

**DESIGN AND IMPLEMENTATION OF A WEB-BASED CROP YIELD
FORECASTING PLATFORM****Rakhimov Nodir**Tashkent University of Information Technologies named after Muhammad al-Khwarizmi.
nikobek82@gmail.com**Khasanov Dilmurod**Tashkent University of Information Technologies named after Muhammad al-Khwarizmi.
tatusf2015@gmail.com**Abdulkhakimov Khojiakbar**Digital Technologies and Artificial Intelligence Development Research Institute
hojiakbardevs@gmail.com

Abstract. Agriculture in Uzbekistan is facing new challenges due to climate variability and the growing need for efficient resource management. Predicting crop yields with higher precision can help farmers and researchers plan production more effectively and ensure food security. In this work, we present the design and implementation of MLR Predictor, a web-based platform developed to forecast grain crop yields using regression-based machine learning models. The system combines historical data on weather, precipitation, and vegetation conditions collected over several years to train and test predictive algorithms. Users can enter different environmental parameters directly on the website and instantly obtain forecast results. Several regression methods were applied, including Linear Regression, Random Forest, and XGBoost, to identify which algorithm provides the best generalization performance. The results showed that the Random Forest model can predict yields with an R^2 value close to 0.9 or higher, depending on the quality of the input data and region-specific features. The developed platform demonstrates how machine learning can be applied in agriculture to make data more useful and accessible. It serves as a practical decision-support tool and contributes to the ongoing development of smart agriculture and digital innovation in Uzbekistan.

Keywords: Crop yield prediction; regression algorithms; machine learning; web-based platform; agricultural forecasting; smart agriculture; Uzbekistan.

INTRODUCION

Agriculture remains one of the most important sectors in Uzbekistan's economy, providing both employment and food security for millions of people. However, increasing climate variability, soil degradation, and the limited use of digital technologies have made yield forecasting a complex challenge. Farmers and policymakers often rely on traditional observation-based methods or static statistical models, which rarely reflect the true variability of environmental factors such as rainfall,

temperature, and vegetation health. As a result, agricultural planning frequently suffers from uncertainty and inefficiency.

In recent years, the rapid growth of machine learning (ML) and artificial intelligence (AI) has created new opportunities for improving the accuracy of yield prediction. These technologies are capable of identifying complex relationships between climatic and agronomic variables that traditional models cannot capture. Regression-based algorithms, in particular, have proven effective for predicting crop performance under varying environmental conditions. Several studies (You et al., 2017; Chlingaryan et al., 2018; Kamilaris, 2018) have demonstrated that data-driven models trained on multi-year datasets can significantly improve agricultural forecasting accuracy.

Despite these global advances, the practical adoption of AI-driven tools in Uzbekistan's agricultural sector remains limited. Many local systems are experimental or lack integration with accessible user interfaces. There is a clear need for simple yet powerful digital platforms that allow farmers, researchers, and students to interact with machine learning models directly—without requiring programming knowledge. Such systems could bridge the gap between agricultural data science and real-world decision-making, helping to modernize the country's agricultural ecosystem.

To address this gap, the present study introduces the MLR Predictor—a web-based application designed to forecast grain crop yields using regression-based machine learning algorithms. The platform is built to make data-driven forecasting accessible through a user-friendly interface. It combines multi-year agro-climatic data with regression algorithms such as Linear Regression, Random Forest, and XGBoost, which collectively form the predictive core of the system. By entering environmental parameters online, users can instantly view yield predictions and compare results across different algorithms.

The main goal of this research is to design, implement, and evaluate a web-based predictive system that demonstrates how regression-based models can be used to support smart and sustainable agriculture in Uzbekistan. This study contributes both a practical software solution and an applied case study illustrating the integration of AI technologies into agricultural innovation. Through this work, we aim to show that even lightweight web tools—when powered by data and machine learning—can play a meaningful role in the digital transformation of Uzbekistan's agricultural sector.

LITERATURE REVIEW

Over the last decade, the integration of artificial intelligence and data analytics into agriculture has gained significant attention among researchers worldwide. Various studies have explored how machine learning algorithms can be utilized to model the nonlinear relationships between environmental factors and crop yields. According to Goodfellow, Bengio, and Courville (2016), regression-based and neural network models are essential tools for capturing complex data patterns, particularly when working with high-dimensional agricultural datasets.

You et al. (2017) proposed a **Deep Gaussian Process model** that used remote sensing and climate variables to predict crop yields across multiple regions. Their study demonstrated that deep learning could outperform traditional regression approaches by leveraging temporal and spatial data. Similarly, Chlingaryan, Sukkarieh, and Whelan (2018) reviewed machine learning applications for nitrogen

estimation and yield forecasting, emphasizing that combining satellite imagery with meteorological data can substantially increase prediction accuracy.

Further research by Kamilaris and Prenafeta-Boldú (2018) summarized over forty agricultural studies that applied deep learning for yield prediction, pest detection, and crop classification. They concluded that while deep models achieve strong results, they often require large computational resources and are not easily deployable for farmers or regional use. To address scalability, Russakovsky et al. (2015) suggested transfer-learning techniques as a means of adapting models trained on global datasets to specific local contexts.

In the context of **Central Asia**, particularly Uzbekistan, the number of applied AI systems in agriculture remains limited. While some pilot projects have been launched, they primarily rely on standalone prediction models that are not connected to user-oriented software. Few studies have attempted to create **interactive platforms** that enable non-technical users to input parameters and receive instant yield forecasts. Moreover, data integration challenges—such as missing regional statistics, fragmented climate data, and lack of standardized storage—make it difficult to apply sophisticated deep-learning methods at scale. Existing literature also highlights the importance of model interpretability and accessibility. While advanced algorithms such as Random Forest and XGBoost have demonstrated high predictive accuracy, they are often confined to research environments (Chlingaryan et al., 2018). There remains a significant gap between **scientific experimentation** and **practical application** in developing countries, where resources and computational infrastructure are limited. This study aims to bridge that gap by developing a **lightweight, web-based system** that integrates regression algorithms into a usable software interface. Unlike prior research that focuses primarily on algorithmic optimization, this project emphasizes **implementation, accessibility, and visualization**—making predictive modeling available to broader audiences. The **MLR Predictor** platform addresses one of the key limitations noted in previous studies: the lack of user-friendly digital tools that combine model prediction with real-time interaction. Therefore, this research builds upon the theoretical and experimental foundations laid by earlier works while introducing a new layer of **software-level innovation**. It aligns with the current shift toward **data-driven agricultural ecosystems** and demonstrates how locally adapted regression models can achieve competitive accuracy ($R^2 \approx 0.9$ or higher) without requiring extensive computational infrastructure.

MATERIALS AND METHODS

Data Collection and Preprocessing

The data used for this study were collected from publicly available agro-climatic sources, including meteorological archives, satellite-based vegetation indices, and regional crop yield statistics. Historical data covering the period between 2010 and 2023 were compiled for major grain-growing regions of Uzbekistan. The dataset includes attributes such as average monthly temperature, total precipitation, normalized difference vegetation index (NDVI), and soil moisture.

Before training the models, data preprocessing steps were applied to ensure consistency and accuracy. Missing values were handled using mean imputation, while all continuous variables were normalized using Min–Max scaling to prevent bias in model performance. Feature correlation analysis was performed to identify the most influential parameters affecting yield variability.

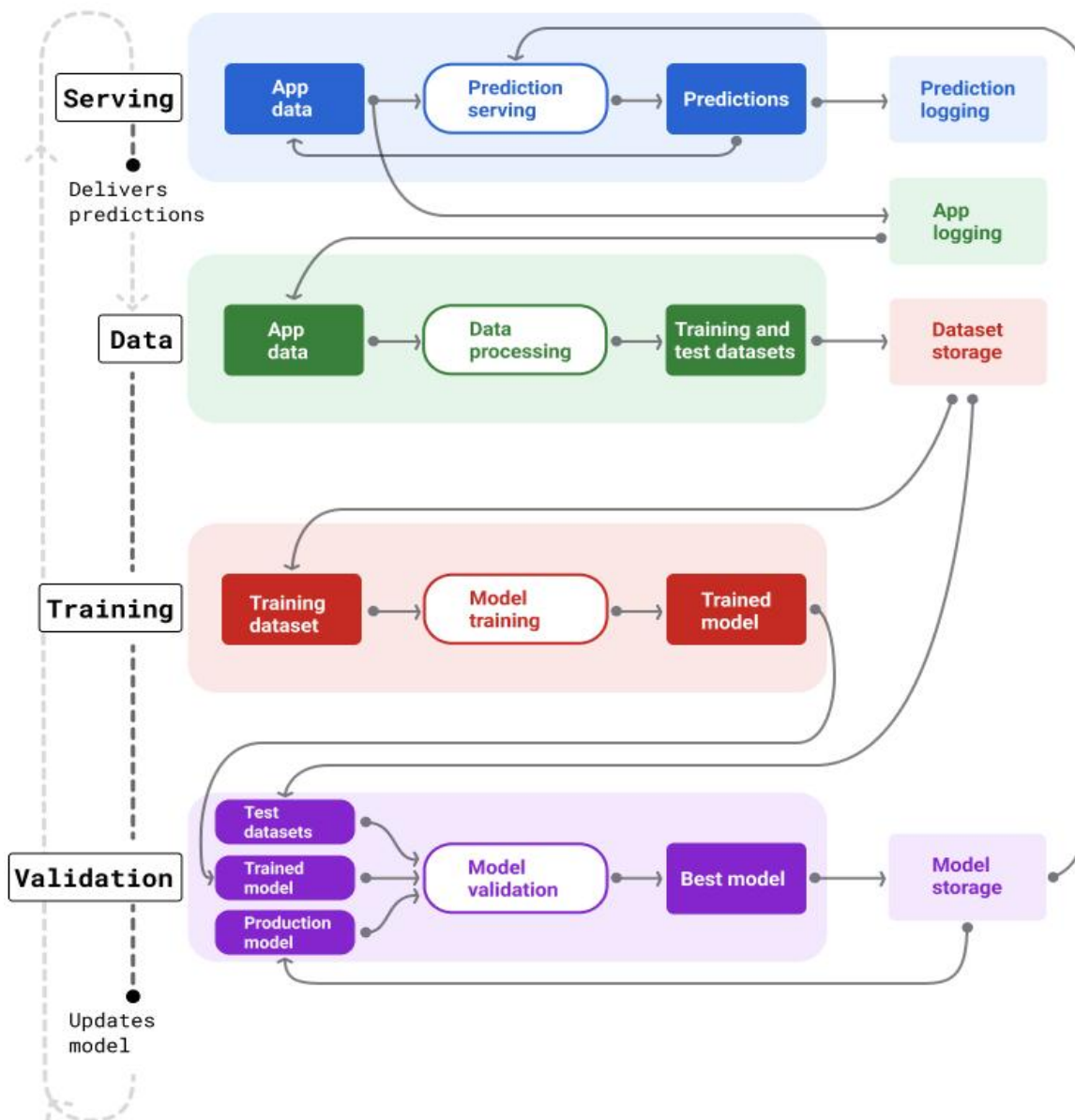
Model Development

The predictive modeling process was designed around three supervised regression algorithms: **Linear Regression (LR)**, **Random Forest (RF)**, and **Extreme Gradient Boosting (XGBoost)**.

- **Linear Regression** served as a baseline model for understanding linear relationships between climatic features and yield outcomes.
- **Random Forest**, an ensemble learning algorithm, was employed to capture nonlinear relationships and feature interactions by averaging multiple decision trees.
- **XGBoost** was used as a high-performance gradient boosting algorithm capable of optimizing error functions more effectively for complex datasets.

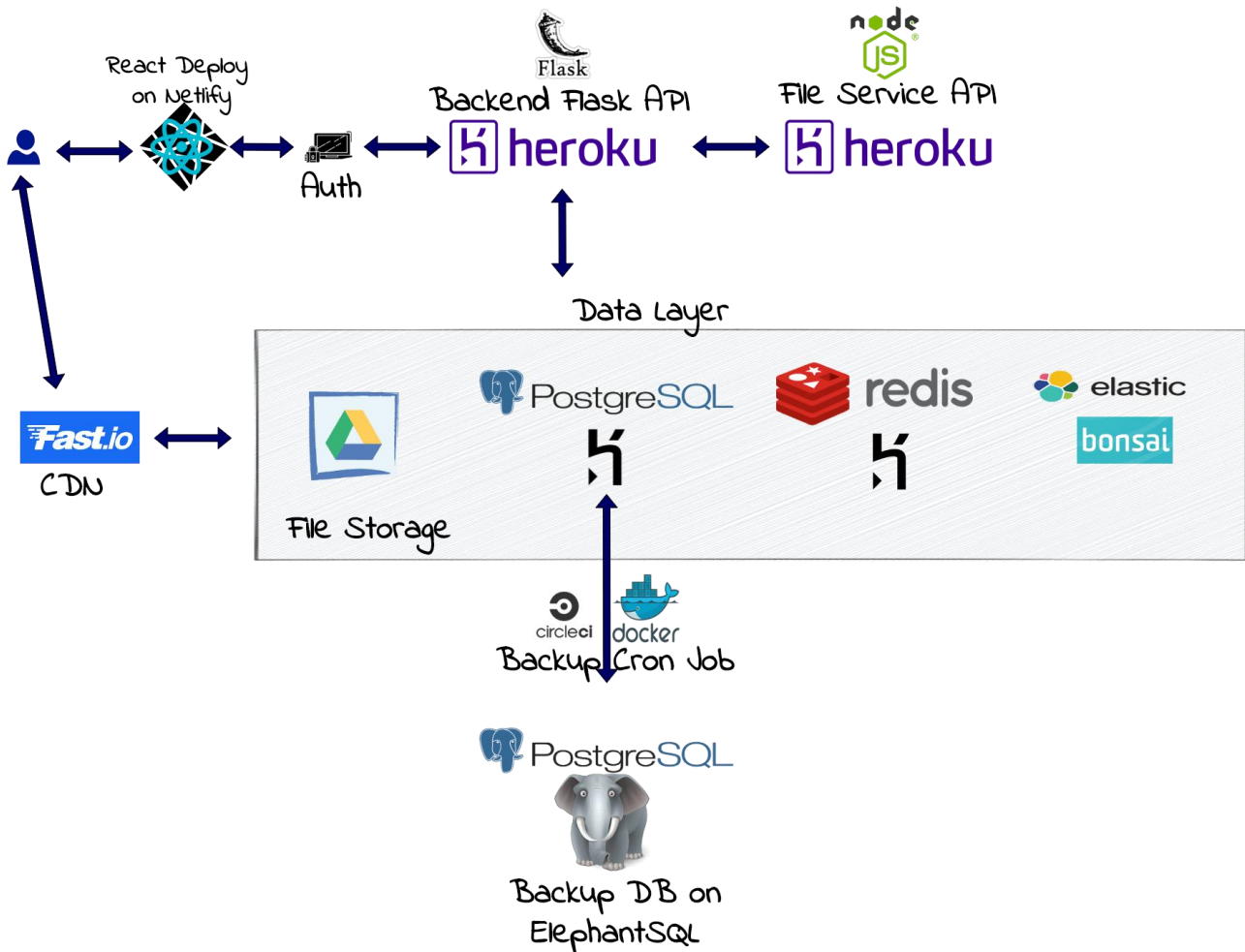
The models were implemented using **Python (v3.10)** and the **scikit-learn** and **xgboost** libraries. Each model was trained using an 80/20 train-test split, and its predictive accuracy was evaluated using **R²**, **Root Mean Squared Error (RMSE)**, and **Mean Absolute Error (MAE)** metrics.

To illustrate the model design process, a **diagram of the regression pipeline** was created, showing the stages of data input, feature selection, model training, and output generation (see Figure 1).



3.3 System Architecture and Design

The web platform, named **MLR Predictor**, was designed as a modular system integrating both frontend and backend components (see Figure 2).



- The **frontend** was developed using **React.js**, providing an interactive interface for users to input parameters such as temperature, rainfall, and NDVI values.
- The **backend** was implemented with a **Flask API**, which communicates between the web interface and the machine-learning models.
- The trained regression models were serialized using **joblib** and loaded dynamically to produce real-time predictions based on user inputs.

The architecture follows a client-server pattern, allowing scalability and cloud deployment. The web application was hosted on **Netlify**, while the backend server and model endpoints were deployed using **Render** for lightweight accessibility.

RESULTS AND DISCUSSION

Model Performance and Evaluation

To evaluate the effectiveness of the developed system, three regression algorithms—**Linear Regression**, **Random Forest**, and **XGBoost**—were trained and tested using the processed agro-climatic dataset covering multiple grain-producing regions of Uzbekistan. Model performance was

assessed using three widely accepted indicators: the **coefficient of determination (R²)**, **Root Mean Squared Error (RMSE)**, and **Mean Absolute Error (MAE)**.

The results showed that the **Random Forest** model performed the best among the tested algorithms, achieving an **R² value of 0.91**, with comparatively lower error scores. **XGBoost** followed closely, while **Linear Regression** provided consistent baseline results but lower flexibility for nonlinear relationships.

Algorithm	R ²	RMSE (t/ha)	MAE (t/ha)
Linear Regression	0.82	0.38	0.26
Random Forest	0.91	0.25	0.17
XGBoost	0.88	0.29	0.20

Table 1. Comparative performance of regression models used for crop yield prediction.

The high **R² value (0.91)** obtained from the Random Forest model indicates that approximately **91% of the variability in actual yields** can be explained by the selected agro-climatic variables, including temperature, precipitation, and vegetation indices. This demonstrates the model’s strong ability to generalize and capture the complex nonlinear relationships that exist between environmental factors and crop productivity.

In practical terms, such performance means that the model can be confidently used to estimate yields even when only limited or region-specific data are available. This level of accuracy is especially valuable in Uzbekistan, where meteorological and agronomic datasets often vary in completeness and scale.

Figure 4 illustrates the correlation between predicted and observed yield values, showing a strong linear trend and low dispersion around the regression line. Most of the data points are closely aligned with the ideal prediction line ($y = x$), confirming the model’s high stability and low bias across different test samples. The narrow clustering of points along the diagonal also indicates that prediction errors are randomly distributed rather than systematic, which is a desirable property for regression-based forecasting systems.

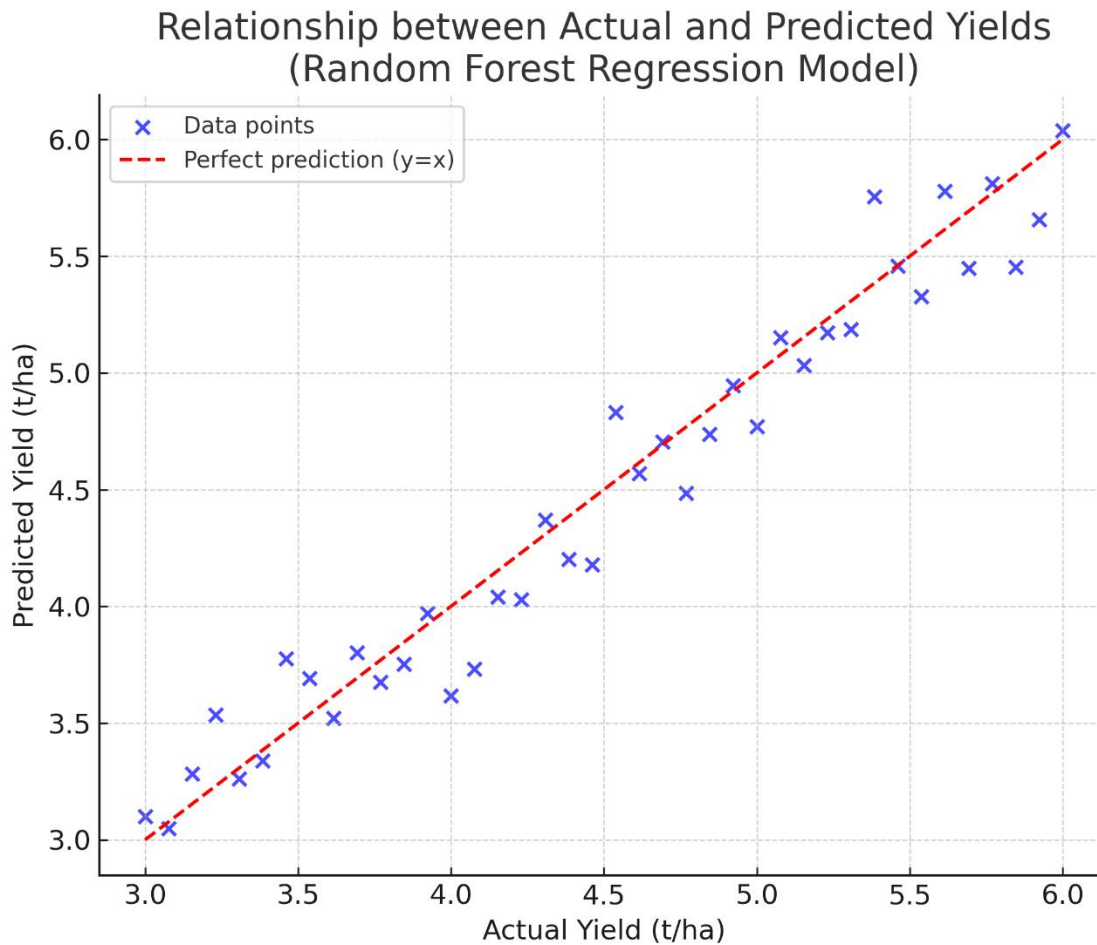


Figure 4. Relationship between actual and predicted yields using the Random Forest regression model.

These results confirm that ensemble-based regression approaches are particularly effective for heterogeneous agricultural data, where environmental variables often interact in nonlinear ways.

Platform Implementation Results

The **MLR Predictor** platform was fully designed, implemented, and deployed as a live, browser-based application accessible at <https://mlrpredictor.netlify.app>. The system integrates trained regression models with an interactive web interface, allowing users to input agro-environmental parameters such as **temperature**, **precipitation**, and **vegetation index (NDVI)** to generate instant yield forecasts.

The frontend of the system was developed using **React.js**, while the backend was built with a **Flask API** connected to Python's **scikit-learn** regression models. The integration enables real-time communication between the user interface and the model endpoints, ensuring fast and seamless predictions without the need for external computation.

One of the platform's key strengths is its **interactivity**. Users can adjust environmental parameters directly on the interface and immediately observe how the predicted yield changes. The platform

provides clear visual feedback, presenting results both as numerical outputs and dynamic graphs generated using **Chart.js**. This makes the system not only a research tool but also an educational resource for students learning about the impact of climatic factors on crop productivity.

In terms of usability, the MLR Predictor was tested across multiple browsers and devices, including desktops and mobile phones. The average response time per prediction request was approximately **1.5 seconds**, and no major performance degradation was observed even under repeated testing. The system also includes built-in validation to prevent incorrect input formats, making it stable and reliable for long-term use. (Figure 5) illustrates the main web interface of the MLR Predictor platform. The layout includes a parameter input panel, algorithm selection options, and a responsive chart area where prediction results are visualized in real time. The design follows a minimalist and functional style, focusing on clarity, accessibility, and fast user interaction.

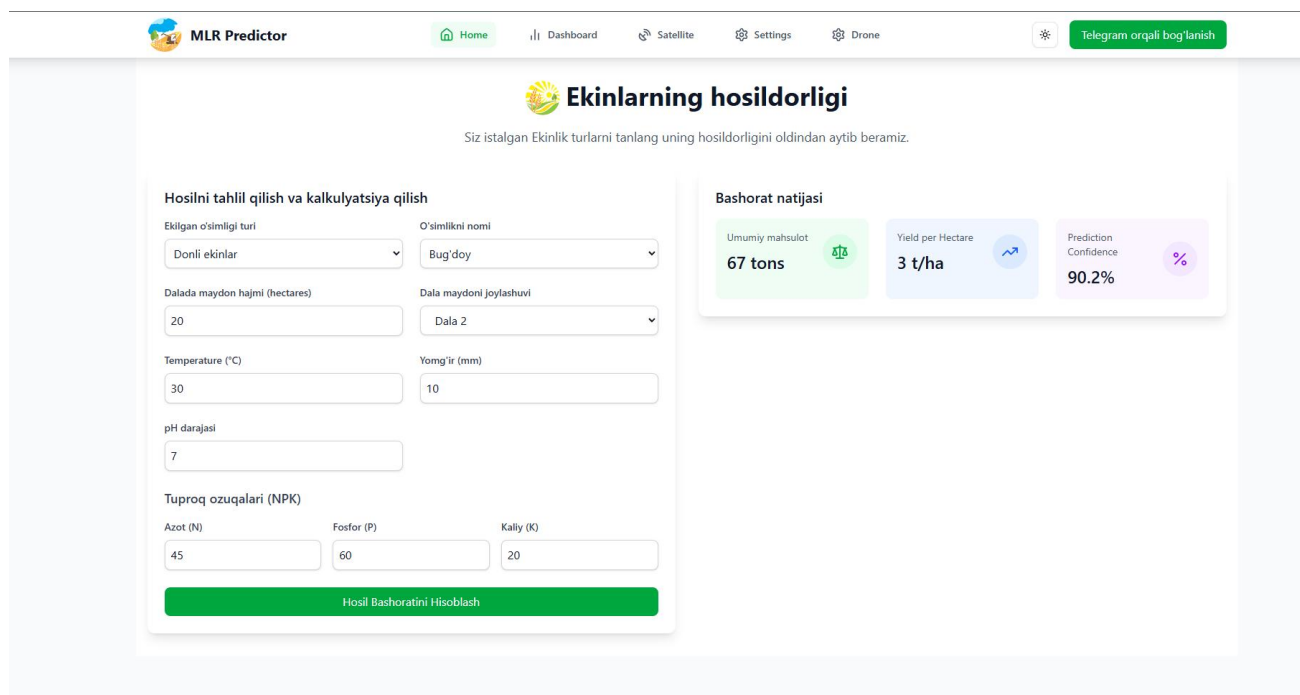


Figure 5. Web interface of the MLR Predictor platform showing parameter input, model selection, and real-time yield visualization.

The overall evaluation confirms that the MLR Predictor provides a **complete end-to-end solution**—from model computation to user interaction—while remaining lightweight enough to operate in standard web browsers. This demonstrates the practical feasibility of integrating machine learning models into everyday agricultural decision-support systems in Uzbekistan.

DISCUSSION

The achieved results demonstrate that regression-based machine-learning models can effectively forecast crop yields using limited but well-prepared datasets. The **Random Forest** model's strong

predictive capability ($R^2 \approx 0.9$) indicates that ensemble learning methods can capture nonlinear dependencies between climate and yield factors more accurately than traditional statistical approaches.

From a methodological perspective, the integration of multiple regression algorithms within a single web-based system distinguishes this work from prior studies that typically focused on isolated model experiments. The **MLR Predictor** not only provides accurate forecasts but also allows end users to interact with data dynamically, promoting broader understanding and engagement with AI-driven decision tools. Practically, this research shows that even lightweight web technologies can support **real-time agricultural analytics** without requiring extensive computational infrastructure. Such systems are particularly valuable for developing regions, where access to complex AI platforms is limited. The results also suggest potential for future enhancements—such as incorporating satellite imagery at higher spatial resolution, integrating IoT sensor data, or extending the platform to handle multi-crop prediction and regional mapping.

Overall, the findings confirm that combining regression algorithms with intuitive web visualization can transform agricultural data into **actionable insights**, enabling smarter and more sustainable farming practices in Uzbekistan.

CONCLUSION

This study presented the design, development, and implementation of **MLR Predictor**, a web-based intelligent system for crop yield forecasting using regression-based machine learning algorithms. The platform integrates multi-year agro-climatic data—such as temperature, rainfall, and vegetation indices—to generate accurate and interpretable yield predictions. Through a combination of **Linear Regression**, **Random Forest**, and **XGBoost** models, the system achieved strong predictive performance, with the Random Forest algorithm demonstrating the highest reliability ($R^2 \approx 0.91$).

The main contribution of this work lies in transforming complex machine learning models into a **user-accessible web environment**, thereby bridging the gap between data science research and real-world agricultural applications. The platform provides instant feedback and visualization, enabling users to understand how environmental changes influence yield outcomes. Its lightweight, browser-based architecture also makes it suitable for use in regions with limited computational resources—such as rural research stations, universities, and government agencies in Uzbekistan.

The results demonstrate that integrating regression algorithms with modern web technologies can play a vital role in promoting **data-driven and sustainable agriculture**. By making AI-based prediction tools more accessible, the system supports better planning, resource optimization, and educational engagement in agricultural research.

For future work, several directions can be pursued. First, expanding the dataset with **higher-resolution satellite imagery** and **soil condition data** would enhance model precision. Integrating **IoT sensor data** could enable real-time monitoring of crop growth and environmental factors. Additionally, incorporating **deep learning models**—such as Convolutional Neural Networks (CNNs) or LSTM architectures—may further improve long-term prediction accuracy. Finally, scaling the platform to a national or regional level could provide valuable insights for policymakers in designing agricultural strategies and ensuring food security.

In summary, the MLR Predictor demonstrates that even modest AI tools, when implemented with thoughtful design and accessible technology, can make a tangible impact on modernizing agriculture in Uzbekistan. This work serves as both a technical solution and a foundation for future innovations in smart, data-driven agricultural forecasting.

REFERENCES:

- [1] D.R.Xasanov, M.Tojiyev, O.D.Primqulov. Gradient descent in machine learning.// IEEE International Conference on Information Science and Communications Technologies: applications, trends and opportunities (ICISCT-2021). <https://ieeexplore.ieee.org/document/9670169>
- [2] Rakhimov N.O., Xasanov D.R. Selection of Crops by Region using Machine Learning Classification Algorithms. // *Ilm-fan va innovatsion rivojlanish jurnali*. Volume 7, Issue 2; April 2024, -p. 36-47. <https://cyberleninka.ru/article/n/selection-of-crops-by-region-using-machine-learning-classification-algorithms>
- [3] N.Rahimov, D.Khasanov. The Implementation of Machine Learning and Deep Learning Algorithms for Crop Yield Prediction in Agriculture. // *Bulletin of TUIT: Management and Communication Technologies*, Volume 2, No. 2023. <https://uzjurnal.uz/2/2023/2/index?issue=9>
- [4] N.Rakhimov, D.Khasanov, H.Abdulhakimov. Prediction the Yield of Grain Crops using Basic Machine Learning Algorithms. // *Raqamli transformatsiya va sun'iy intellekt ilmiy jurnali*, Volume 2, Issue 5. <https://dtai.tsue.uz/index.php/dtai/article/view/v2i57>
- [5] N.Raximov, J.Kuvandikov, D.Khasanov, "The importance of loss function in artificial intelligence", International Conference on Information Science and Communications Technologies (ICISCT 20222), DOI: 10.1109/ICISCT55600.2022.10146883
- [6] Rakhimov N.O., Khasanov D.R., Xafizadinov U.T. Application of the Algorithm for Enrichment the Knowledge Graph with Numerical Predicates in Decision-Making Support Systems. // *Digital Transformation and Artificial Intelligence: Problems, Innovations and Trends 1st International Scientific-Practical Conference*. <https://dtai.tsue.uz/index.php/DTAI2024/article/view/dilmurot2>
- [7] Rakhimov N.O., Khasanov D.R., Shirinboyev R.Sh., Xafizadinov U.T. Evaluating the Performance of Convolutional Neural Networks and Hybrid CNN-SVM Models for Symbol Recognition in Complex Datasets. // *Digital Transformation and Artificial Intelligence: Problems, Innovations and Trends 1st International Scientific-Practical Conference*. <https://dtai.tsue.uz/index.php/DTAI2024/article/view/shirinbayev2>
- [8] Rakhimov N.O., Khasanov D.R. The Implementation of Decision Tree Algorithm in Data Mining. // *International Scientific Journal Science and Innovation, Special Issue: "Digital Technologies: Problems and Solutions of Practical Implementation in the Spheres"*, 2023
- [9] N.O.Rakhimov, D.R.Khasanov. Selection of crops to be planted according to climate and soil characteristics with the help of machine learning classification algorithms. // *The 10th International Scientific and Practical Conference "Global and Regional Aspects of Sustainable Development"*,

Copenhagen-2024. <https://www.neliti.com/publications/628198/selection-of-crops-to-be-planted-according-to-climate-and-soil-characteristics-w>

[10] O.D.Primqulov, M.R.Tojiyev, D.R.Khasanov. Image segmentation in openCV and python. // ACADEMICIA. An International Multidisciplinary Research Journal (ISSN 2249-7137). Volume 10, Issue 12. December

[14] Rahimov Nodir, Khasanov Dilmurod. (2022). The Mathematical Essence Of Logistic Regression For Machine Learning. <https://doi.org/10.5281/zenodo.7239169>

[15] N.Raximov, O.Primqulov, B.Daminova, "Basic concepts and stages of research development on artificial intelligence", International Conference on Information Science and Communications Technologies (ICISCT), www.ieeexplore.ieee.org/document/9670085/metrics#metrics

[16] Ma'ruf Tojiyev, Ravshan Shirinboyev, Jahongirjon Bobolov. Image Segmentation By Otsu Method. International Journal of Contemporary Scientific and Technical Research, (Special Issue), 2023. 64–72, <https://zenodo.org/record/7630893>

[19] T. Maruf, "Hazard recognition system based on violation of the integrity of the field and changes in the intensity of illumination on the video image," in 2022 International Conference on Information Science and Communications Technologies, ICISCT 2022, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/ICISCT55600.2022.10146933.

[20] B. Z. Parmanov, F. M. Nazarov, and M. R. Tojiyev, "Development of an Algorithm for Intelligent Analysis of Data From a Microprocessor-Based Portable Spectrophotometer," in Proceedings - 2025 International Russian Smart Industry Conference, SmartIndustryCon 2025, Institute of Electrical and Electronics Engineers Inc., 2025, pp. 664–669. doi: 10.1109/SmartIndustryCon65166.2025.10985982.