, Y.-W. Lee // Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography. – 2016. – № 34(4). – P. 383-390.
3. Chang J. Identifying Factors for Corn Yield Prediction Models and Evaluating Model Selection Methods / J. Chang, D.E. Clay // Korean Journal of Crop Science. – 2005. – № 50(4). – P. 268-275.
4. Boosted random forest / Y. Mishina et al // IEICE Transactions on Information and Systems. – 2005. – № E98.D(9). – P. 1630-1636. https://doi.org/10.1587/transinf.2014OPP0004.
5. Li.Y. Regional segmentation of field images based on convolutional neural network for rice combine harvester. / Y. Li., M. Iida, M. Suguri, R. Masuda // Journal of the Japanese Society of Agricultural Machinery and Food Engineers. – 2020. – № 82(1). – P. 47-56. https://doi.org/10.11357/jsamfe.82.1_47.
6. Park H. Satellite-based cabbage and radish yield prediction using deep learning in Kangwon-do / H. Park, Y. Lee, S. Park // Korean Journal of Remote Sensing. – 2023. – Vol. 39. – P. 1031-1042. https://koreascience.kr/article/JAKO202331857673593.page.
7. Sung J.H. Rice yield prediction based on the soil chemical properties using neural network model / J.H. Sung, D.H. Lee // Journal of Biosystems Engineering. – 2004. – № 29(2). – P. 123-135.
8. Relationships between global climate indices and rain-fed crop yields in highland of South-Central Java, Indonesia / B.D.A. Nugroho et al // Journal of Geography (Chigaku Zasshi). – 2013. – № 122(3). – P. 438-447. https://doi.org/10.5026/jgeography.122.438.
9. Purwanto M.Y.J. Crop yield prediction by stress day indices under both excessive and deficient soil water conditions / M.Y.J. Purwanto, S. Hardjoamidjojo, R. Nakamura, N. Kubo // Journal of

Irrigation Engineering and Rural Planning. – 1993. – № 25. – P. 31-41. https://doi.org/10.11408/jierp1982.1993.25_31.
10. Jun S.H. A differential evolution based support vector clustering / S.H. Jun // Journal of the Korean Institute of Intelligent Systems. – 2007. – № 17(5). – P. 679-683. https://doi.org/10.5391/JKIIS.2007.17.5.679.

Д.В.Сон
Астана IT Университет,
010000, Республика Казахстан, г. Астана, пр. Мәңгілік Ел, С1
e-mail: qwerty.01.qwerty0op@gmail.com

ПРОГНОЗИРОВАНИЕ УРОЖАЙНОСТИ ПОЧВЫ

Эта исследовательский проект представляет собой комплексный метаанализ в области сельскохозяйственной науки, в котором особое внимание уделяется прогнозированию урожайности сельскохозяйственных культур. Это исследование включает в себя сопоставление и синтез результатов различных исследований и статей, в которых рассматриваются различные методологии и модели прогнозирования сельскохозяйственной продукции. Целью этого всеобъемлющего обзора является выявление тенденций, методологий и ключевых факторов, которые последовательно влияют на прогнозы урожайности сельскохозяйственных культур в рамках различных исследований.

В нем обобщены методологии из различных исследований, особое внимание уделяется методам машинного обучения (ML), таким как методы опорных векторов (SVM), случайный лес (RF) и сверточные нейронные сети (CNN). Эти исследования объединяют спутниковые снимки высокого разрешения с экологическими показателями, такими как NDVI, EVI и LAI. Химические свойства почвы (pH, питательные вещества) и полученные со спутника данные были использованы для улучшения прогнозирования урожайности различных культур. Полученные результаты свидетельствуют о сравнительной эффективности различных моделей при обработке пространственной и временной изменчивости как наземных, так и подземных данных, что повышает точность прогнозирования в различных условиях окружающей среды и почвы.

Благодаря этому теоретическому анализу исследование подчеркивает потенциал передовых аналитических моделей для преобразования сельскохозяйственного мониторинга и прогнозирования, предоставляя важную информацию, которая может помочь в оптимизации сельскохозяйственной политики и управлении ресурсами.

Ключевые слова: *Прогнозирование урожайности сельскохозяйственных культур, Спутниковые снимки, Машинное обучение, Сверточные нейронные сети (CNN), Вегетационные индексы, Химические свойства почвы.*

Д.В. Сон
Астана IT Университеті,
010000, Қазақстан Республикасы, Астана, Мәңгілік Ел даңғылы, С1
e-mail: qwerty.01.qwerty0op@gmail.com

ТОПЫРАҚ ӨНІМДІЛІГІН БОЛЖАУ

Бұл зерттеу жобасы ауыл шаруашылығы ғылымы саласындағы кешенді мета-анализ ретінде қызмет етеді, атап айтқанда дақылдардың өнімділігін болжауға бағытталған. Бұл әрекет ауыл шаруашылығының өнімділігін болжаудың әртүрлі әдістемелері мен үлгілерін зерттеген әртүрлі зерттеулер мен мақалалардың нәтижелерін салыстыруды және синтездеуді қамтиды. Бұл жан-жақты шолудың мақсаты әртүрлі зерттеулердегі дақылдардың өнімділігін болжауға дәйекті түрде әсер ететін тенденцияларды, әдістемелерді және негізгі факторларды анықтау болып табылады.

Ол Әртүрлі зерттеулердің әдістемелерін синтездейді, Векторлық Машиналарды (SVM), Кездейсоқ Ормандарды (RF) Және Конволюциялық Нейрондық Желілерді (CNN) Қолдау сияқты машиналық оқыту (ML) әдістеріне баса назар аударады. Бұл зерттеулер жоғары ажыратымдылықтағы спутниктік суреттерді NDVI, EVI және LAI сияқты экологиялық көрсеткіштермен біріктіреді. Топырақтың химиялық қасиеттері (рн, қоректік заттар) және спутниктік деректер әртүрлі дақылдардың өнімділігін болжауды жақсарту үшін пайдаланылды. Нәтижелер әртүрлі модельдердің жер үсті және жер асты деректерінің кеңістіктік және уақыттық өзгергіштігін өңдеудегі салыстырмалы тиімділігін көрсетеді, қоршаған орта мен топырақтың әртүрлі жағдайларында болжау дәлдігін жақсартады.

Осы теориялық талдау арқылы зерттеу ауыл шаруашылығы саясатын оңтайландыруға және ресурстарды басқаруға көмектесетін маңызды ақпаратты ұсына отырып, ауыл шаруашылығы мониторингі мен болжамын өзгерту үшін озық аналитикалық модельдердің әлеуетін көрсетеді.

Түйін сөздер: *Дақылдардың Өнімділігін Болжау, Спутниктік Суреттер, Машиналық Оқыту, Конволюциялық Нейрондық Желілер(CNN), Өсімдік Жамылғысының Көрсеткіштері, Топырақтың Химиялық Қасиеттері.*

Information about the authors

Dmitriy Vladislavovich Son – Master's Student, Astana IT University, Republic of Kazakhstan, Astana; e-mail: qwerty.01.qwerty0op@gmail.com.

Сведения об авторах

Дмитрий Владиславович Сон – магистрант, Astana IT University, Республика Казахстан, г. Астана; e-mail: qwerty.01.qwerty0op@gmail.com.

Авторлар туралы мәліметтер

Дмитрий Владиславович Сон – Магистрант, Astana IT University; Қазақстан Республикасы, Астана қ.; e-mail: qwerty.01.qwerty0op@gmail.com.

Received 23.10.2024
Revised 04.11.2024
Accepted 13.11.2024

https://doi.org/10.53360/2788-7995-2024-4(16)-11

IRSTI: 53.81.31

B.K. Kenzhaliyev[1], S.Zh. Aibagarov[2], S.D. Bekele[2*], S.A. Kvyatkovskiy[1], M.A. Dyussebekova[1]
[1]Institute of Metallurgy and Ore Beneficiation, Satbayev University,
050010, Kazakhstan, Almaty, Shevchenko Street, 29,
[2]Al-Farabi Kazakh National University,
050040, Kazakhstan, Almaty, Al-Farabi Avenue, 71,
*e-mail: samsondawitb@gmail.com

DEVELOPMENT OF A COMPREHENSIVE SOFTWARE SOLUTION FOR PROCESSING HIGH-SULFUR, COPPER-POOR CONCENTRATES IN THE COPPER SMELTING INDUSTRY

Abstract: *The depletion of high-grade copper ores and the increasing prevalence of low-grade, high-sulfur copper concentrates present significant challenges to the copper smelting industry. Traditional smelting processes struggle to maintain economic viability and comply with environmental regulations when processing these complex ores. This paper details the development of a comprehensive software solution designed to simulate the smelting process for high-sulfur, copper-poor concentrates. The software employs a detailed mathematical model to predict the yields and compositions of products, including valuable metals, during the smelting process. It integrates multiple modules, such as ore input, smelting simulation, and results presentation, providing a user-friendly platform for optimizing smelting operations. Critical parameters like ore composition, smelting temperature, and flux addition are incorporated into the model, enabling accurate predictions of matte and slag outputs. By analyzing these outputs, the software aids in optimizing metal recovery and reducing losses, ultimately enhancing the efficiency and sustainability of copper production. This tool is particularly relevant for large sulfide copper ore deposits, such as those in Kazakhstan, which have high sulfur content and low copper levels. The software's ability to simulate different processing scenarios provides valuable insights for industrial applications, supporting the development of more efficient and eco-friendly smelting technologies. The comprehensive software solution not only addresses the technical challenges of processing high-sulfur, low-copper ores but also contributes to the industry's efforts to reduce environmental impact and improve resource management. This innovation represents a significant step forward in the optimization of copper smelting operations, promoting sustainability and efficiency in the face of declining ore quality.*

Key words: *copper smelting, high-sulfur ores, smelting simulation, metal recovery, optimization.*

Introduction

The copper smelting industry is transforming due to the depletion of high-grade ores and the rise of complex, low-grade ores [1, 2]. These ores, high in sulfur and low in copper, challenge