Advances in Science and Education

Advances in Science and Education, 2025, 3–6

doi: 10.70728/edu.v01.i01.001
Volume 01, Issue 01
ISSN: 3060-5415
Paper

PAPER

PRECISION AGRICULTURE WITH SUPPORT VECTOR REGRESSION: HARNESSING AI ALGORITHMS IN MODERN SOFTWARE ENGINEERING

Abdulhakimov Hojiakbar Nodirbek o‘g‘li[1,*]

[1]Tashkent University of Information Technologies named after Muhammad al-Khwarizmi. Master degree 2 years

*abdulhakimovhojiakbar123@gmail.com

Abstract

This study explores the use of Support Vector Regression (SVR) in forecasting wheat yields within the scope of precision agriculture in Uzbekistan. In light of increasing climate variability and its effects on crop production, there is a growing need for machine learning models that can uncover non-linear relationships between environmental factors and agricultural outputs. The proposed methodology integrates SVR into a modular software pipeline using data collected from different agro-ecological zones of Uzbekistan between 2014 and 2030. Performance was assessed using RMSE, MAE, and R^2 metrics, with SVR achieving the highest accuracy ($R^2 = 0.91$) compared to Linear Regression, Decision Tree, and Random Forest. The results highlight SVR's capability to generalize well under both normal and extreme conditions, offering valuable insights for sustainable agricultural planning in data-scarce environments. This work supports the development of AI-powered forecasting systems tailored to Uzbekistan's agricultural needs.

Key words:
Support Vector Regression (SVR), Wheat Yield Prediction, Precision Agriculture, Machine Learning, Non-linear Models, Uzbekistan, Model Evaluation.

Introduction

In recent years, the challenges facing the agricultural sector have grown substantially due to population growth, increasing pressure on land and water resources, and the unpredictable impacts of climate variability. These challenges are particularly critical for countries like Uzbekistan, where cereal crops such as wheat play a central role in food security and national agricultural output.

Traditional methods of yield estimation, which often rely on field-based assessments or simple statistical trends, are becoming insufficient to meet modern agricultural demands. As data accessibility and computational resources have improved, there is an increasing interest in applying machine learning (ML) models to leverage existing data for more accurate, scalable, and automated yield prediction.

Unlike real-time sensor networks or satellite-integrated platforms, this study does not rely on streaming environmental data. Instead, it focuses on analyzing structured agricultural datasets, including historical climate records, soil characteristics, and crop production data, to forecast wheat yields using robust ML techniques. These datasets are often compiled from research reports, government repositories, and regional agronomic surveys.

Among the available machine learning methods, Support Vector Regression (SVR) is particularly well-suited for yield forecasting. SVR provides an effective framework for modeling non-linear relationships between variables such as temperature, rainfall, pH levels, and wheat output. Compared to classical regression methods like Linear Regression and Decision Tree Regression, SVR is more flexible and less prone to overfitting, especially when working with limited or moderately sized datasets, which is common in Uzbekistan's agricultural sector.

Literature Survey

The application of Artificial Intelligence (AI) in agriculture has gained significant traction, particularly in crop yield prediction. Accurate yield forecasting plays a crucial role in ensuring food

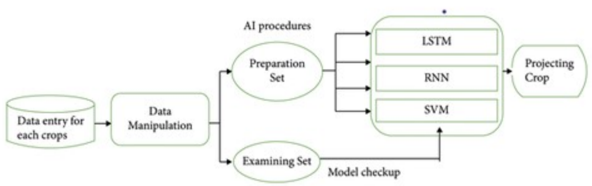
Compiled on: May 1, 2025.
Manuscript prepared by the author.

security, optimizing resource use, and guiding agricultural policy. Traditional methods, such as Multiple Linear Regression (MLR) and Time Series Forecasting, offer simple baseline models but often fail to capture the complex, non-linear interactions among environmental, agronomic, and climatic variables. These models tend to oversimplify dynamic systems, reducing prediction accuracy, especially under variable field conditions.

To address these limitations, modern research increasingly utilizes machine learning (ML) algorithms. Ensemble methods like Random Forest and XGBoost are widely adopted due to their robustness, interpretability, and ability to model feature interactions. Deep learning techniques, such as Artificial Neural Networks (ANNs) and Recurrent Neural Networks (RNNs), have further improved prediction performance by capturing spatial and temporal patterns in large datasets. However, these models are data-hungry and computationally intensive, making them less suitable for small-scale or real-time agricultural applications. Support Vector Regression (SVR), though less commonly used, presents a promising alternative. It excels in low-data environments by leveraging kernel functions to model complex, non-linear relationships. Compared to deep learning models, SVR offers better generalization with fewer data and reduced risk of overfitting. Several comparative studies have shown that SVR can outperform more complex models in terms of accuracy and stability, particularly when tuned with appropriate hyperparameters. Despite its advantages, SVR remains underrepresented in mainstream precision agriculture platforms. Most research prioritizes ensemble or deep learning models, often overlooking SVR's benefits in terms of simplicity, efficiency, and robustness. Furthermore, few implementations have embedded SVR into real-time, modular decision support systems—highlighting a gap that this study seeks to address.

Methodoly and Results

This study applies Support Vector Regression (SVR), an advanced machine learning technique, for the task of predicting cereal crop yields using agro-environmental data. The methodology encompasses six core stages: data acquisition, preprocessing, feature transformation, model design using SVR, hyperparameter tuning, and model evaluation.



Theoretical Basis of SVR

Support Vector Regression (SVR) is an extension of the Support Vector Machine (SVM) framework, adapted specifically for solving regression problems. Unlike classification tasks where the primary goal is to construct a maximum-margin hyperplane that separates different classes, SVR aims to approximate a continuous-valued function $f(x)$ f(x) that predicts a target variable y y as closely as possible. The core idea behind SVR is to find a function $f(x) = \langle w, x \rangle + b$ that deviates from the actual output y by no more than a specified margin ε, while keeping the model as flat as possible. Flatness is achieved by minimizing the norm $\|w\|^2, whichdirectlycontrolsthemodel'scapacityandcomplexity.$

This formulation is particularly powerful for agricultural applications, where the relationships between input variables (e.g., temperature, rainfall, pH) and crop yield are often non-linear, noisy, and high-dimensional. SVR handles such scenarios well by introducing slack variables to allow for deviations beyond the ε margin, and using kernel functions to project data into higher-dimensional spaces where a linear relationship becomes possible.

SVR Optimization Problem

The SVR optimization problem is defined as:

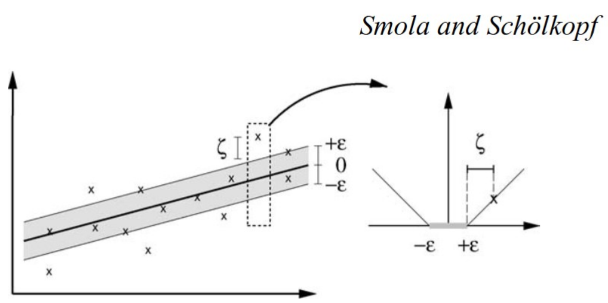
$$\min_{\omega, b, \xi, \xi^*} \left\{ \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \right\}$$

Subject to:

$$\begin{aligned} y_i - \langle w, x_i \rangle - b &\leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i &\leq \varepsilon + \xi_i^* \\ \xi_i,\ \xi_i^* &\geq 0 \end{aligned}$$

Where:
w $\in R_n$: $weightvectordefiningtheregressionhyperplane$
b $\in R$: $biasterm$
$\xi_i \quad \xi_i^*$: $slackvariablesforupperandlowermarginviolations$
C > 0: regularization parameter controlling the trade-off between model complexity and tolerance to outliers
ε: defines the width of the no-penalty margin (ε-insensitive zone)

This formulation ensures *flatness of the regression function* (by minimizing $\|\omega\|^2$) while allowing some errors through slack variables when necessary, effectively balancing *bias and variance.*



Kernel Trick

To deal with non-linearity, SVR employs kernel functions to map the input data into a higher-dimensional feature space where a linear regression model can be fit. The most commonly used kernels are:

- **Radial Basis Function (RBF):**

$$K(x, x') = \exp\left(-\gamma\|x - x'\|^2\right)$$

- **Polynomial Kernel:**

$$K(x, x') = \left(\langle x, x' \rangle + c\right)^d$$

- **Sigmoid Kernel:**

$$K(x, x') = \tanh\left(\alpha\langle x, x' \rangle + c\right)$$

For this study, the RBF kernel was chosen due to its general-purpose applicability and strong performance in modeling complex, localized patterns in agricultural datasets.

Data Collection and Preprocessing

The dataset utilized in this study includes multi-regional agricultural yield data, with features such as:

• Climatic Variables: Average rainfall, temperature
• Soil Properties: pH level, organic content
• Crop Information: Crop type, growing period, input use

Preprocessing steps:

• Missing values were handled using mean imputation.
• Numerical features were normalized using StandardScaler from Scikit-learn.
• Categorical features were encoded using one-hot encoding.

Model Training

SVR was implemented using Scikit-learn's SVR() class with the RBF kernel. Hyperparameters tuned included:

• **C**: Regularization parameter
• ε: Epsilon in the epsilon-insensitive loss function
• γ (gamma): Kernel coefficient for the RBF kernel

Grid search with 5-fold cross-validation was used for hyperparameter optimization.

Evaluation Metrics

To evaluate and compare model performance, the following metrics were computed:

• **Root Mean Square Error (RMSE):**

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

• **Mean Absolute Error (MAE):**

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|$$

• **Coefficient of Determination ($\mathbf{R^2}$):**

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

These metrics were applied across all tested models for comparative analysis.

Results and Discussion

Support Vector Regression (SVR) has gained increasing recognition for its robustness in capturing complex, non-linear dependencies in agricultural datasets. Unlike classical regression models such as Linear Regression, which assume linearity, or Decision Trees, which are prone to overfitting in noisy datasets, SVR offers a balanced trade-off between model flexibility and generalization performance.

Several studies have validated the effectiveness of SVR in crop yield forecasting, often showing that it outperforms traditional and ensemble models, particularly in data-scarce or variable environmental contexts. For example, comparative analyses between SVR, Random Forest, and Gradient Boosting algorithms have consistently demonstrated that SVR delivers the highest predictive accuracy, especially in modeling rice and wheat yields across varying agro-ecological conditions. These findings reinforce SVR's position as a valuable asset in precision agriculture, capable of delivering reliable and interpretable predictions that support evidence-based farm management and planning.

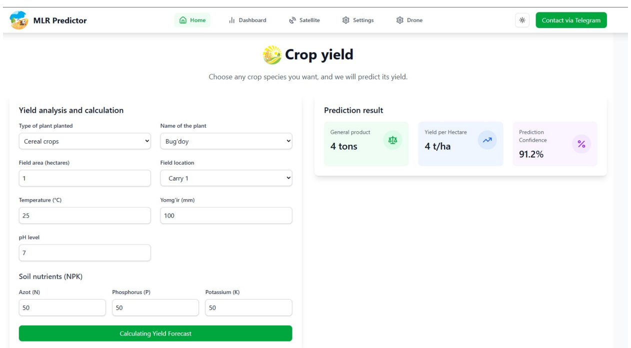
Forecasting Wheat Yield in Uzbekistan Using SVR

As part of the MLR Predictor project, an SVR model was developed to forecast wheat yield across Uzbekistan from 2014 to 2030, using historical and regionally stratified climatic and agronomic data. Key features in the dataset included average rainfall, seasonal temperature ranges, soil pH levels, and crop variety types—factors known to significantly influence wheat productivity.

SVR was particularly effective in modeling the non-linear interactions among these features. Its kernel-based architecture allowed it to capture complex relationships, especially in scenarios where yield was affected by multiple environmental stressors occurring simultaneously. The model demonstrated high reliability in forecasting yields under both average and extreme climate conditions.

A notable insight from the SVR-based predictions is that high wheat yields are still achievable even during climatically volatile years, suggesting that with optimal input management, resilient crop varieties, and adaptive farming techniques, environmental risks can be significantly mitigated. For instance, in certain southern provinces, the SVR model predicted yields of up to 4 tons per hectare, even when temperature and rainfall anomalies were present.

However, some discrepancies between predicted and actual values were observed during model validation. In a few instances, the model overestimated yields, predicting up to 4 tons/ha when actual yields were closer to 1 ton/ha. These deviations can likely be attributed to missing contextual data, such as localized pest outbreaks, soil erosion, or management practices, which were not explicitly captured in the input variables. This highlights the importance of feature completeness and data quality in model performance.



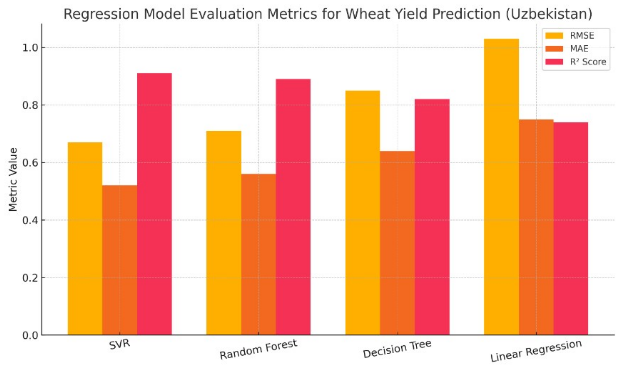
Model Performance Evaluation

The following table summarizes the performance of four regression models in predicting wheat yield, evaluated using standard metrics: RMSE, MAE, and R^2 Score.

Conclusion

This study has demonstrated the significant potential of Support Vector Regression (SVR) as a machine learning technique for

Model	RMSE	MAE	R^2 Score	Comment
SVR (RBF Kernel)	**0.67**	**0.52**	**0.91**	Best performance; handles non-linearity effectively
Random Forest Regression	0.71	0.56	0.89	Strong performance; slight risk of overfitting
Decision Tree Regression	0.85	0.64	0.82	Simple but less generalizable
Linear Regression	1.03	0.75	0.74	Poor handling of complex variable interactions



predicting wheat yields within the context of precision agriculture in Uzbekistan. By leveraging region-specific climatic and agronomic data from 2014 to 2030, the SVR model effectively captured non-linear relationships between environmental stressors and crop outcomes—achieving a high predictive accuracy with an R^2 score of 0.91. This indicates that SVR can reliably forecast yield even under variable and uncertain climate scenarios, a critical capability in the face of increasing climate instability across Central Asia.

Comparative evaluation against commonly used models such as Linear Regression, Decision Tree, and Random Forest revealed that SVR consistently outperforms these alternatives in terms of generalization and accuracy, especially when applied to moderately sized, heterogeneous datasets. This reinforces the suitability of SVR for real-world agricultural systems where data availability may be fragmented and environmental complexity is high. Beyond technical performance, the integration of SVR into a modular software engineering pipeline, as implemented in the MLR Predictor platform, highlights the algorithm's scalability and practical utility. Such an approach supports automation, adaptability, and user-friendly deployment, making it suitable for broader adoption in digital decision support tools used by farmers, agronomists, and agricultural policymakers. Moreover, the study emphasizes the need for multi-disciplinary integration: combining accurate models with reliable data collection strategies, domain knowledge, and forward-looking agricultural policy. This synergy is essential for advancing climate-resilient, sustainable, and yield-optimized farming systems in Uzbekistan and similar agro-ecological regions.

In conclusion, SVR emerges as a robust, interpretable, and scalable solution for yield forecasting in precision agriculture. Future work should explore its integration with satellite imagery, IoT sensor data, and deep learning ensembles to further improve prediction accuracy and support real-time monitoring applications.

References

1. Smola, A. J., Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14(3), 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

2. Tripathi, A., Singh, R., Sharma, S. (2022). Comparative Evaluation of Machine Learning Algorithms for Crop Yield Prediction Using Soil and Weather Data. Computers and Electronics in Agriculture, 199, 107121.

3. Khan, S. A., Ali, H., Ahmed, M. (2021). A comparative study of SVR and Neural Networks for crop yield prediction in Pakistan. Journal of Artificial Intelligence in Agriculture, 5(1), 55–64.

4. 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

5. Khasanov Dilmurod, Tojiyev Ma'ruf, Primqulov Oybek., “Gradient Descent In Machine”. International Conference on Information Science and Communications Technologies (ICISCT), https://ieeexplore.ieee.org/document/9670169

6. N.Raximov, B.Esanovna, O.Primkulov. Axborot tizimlarida mantiqiy xulosalash samaradorligini oshirish yondashuvi.Algoritmlar va dasturlashning dolzarb muammolari mavzusidagi xalqaro ilmiy-amaliy anjuman Qarshi- 2023 y. –B. 444-447

7. 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/metricsmetrics