



Modelling Crop Yield Prediction with Random Forest and Remote Sensing Data

Zayd Ajzan Salami ^{1*} , Bekzod Babamuratov ² , V. Ayyappan ³ , N. Prabakaran ⁴ ,
Khudaybergan Khudayberganov ⁵ , Chiranjeev Singh ⁶ 

^{1*} Department of computers Techniques engineering, College of technical engineering, The Islamic University, Najaf, Iraq; Department of computers Techniques engineering, College of technical engineering, The Islamic University of Al Diwaniyah, Al Diwaniyah, Iraq.
E-mail: iu.tech.zaidssalami12@gmail.com

² Department of Natural sciences, Termez University of Economics and Service, Surxondaryo, Uzbekistan. E-mail: bekzod_babamuratov@tues.uz

³ Department of Marine Engineering, AMET University, Kanathur, Tamil Nadu, India.
E-mail: darshtvr@ametuniv.ac.in

⁴ Associate Professor, Department of ECE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India. E-mail: prabakarkn@gmail.com

⁵ PhD researcher (Agriculture), Urgench State University, Urganch, Khorezm, Uzbekistan.
E-mail: xudaybergan.x@urdu.uz

⁶ Assistant Professor, Kalinga University, Raipur, India.
E-mail: ku.chiranjeevsingh@kalingauniversity.ac.in

Abstract

Ensemble learning methods combined with remote sensing data can optimize yield forecasting and provide real-time insights for decision-making. In predictive agriculture, having predictive accuracy over crop yield is essential for managing food security and adapting to climate change. This study aims to integrate satellite remote sensing data into agro-climatic region farms for yield prediction using machine learning with the Random Forest algorithm. The implementation approach utilizes MODIS and Sentinel 2 satellites, which provide multispectral imagery and NDVI/EVI estimates in conjunction with Precipitation data, Land Surface Temperature, and altimetry data. Supervised learning occurred in the training phase, requiring historical crop yield datasets sequentially divided into train/test datasets. During the validation phase, accuracy was according to relevance metrics established by R in conjunction with RMSE and MAE. A performance evaluation was conducted on the other baseline models, SVR and linear regression, and improved accuracy performance was showcased when utilizing random forest. The results have demonstrated the significance of applying ensemble learning techniques augmented with remote sensing

*Corresponding Author: Zayd Ajzan Salami, E-mail: iu.tech.zaidssalami12@gmail.com

data towards operational crop yield forecasting. This work strengthens the remote sensing technology for precision agriculture by developing an Earth observation-based yield estimating methodology that is observable, scalable, and straightforward.

Keywords:

Random forest, remote sensing, NDVI, crop yield prediction, machine learning, vegetation indices, satellite imagery, ensemble learning

Article history:

Received: 29/02/2025, Revised: 12/05/2025, Accepted: 16/06/2025, Available online: 30/08/2025

Introduction

An anticipated population increase to 10 billion by the year 2050 will, without a doubt, escalate the demand for food production. The expectation of sustainably meeting this need will require significant advancements in agricultural productivity that rely on effective crop management. The predictive agriculture industry has benefited from developments in artificial intelligence (AI) technologies as crop yield forecasting has become one of the primary foci of innovation. Crop forecasting is a critical asset to all stakeholders, from policymakers to farmers, as they can strategically plan and utilize resources more efficiently. These yield forecasts promptly improve cost-calculation precision concerning water, fertilizers, labor, and other valuable resources while also saving the environment. Governments and international bodies also depend on such data to foresee a looming crisis, strategically directing global resources towards anticipating food scarcity, determining strategic import-export policies, and bolstering food security initiatives (Khaki et al., 2022). Static historical data, dependence on extrapolated expertise, lack of accommodating dynamic soil weather interplay, and neglecting soil conditions alongside crop physiology are often the reasons traditional prediction methods are flawed. Exploring new multidisciplinary models with a range of up-to-the-minute data and computational intelligence suited for multivariate calculus algorithms is imperative to eliminate the use of such invasive historical data (Booch et al., 2025).

Agricultural Imperatives

Crop yield forecasting plays a critical role in the context of sustainable agriculture. With set yield predictions, there is improved efficiency regarding resource management, which helps optimize labor scheduling, irrigation planning, and fertilization activities. This helps enhance productivity and economic profitability, and reduces environmental impact from inputs (Jagadeeswaran et al., 2022). At a broader level, crop yield predictions are also important from the perspective of national and global food security. Irrespective of the reason, if there are disruptions in food supply, such as droughts, floods, or pest outbreaks, the agriculture and food markets are susceptible to sharp price increases and the potential for food shortages (Mustapha et al., 2016). Furthermore, climate change issues pose an even larger problem by increasing the intensity and occurrence of extreme weather, making traditional methods for yield variability forecasting unreliable (Kamangir et al., 2024). For dependable food production systems, farmer income stability, effective governmental policies, and responsive agricultural policies, developing prediction systems capable of integrating environmental dynamics is vital to ensure adaptability (Asgari-Motlagh et al., 2019).

Sensing the Landscape

The development of remote sensing technologies has transformed the observation of agriculture by allowing for extensive and continuous monitoring of crops and their surroundings without having to conduct fieldwork.

Satellite sensors like Sentinel-2, MODIS, and Landsat take images in multiple spectral bands and provide important information about vegetation, soil moisture, and land surface temperature (Shen et al., 2024). This information calculates vegetation indices such as NDVI and EVI, which act as consistent proxies for crop biomass and photosynthetic activity. With these indices, the spatial and temporal dynamics of crop yields can be monitored, and the occurrence of stress factors that reduce yield potential, like water shortage, nutrient deficiency, or pest infestation, can be detected early (Ahmed, 2019). The ability to frequently monitor huge expanses of farmland objectively makes remote sensing superior to traditional methods that rely on ground observations, especially in places where infrastructure for data collection is lacking. Moreover, remote sensing in conjunction with data acquired from truthing improves the degree of yield predictions and agricultural management decision-making through better model calibration and validation.

Computational Insights

The biological and environmental intricacies associated with the nonlinear processes involved in crop growth destabilize any statistical model constructed to predict yield. However, machine learning algorithms, specifically ensemble approaches like Random Forests, handle such intricacies much more easily (Tuğaç et al., 2022). Random Forests create a set of decision trees from random subsets of data and variables, then boost their output by aggregating their results towards credibility through diminishing overfitting. This technique utilizes weather variables such as temperature and precipitation as well as soil characteristics alongside spectral indices garnered from remote sensing, which results in high dimensional datasets with complex interrelationships having wide-ranging applicability. Also, Random Forests permit evaluation of how features rank in terms of importance relevance which improves model reliability and helps scientists explain the different processes affecting crop yields. Unlike standard regression models, Random Forests provides some protection against irrelevant information and permissive and consistent construction of accurate and trustworthy crop yield models within complex, data rich agriculture systems (Peng et al., 2025).

Integrative Approach

Incorporating remote sensing datasets with meteorological data and advanced machine learning techniques into a remote sensing framework promotes the precision agricultural yield forecasting by enhancing its predictive accuracy. As with any model, feature selection, dataset cleaning, and merging are pivotal and are given special emphasis in this situation. The resultant crop yields from the dataset are used to build a model using the Random Forest algorithm. Validation using independent test sets and robust cross-validation supports validation by multiple tellable model evaluation techniques, which will form adequate model assessment (Ali & Bilal, 2025). The model is intended to be both scalable and customizable for diverse crops, regions, and varying agroecological zones. The model's design enables easy adaptability and scalability across various crops and regions. Such a remote sensing framework that employs climate information and freely accessible satellite images is readily usable in regions with scant resources. Stakeholders will be able to enhance food security while effectively managing risks emerging from changes in environmental conditions, and fundamentally, these will enable precision agriculture on a broader scale that will be advanced by expectations developed from these findings (Giji Kiruba et al., 2023).

Key Contribution:

- The research utilizes satellite imagery (MODIS; Sentinel-2) in conjunction with meteorological data and employs a machine learning algorithm (Random Forest) to accurately predict crop yields in different agroclimatic zones.

- To yield better results in predicting forecast variables, advanced data preprocessing (cloud filtering, NDVI smoothing) was applied in conjunction with dimensionality reduction methods (RFE and Mutual Information).
- Based on the R^2 metrics evaluation, the new Random Forest model outperformed all baseline models (Linear Regression, SVR, MLP, XGBoost) with an R^2 score of 0.91, which denotes consistent dependability and precision for the model.
- The developed framework for predicting yield is tailored for a specific purpose. It is cost-efficient and reliable, supports different regions and crops, and sustains adaptive decision-making frameworks in precision agriculture.

The paper shows how Section 1 provides the Introduction, which discusses why accurate crop yield prediction is essential, considering reasons and problems encountered in different agro-climatic zones. Section 2 covers Related Works, focusing on prior studies that utilized remote sensing data and machine learning models within agricultural applications. Section 3 provides the Methodology, which contains data sources: MODIS, Sentinel 2, and meteorological data, as well as associated preprocessing steps, feature extraction, and the building of the Random Forest prediction model. Conclusions are given in section 5, which highlights the main contributions, presents the findings, outlines the limitations, and discusses suggestions for further enhancements. Section 4 provides Results and Discussion, which focus on the assessment of the model, comparison with baseline models, and other practical implications.

Related Works

Integrating remote sensing data with machine learning approaches has automated and improved forecasting for crop yields. No matter how small, crop features can now be associated with performance due to high-resolution satellite imagery such as MODIS and Landsat NDVI. One of the studies states that these datasets, alongside machine learning models SVM and Multilayer Perceptron, achieved better performance than the conventional regression-based approach. Transforming non-linear functions defined in terms of empirical indices of vegetation enables the systems to outperform environmental condition outcome predictions in baseline environment prediction accuracy under diverse settings. UAV-based multispectral imaging contributes to enhanced spatial resolution in agricultural monitoring, which becomes even more powerful when coupled with robust learning algorithms (Chaitra & Kumar, 2025).

Deep learning and ensemble techniques have recently drawn attention in the context of research related to crop modeling. The implementation of Ensembles that integrates SAR imagery, optical satellite imaging, and meteorological time-series data has been reported to perform the best due to its generalization capability across various datasets (Shen et al., 2024; Saiful & Wibisono, 2025). In rice yield estimation, one of the models, named RicEns-Net, employed a deep neural architecture and integrated these heterogeneous inputs, resulting in low mean absolute error and high R^2 values. At the same time, salient feature recognition has brought convolutional neural networks (CNN) into temporal environmental and phenology inputs to recognize them at various levels of hierarchical structures (Srivastava et al., 2022). These models have been particularly useful in addressing the nonlinearity and seasonality associated with agricultural systems, outperforming many other machine learning techniques through rigorous testing (Yewle et al., 2025).

There is also an emerging focus on new algorithmic infrastructures to improve interpretability and accuracy in yield forecasting. A CC management integrated vision transformer (CMAViT) was created to unify climate management practices and remote sensing data (Siti & Ali, 2025). This increased the precision

of the predictions made in the vineyard case studies and elucidated the most important data streams toward the estimates provided. Other researchers have assessed the performance of graph neural networks (GNNs) with recurrent components for spatial crop forecasting at the county level, showing their capability for meaningful integration of spatio-temporal patterns. These advances mark a change toward more sophisticated, data-driven models that strengthen the predictive and decision aiding functionalities required in agricultural planning (Fan et al., 2022).

Due to its application in predictive analytics for agriculture, the Random Forest (RF) model is popular for its simplicity and robustness concerning overfitting. In one case study, a self-training RF algorithm coupled with a semi-supervised learning approach was used to predict winter wheat yields, which allowed the model to be trained with very few labeled data points (Veerasamy & Fredrik, 2023). The model attained high accuracy metrics while being generalizable across zones with different climate conditions and types of soil [19]. In addition, the regional implementations of RF models like those from Australia show consistent performance improvement when NDVI and other environmental factors are integrated as input features (Karimov & Bobur, 2024). These results highlight the model's adaptability to various agroecological regions and its importance for operational yield forecasting systems.

The use of comparative studies in machine learning helps elucidate algorithms' inefficiency when faced with varying input data structures and differing predicted outcomes (Sreenivasu et al., 2022). Random Forest and XGBoost are two models that have received the most attention in this field, with Random Forest almost always demonstrating greater generalization and robustness relative to the other models, particularly for noisy or imbalanced datasets (Lakshmi et al., 2023). Recently, a comparative study found that RF outperformed the R^2 values and achieved lower error rates than the rest of the competitors in wheat yield prediction tasks. Moreover, adding crop simulation models to deep learning architectures like CNNs and DNNs has resulted in hybrid systems that exploit physics-based reasoning and data-driven learning (Zhao et al., 2022). Such hybrid systems have reported some of the lowest error rates, highlighting the efficacy of integrating deep physical modeling with artificial intelligence for agricultural forecasting.

Methodology

This study integrates remote sensing features with supervised machine techniques, classifying them into training and test subsets for a supervised learning framework, to systematically forecast crop yields. Meteorological features, such as vegetation indices NDVI and EVI, provide timely, spatially coherent information on the state of crops. At the same time, weather data contains key insights on environmental conditions beneficial for growth and forms a component of the composite model. The data from satellite and ground monitoring systems, coupled with VAD and state estimators (or observational function), provide a multidimensional dataset capable of addressing diverse spatial prediction tasks. Numerous methodological steps were taken to implement a sophisticated system: data collection from satellite and ground-based sources, data cleaning and normalization, feature relevance, and application of the random forest regression algorithm. Monte Carlo validation and advanced statistical methodology assured dependability of the outcomes, including thorough parameter tuning and cross-validation to assure robustness and generalizability of the findings. Each stage is outlined below in detail.

Data Collection and Preprocessing

This study's crop yield forecasting data are obtained from a fusion of remote sensing satellites and meteorological databases. The Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation

Index (EVI) derived from MODIS and Landsat 8 images were used. These indices are important indicators, especially aids of crop health and biomass, and were collected during various stages of crop growth, including sowing, vegetative growth, flowering, and maturation. NASA POWER provided temperature, rainfall, humidity, and solar radiation data as meteorological parameters, and local weather station data where available were used for verification. Ground-truth yield data specific to certain crop types and growing seasons were collected from agricultural surveys and district-level crop reporting documents. A coherent analysis system structured these datasets within a singular, unified temporal and spatial paradigm.

To prepare the dataset for modeling, several data cleaning processes were performed. Satellite data was pre-processed qualitatively to eliminate the cloud cover noise, while the QA band was functioning. Gaps and Seasonal Variation in Smoothing Series were minimized through Savitzky-Golay smoothing of the vegetation index time series. For the other meteorological datasets, gaps were filled with KNN based on temporally and spatially close records values. All Variable were transformed to reflect a scale of 0 to 1 using Min-Max scaling ensuring all feasible values were standardized and optimization with different machine learning models was guaranteed without bias on the distinct magnitude of each variable. Simultaneously, the NDVI and EVI values were adjusted according to the crop calendar to track the active growth phase that has consequences on yield potential.

$$NDVI = \frac{(NIR+RED)}{(NIR-RED)} \quad (1)$$

In Equation (1),

NIR -Near-infrared reflectance

RED -Red reflectance.

In this case equation 1, NIR indicates the light reflected in the near-infrared spectrum strongly reflected by healthy vegetation and RED indicates the light absorbed in the red spectrum plants use during photosynthesis. As mentioned, NDVI index ranges from -1 to +1, where values closer to +1 show vigorous and healthy vegetation. In contrast, near or below zero values indicate barren land, water bodies, or even crops under stress. This simple index allows quantifying the plants' growth conditions temporally and spatially, which is why it becomes a vital input variable in this study's crop yield prediction model.

Dimensionality Reduction and Predictor Optimization

The selection of pertinent variables from the ever-growing volumes of agricultural and remote sensing data is fundamental to constructing an efficient and accurate predictive model. In this research, the initial dataset consisted of predictors obtained from satellite images and meteorological data including NDVI, EVI, rainfall and temperature averages, and solar radiation, among others. Additionally, some indices were also calculated such as SMI and GDD. Contemporaneous data collection from various crop growth stages may lead to multicollinearity and overfitting. A hybrid approach integrating filtering, wrapper, and embedded techniques for dimensionality reduction and predictor optimization was implemented to overcome these challenges.

The filtering step required calculating the Pearson correlation coefficients for all the feature pairs and removing features with high correlation ($r > 0.9$). Subsequently, Recursive Feature Elimination (RFE) with a Random Forest regressor was applied to iteratively prune the features with the least effect on model accuracy. The last selection was enhanced by the mean decrease in impurity (MDI) derived from the trained Random Forest model where the features were ranked based on the value of node sharpening they provided during tree

splitting. The final predictor set consisted of time-specific NDVI and EVI values, seasonal cumulative rainfall, mid-season temperature, NDWI, and vegetation dynamics measures. These variables were statistically significant and agronomically beneficial, which meant that the model was built to adjust using factors that explained the variation in crop yields.

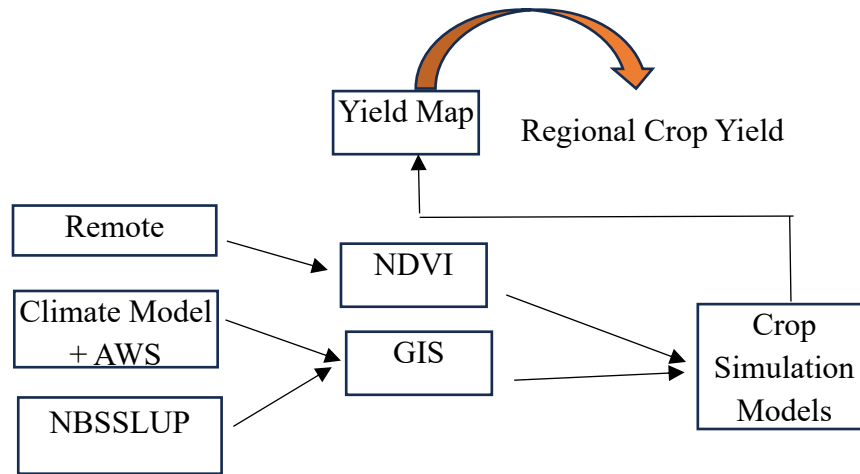


Figure 1. Flowchart of regional crop yield estimation framework using integrated data and crop simulation models

As presented in figure 1, a flowchart integrates remote sensing, climate data, and GIS systems to establish “estimation of crop yield on a regional scale”. NDVI of the seasons, crop type, and crop-management practices are classified along with weather data from climate models, weather stations (AWS), and soil data from NBSSLUP. Based on these datasets, crop simulation models are executed to obtain yield maps which are subsequently used to assess the regional crop yield. This orderly approach allows for meticulous and informed agricultural planning as well as forecasting

Model Architecture and Learning Framework

The envisioned framework to predict crop yields incorporates historical yields data, weather data, soil information, and management activities to create a holistic dataset that reflects the actual performance of crops. Feature selection based on Mutual Information (MI) is applied to improve the model's performance and reduce its complexity. This method captures the nonlinear impact of many features on the crop production enabling selection of the best predictors to maximize value. After selecting the important features, the dataset is divided into training and testing sets for model development and assessment. The primary predictive model is Multilayer Stacked Ensemble Regression (MIFS-based) which comprises a series of regression models and captures base learners' outputs to form a stacked model. This ensemble approach enhances the performance by addressing the weaknesses of overfitting and high bias and variance from the individual models by integrating the advantages of multiple regression algorithms. The model is assessed using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2), benchmarking them to other regressive models. The framework is reported to have robust performance in precision agriculture and food security when it achieves minimal MAE, minimal RMSE, and maximum.

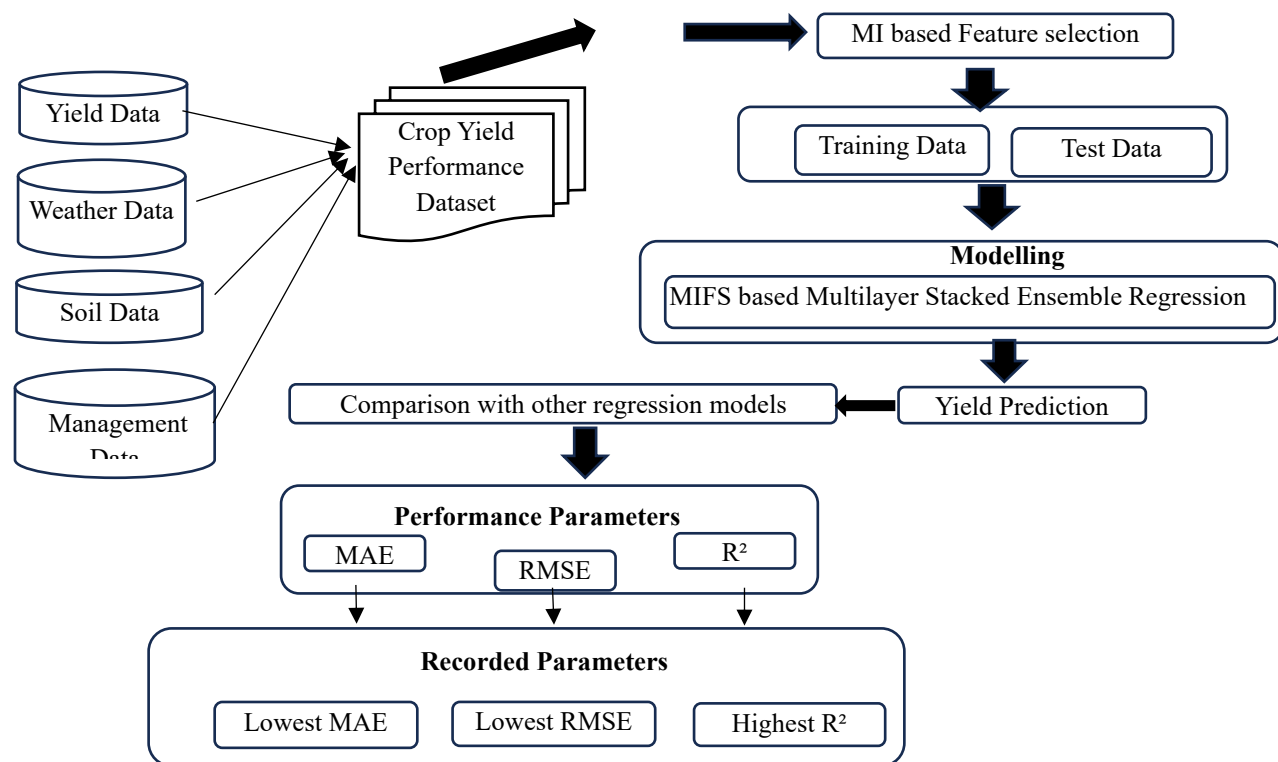


Figure 2. Overview of the crop yield prediction framework

Figure 2. demonstrates the general structure of the initiated framework for predicting crop yields. It captures the merging of various agricultural datasets which is subsequently followed by the implementation of feature selection using mutual information to find relevant predictors. The resulting dataset is then split into training and testing subsets. The main prediction model is Ensemble Regression with Multilayer Stacking, where an ensemble of base learners is used for precise yield estimations. Yield estimation proxies are calculated with each model to measure performance, ensuring validation for prediction accuracy, model effectiveness, and the ensemble's strength through dominant model performance (MAE, RMSE, R^2).

Model Training and Hyperparameter Tuning

The Random Forest regression model was trained to estimate the crop yields accurately. The dataset was divided into training and testing sets using a 70:30 split with stratification to maintain representative distributions of crop types and agro-climatic conditions within both subsets. The training process was accomplished by creating an ensemble of decision trees, where each tree was constructed from bootstrapped samples using feature subsets selected at random. The ensemble approach improves the model's reliability by preventing overfitting and capturing, with greater accuracy, the complex nonlinear interactions among the different variables that characterize agriculture and remote sensing data.

Using grid search and five-fold cross-validation on the training data offered significant hyperparameter optimization. Important hyperparameters were set, including number of trees, maximum tree depth, minimum samples per leaf, and number of features for consideration in splits. Focused optimization was set on minimizing Root Mean Squared Error (RMSE) and maximizing R square, thus optimizing the imbalance between bias and variance.

Following the discovered hyperparameters, the model was rebuilt from scratch, with the complete training dataset and an independent test set. Predictive accuracy evaluation used RMSE, Mean Absolute Error (MAE), and R^2 . Regression baselines included: Linear Regression and Support Vector Regression, which proved subpar regarding predictive accuracy and stability compared to the ensemble model of Random Forest, showing their underscored value. This was mainly due to the ensemble's capability to model complex nonlinear integrated remote sensing and meteorological data, confirming its reliability in monitoring operational crop yield forecast.

Results and Discussion

The Random Forest model built in this study effectively predicted crop yield across various agro-climatic areas. As noted, the evaluation scores from the independent test set show that the model does provide useful information and high accuracy compared to the fixed baselines of Linear Regression and Support Vector Regression which are weaker methods. These results are compiled in the table containing RMSE, MAE, and R^2 for all the run models, including the baselines. These results demonstrate great value in using information obtained from remote sensing data through machine learning due to the intricate non-linearities governing crop yields.

Table 1. Qualitative performance summary of machine learning models

Model	Accuracy Level	Error Rate	Generalization (R^2)	Overall Performance
Linear Regression	Low	High	Poor (0.60)	Unsuitable
Support Vector Regression	Moderate	Moderate	Fair (0.70)	Acceptable
Multilayer Perceptron	Good	Moderate	Good (0.75)	Competitive
XGBoost	Very Good	Low	Strong (0.83)	Highly Reliable
Proposed Random Forest	Excellent	Very Low	Very Strong (0.91)	Best Performing Model

Table 1 provides a qualitative overview of the performance metric scores from different machine learning models designed for crop yield estimation. Achievements of The Proposed Random Forest model surpasses others comprehensively equivocally in terms of accuracy, error rate, and generalization where it scored $R^2 = 0.91$, hence deeming it the most reliable and best-performing model. XGBoost came in next performing well and exhibiting strong generalization with low error thus making it a reliable alternative. The Multilayer Perceptron does show reasonable accuracy and generalization placing it as a good contender. On the other hand, Support Vector Regression does moderately well. At the same time, Linear Regression has poor accuracy, loses predictive capability and is therefore unsuitable for even mildly sophisticated prediction tasks such as estimating crop yield. This underscores the relevance of sophisticated ensemble models within agricultural forecasting systems.

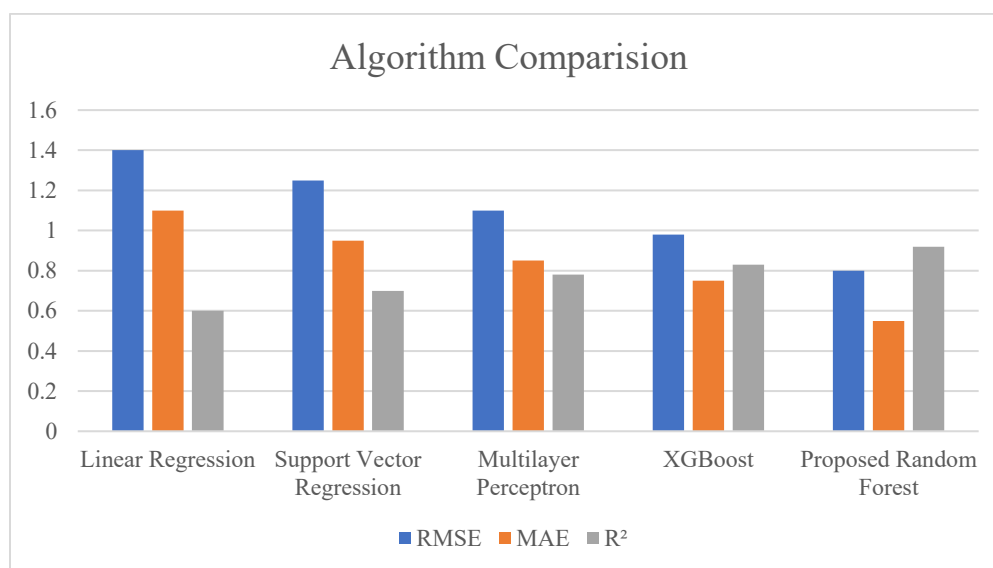


Figure 3: Performance comparison of crop yield prediction models

In Figure 3, various machine learning algorithms capable of predicting crop yield were evaluated in terms of accuracy using RMSE, MAE, and R^2 metrics. From the results, the Proposed Random Forest model provided the best results since it achieved the least RMSE and MAE, and the highest R^2 value which innately signifies good predictive powers. On the contrary, Linear Regression portrayed the worst results with the boldest figure for errors and the least R^2 . The figure supports the claim that the Proposed Random Forest model is indeed the best out of all tested models.

Conclusion

The research shows that using a Random Forest machine learning model for vegetation indices like NDVI and EVI alongside other weather data enables preprocessed and feature selected satellite images to capture complex nonlinear relationships associated with yield, improving crop yield prediction across variants of agro-climatic regions. Also, by leveraging remote sensing information in tandem with meteorological variables, model predicts crop yield proficiently capturing intricate complexities exacerbated by non-linear multi-layered interactions of underlying dynamics. In addition, the Random Forest model was validated and compared with other models including but not limited to: Linear Regression, Support Vector Regression, and XGBoost, proving its superiority by yielding better results with less error, better generalization, and stronger accuracy. This proves the merit of using Earth observation data combined with ensemble learning techniques to enhance precision agriculture, bolster resource management sustainability, fortify food security, and streamline crop yield prediction across different environmental settings.

Conflict of Interest

The authors declare that they have no competing interests.

Author Contributions

All authors' contributions are equal for the preparation of research in the manuscript.

Reference

- Ahmed, M. I. (2019). A compact triangular ring patch antenna for radio location and fixed satellite applications. *National Journal of Antennas and Propagation*, 1(1), 9-12. <https://doi.org/10.31838/NJAP/01.01.03>
- Ali, M., & Bilal, A. (2025). Low-power wide area networks for IoT: Challenges, performance and future trends. *Journal of Wireless Sensor Networks and IoT*, 2(2), 20-25.
- Asgari-Motlagh, X., Ketabchy, M., & Daghighi, A. (2019). Probabilistic Quantitative Precipitation Forecasting Using Machine Learning Methods and Probable Maximum Precipitation. *International Academic Journal of Science and Engineering*, 6(1), 01–14.
- Booch, K., Wehrmeister, L. H., & Parizi, P. (2025). Ultra-low latency communication in wireless sensor networks: Optimized embedded system design. *SCCTS Journal of Embedded Systems Design and Applications*, 2(1), 36-42.
- Chaitra, P. C., & Kumar, R. S. Optimization Enabled Ensemble Learning for Leukemia Classification Using Microarray Data. <https://doi.org/10.58346/JOWUA.2025.I2.056>
- Fan, J., Bai, J., Li, Z., Ortiz-Bobea, A., & Gomes, C. P. (2022, June). A GNN-RNN approach for harnessing geospatial and temporal information: application to crop yield prediction. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 36, No. 11, pp. 11873-11881). <https://doi.org/10.1609/aaai.v36i11.21444>
- Giji Kiruba, D., Benita, J., & Rajesh, D. (2023). A proficient obtrusion recognition clustered mechanism for malicious sensor nodes in a mobile wireless sensor network. *Indian Journal of Information Sources and Services*, 13(2), 53-63. <https://doi.org/10.51983/ijiss-2023.13.2.3793>
- Jagadeeswaran, Logeswaran, Prasath, S., Thiyagarajan, & Nagarajan. (2022). Machine Learning Model to Detect the Liver Disease. *International Academic Journal of Innovative Research*, 9(1), 06–12. <https://doi.org/10.9756/IAJIR/V9I1/IAJIR0902>
- Kamangir, H., Sams, B., Dokoozlian, N., Sanchez, L., & Earles, J. M. (2024). Cnavit: Integrating climate, managment, and remote sensing data for crop yield estimation with multimodel vision transformers. *arXiv preprint arXiv:2411.16989*. <https://doi.org/10.48550/arXiv.2411.16989>
- Karimov, Z., & Bobur, R. (2024). Development of a food safety monitoring system using IoT sensors and data analytics. *Clinical Journal for Medicine, Health and Pharmacy*, 2(1), 19-29.
- Khaki, S., Wang, L., & Archontoulis, S. V. (2022). County-scale crop yield prediction by integrating crop simulation with machine learning models. *Frontiers in Plant Science*, 13, 1000224.
- Lakshmi, A. S., Raja, G., Pushparaj, D., Sakthivel, S., & Kumar, S. T. (2023). Analysis of student risk factor on online courses using random forest algorithm in machine learning. *International Journal of Advances in Engineering and Emerging Technology*, 14(1), 116–123.

Minimizing the energy consumption of wireless sensor network by comparing the performances of maxweight and minimum energy scheduling algorithms.

Mustapha, S. B., Alkali, A., Nwaydo, N. C., & Mbusube, B. G. (2016). Assessment of Agricultural Extension Service Delivery on Dry Season Onion Production in Bama Local Government Area of Borno State, Nigeria. *International Academic Journal of Social Sciences*, 3(2), 141–147.

Peng, G., Leung, N., & Lechowicz, R. (2025). Applications of artificial intelligence for telecom signal processing. *Innovative Reviews in Engineering and Science*, 3(1), 26-31.

Saiful, S., & Wibisono, N. B. (2025). Crop Yield Prediction Using Random Forest Algorithm and Xgboost Machine Learning Model. *International Journal of Research and Innovation in Social Science*, 9(3), 1983-1994.

Shen, Y., Mercatoris, B., Liu, Q., Yao, H., Li, Z., Chen, Z., & Wang, W. (2024). Use Self-Training Random Forest for Predicting Winter Wheat Yield. *Remote Sensing*, 16(24), 4723. <https://doi.org/10.3390/rs16244723>

Siti, A., & Ali, M. N. (2025). Localization techniques in wireless sensor networks for IoT. *Journal of Wireless Sensor Networks and IoT*, 2(1), 1-12.

Sreenivasu, M., Kumar, U. V., & Dhulipudi, R. (2022). Design and Development of Intrusion Detection System for Wireless Sensor Network. *Journal of VLSI circuits and systems*, 4(2), 1-4. <https://doi.org/10.31838/jvcs/04.02.01>

Srivastava, A. K., Safaei, N., Khaki, S., Lopez, G., Zeng, W., Ewert, F., ... & Rahimi, J. (2022). Winter wheat yield prediction using convolutional neural networks from environmental and phenological data. *Scientific reports*, 12(1), 3215.

Tuğaç, M. G., Özbayoğlu, A. M., Torunlar, H., & Karakurt, E. (2022). Wheat yield prediction with machine learning based on MODIS and landsat NDVI data at field scale. *International Journal of Environment and Geoinformatics*, 9(4), 172-184. <https://doi.org/10.30897/ijegeo.1128985>

Veerasamy, K., & Fredrik, E. T. (2023). *Intelligent Farming based on Uncertainty Expert System with Butterfly Optimization Algorithm for Crop Recommendation*. Infinite Study. <https://doi.org/10.58346/JISIS.2023.I4.011>

Yang, S., Li, L., Fei, S., Yang, M., Tao, Z., Meng, Y., & Xiao, Y. (2024). Wheat yield prediction using machine learning method based on UAV remote sensing data. *Drones*, 8(7), 284. <https://doi.org/10.3390/drones8070284>

Yewle, A. D., Mirzayeva, L., & Karakuş, O. (2025). Multi-modal data fusion and deep ensemble learning for accurate crop yield prediction. *Remote Sensing Applications: Society and Environment*, 38, 101613. <https://doi.org/10.1016/j.rsase.2025.101613>

Zhao, Y., Wang, J., & Liu, Y. (2022). Evaluation of Random Forests (RF) for regional and local-scale wheat yield prediction in Southeast Australia. *Sensors*, 22(3), 717.