Scientia Review

Website: www.scientiafoundation.in

POPULAR SCIENCE ARTICLE

Harnessing Machine Learning Techniques for Forecasting Crop Yields under Changing Climatic Conditions

Sarmistha Borgohain[1*], Chayanika Nath[2]
[1]Krishi Vigyan Kendra, AAU, Sonitpur, Assam- 784028
[2]Krishi Vigyan Kendra, AAU, Darrang, Assam- 784125
*Email: sarmistha.borgohain@aau.ac.in

Received: 22 October 2025
Revised: 25 October 2025
Accepted: 29 October 2025
Published online: 07 November 2025

Article ID: SR01042

Citation: Borgohain, S. & Nath, C. (2025). Harnessing Machine Learning Techniques for Forecasting Crop Yields under Changing Climatic Conditions. *Scientia Review*, 1(5), 12-15

Abstract

Global agriculture is facing increasing uncertainty due to climatic variability, which significantly affects crop growth and productivity. Accurate crop yield prediction under changing environmental conditions has become essential for ensuring food security and sustainable agricultural development. Traditional statistical and mechanistic models have been extensively used for yield estimation, yet they often fail to capture complex, nonlinear relationships among climatic, soil, and management factors. Machine learning techniques have emerged as powerful data-driven tools that can process large and heterogeneous datasets to predict crop yields with higher precision. Various algorithms such as regression models, random forests, support vector machines, artificial neural networks, and deep learning architectures have been successfully implemented for yield forecasting under variable climatic conditions. These methods demonstrate superior adaptability and accuracy in modelling the intricate interactions between environmental variables and crop performance. Integration of machine learning with remote sensing data, precision agriculture technologies, and climate simulation models can further improve the reliability and spatial resolution of yield predictions. The adoption of these advanced computational methods holds significant potential for enhancing agricultural resilience and facilitating informed decision-making in the face of climatic challenges.

Key words: Machine learning, crop yield prediction, climate variability, artificial intelligence, deep learning, precision agriculture

Introduction

Agricultural production across the world is increasingly influenced by fluctuations in climatic conditions. Variations in temperature, precipitation, and solar radiation, along with the growing frequency of droughts and floods, have created considerable uncertainty in crop productivity. Accurate and timely prediction of crop yield under variable climatic scenarios is essential for strategic planning, resource management, and policy formulation aimed at achieving food security (Lobell and Burke, 2010).

Conventional yield forecasting methods, such as empirical statistical models and process-based crop simulation models, have been applied for decades. However, their performance is often limited by the complexity of interactions among environmental factors and by the inability to effectively handle large volumes of heterogeneous data (Van Ittersum *et al*., 2003). In contrast, machine learning provides a data-driven approach capable of learning from historical datasets to uncover hidden patterns and nonlinear dependencies between climatic variables and yield outcomes (Liakos *et al.*, 2018).

By integrating diverse data sources, including climatic records, soil characteristics, remote sensing imagery, and management practices, machine learning models have demonstrated strong predictive performance across different agroecological zones. Their ability to generalize over complex environmental interactions makes them particularly suited for yield prediction under dynamic climatic conditions.

Machine Learning Framework in Yield Prediction

Machine learning involves computational algorithms that automatically learn relationships between input features and target variables through experience and data exposure. In the context of yield forecasting, the process generally includes several stages data acquisition, preprocessing, feature extraction, model training, validation, and prediction. Input features typically consist of climatic parameters such as temperature, rainfall, relative humidity, and solar radiation, in addition to soil fertility indicators and agronomic practices (Jeong *et al.,* 2016).

The performance of a machine learning model depends on the representativeness of the training data and the relevance of input variables. Models are usually evaluated using statistical metrics such as root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) to assess predictive accuracy and robustness under different climatic scenarios.

Machine Learning Algorithms for Crop Yield Prediction

Linear and Nonlinear Regression Methods

Regression-based models are among the simplest machine learning methods applied to crop yield prediction. Linear regression assumes a direct proportional relationship between input variables and yield, which is suitable for systems with low variability. However, crop growth is governed by nonlinear physiological processes influenced by multiple interacting factors, which limits the effectiveness of linear approaches (Drummond *et al*., 2003). Nonlinear regression and kernel-based regression models improve flexibility by modelling curvature in relationships, though they require careful parameter tuning to prevent overfitting.

Decision Tree and Random Forest Models

Decision trees classify input data based on a hierarchical structure of feature thresholds, offering a transparent approach to yield prediction. Random forest, developed as an ensemble of decision trees, enhances prediction stability and reduces variance by averaging multiple tree outputs (Breiman, 2001). Random forest models have demonstrated superior accuracy in estimating yields for crops such as maize, rice, and wheat under diverse climatic conditions.

These models can capture complex interactions between variables and identify the most influential climatic features affecting yield variability. Their robustness and interpretability make them valuable tools for agricultural decision support systems.

Support Vector Machine Models

Support vector machines (SVM) are effective supervised learning algorithms that find an optimal hyperplane separating different data classes or regression outputs in a high-dimensional feature space (Vapnik, 1995). In yield prediction, SVMs are particularly useful when datasets are limited but contain nonlinear relationships between inputs and outputs. Research has shown that SVM models outperform traditional regression techniques when predicting crop yields in regions with unstable weather conditions (Pantazi *et al*., 2016).

Artificial Neural Networks

Artificial neural networks (ANNs) emulate the human brain's neuron connections to process complex and nonlinear relationships among variables. ANNs consist of input, hidden, and output layers where each neuron transforms inputs using weighted connections and activation functions (Kaul *et al*., 2005). When applied to crop yield forecasting, ANNs can capture intricate dependencies between climate parameters such as temperature and rainfall, and crop growth responses (Hansen *et al*., 2006).

Their adaptability makes them effective in predicting yields for diverse crops, though they require large, well-distributed datasets and significant computational resources to achieve high accuracy.

Ensemble Learning Approaches

Ensemble learning combines multiple base learners to form a stronger predictive model. Methods such as bagging, boosting, and stacking improve performance by reducing bias and variance inherent in individual models. Gradient boosting machines (GBM) and extreme gradient boosting (XGBoost) are widely used ensemble methods that have shown exceptional performance in crop yield forecasting due to their ability to handle nonlinear data and heterogeneous feature sets (Chen and Guestrin, 2016).

These techniques are particularly effective when integrating diverse data sources such as climatic records, soil attributes, and remote sensing-derived indices, producing more stable and accurate yield predictions.

Deep Learning Architectures

Deep learning extends neural network architecture by adding multiple hidden layers capable of automatically learning hierarchical representations of data. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are the most common deep learning models used for agricultural applications. CNNs extract spatial features from satellite or aerial imagery to estimate vegetation health and biomass, while RNNs, including long short-term memory (LSTM) networks, analyze sequential climatic data to understand temporal dependencies (Khaki and Wang, 2019).

When trained with remote sensing and climatic datasets, deep learning models have achieved high accuracy in large-scale yield forecasting, demonstrating potential for operational applications in climate-smart agriculture.

Integration of Climatic and Remote Sensing Data

The reliability of yield prediction depends strongly on the quality and diversity of input data. Climatic variables such as rainfall distribution, temperature fluctuations, and solar radiation are fundamental determinants of crop productivity. Integrating these with remote sensing and soil datasets enhances model robustness and accuracy.

Remote sensing technologies such as Landsat, MODIS, and Sentinel satellites generate continuous spatial and temporal information about vegetation dynamics through indices like NDVI (Normalized Difference Vegetation Index) and EVI (Enhanced Vegetation Index). Machine learning models trained on such data can detect subtle changes in crop growth and forecast yields more effectively (Lobell *et al*., 2015).

For example, CNN models using MODIS-derived indices have achieved strong predictive performance in mapping maize yields across different agroecological regions. Combining climatic projections with remote sensing data also enables assessment of future agricultural risks under climate change scenarios.

Challenges in Machine Learning-based Yield Forecasting

Data Limitations

Agricultural data often suffer from incompleteness, inconsistency, and insufficient spatial or temporal resolution. In many developing regions, reliable ground-truth yield records are scarce, limiting the accuracy of machine learning models.

Model Generalization

Models trained on data from one geographic region may not perform well in another due to differences in climatic regimes, soil types, and management practices. Improving model transferability requires domain adaptation techniques and larger, diverse datasets.

Computational and Technical Constraints

Advanced algorithms, especially deep learning models, require substantial computational resources and expert parameter tuning, posing challenges for small-scale or resource-limited applications.

Model Interpretability

Complex machine learning models often act as black boxes, making it difficult to interpret how input variables influence yield outcomes. Model explainability tools such as SHAP values and feature importance rankings can help overcome this limitation (Lundberg and Lee, 2017).

Uncertainty in Climatic Inputs

Climate projections are inherently uncertain, and inaccuracies in input variables can propagate through models, affecting prediction reliability.

Emerging Directions in Yield Prediction

Hybrid Modeling Approaches

Integrating process-based crop growth models with machine learning methods can combine physical understanding of plant growth with data-driven adaptability. These hybrid frameworks have shown promise in improving prediction reliability across diverse climatic conditions.

Transfer Learning and Domain Adaptation

Using pretrained models developed in data-rich regions and adapting them for regions with limited data enhances model scalability. This approach minimizes the need for extensive new datasets and accelerates deployment.

Integration with Precision Agriculture Systems

Machine learning models embedded within precision agriculture platforms enable real-time monitoring and decision-making using IoT devices and sensor data. This integration supports timely interventions such as irrigation scheduling and nutrient management.

Open-access and Cloud-based Platforms

Cloud computing and open-source machine learning frameworks such as TensorFlow and PyTorch have increased accessibility to powerful tools for researchers and practitioners, facilitating collaborative agricultural innovation.

Policy and Decision Support

Accurate yield forecasts can inform government agencies, supply chain managers, and farmers about production trends, enabling proactive responses to climatic variability and market fluctuations.

Conclusion

Machine learning provides a powerful framework for predicting crop yields under changing climatic conditions. By leveraging data-driven algorithms capable of learning complex, nonlinear relationships between environmental, soil, and management factors, these approaches surpass the predictive capacity of traditional models. Techniques such as random forest, gradient boosting, artificial neural networks, and deep learning have demonstrated strong potential in accurately forecasting crop yields across diverse agroclimatic settings.

Nevertheless, challenges related to data availability, model transferability, and interpretability remain significant. Addressing these limitations through hybrid modelling, explainable AI, and improved data infrastructure can enhance the reliability and adoption of machine learning-based forecasting systems. The integration of these advanced methods with climate models and remote sensing data offers a promising pathway toward resilient, efficient, and climate-smart agricultural systems capable of ensuring food security in a rapidly changing environment.

References

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Chen, T. & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

Drummond, S. T., Sudduth, K. A., Joshi, A., Birrell, S. J. & Kitchen, N. R. (2003). Statistical and neural methods for site–specific yield prediction. *Transactions of the ASAE*, *46*(1), 5.

Hansen, J. W., Challinor, A., Ines, A., Wheeler, T. & Moron, V. (2006). Translating climate forecasts into agricultural terms: advances and challenges. *Climate research*, *33*(1), 27-41.

Jeong, J.H., Resop, J.P., Mueller, N.D., Fleisher, D.H., Yun, K., Butler, E.E., Timlin, D.J., Shim, K.M., Gerber, J.S., Reddy, V.R. & Kim, S. H. (2016). Random forests for global and regional crop yield predictions. *PloS one*, *11*(6), e0156571.

Kaul, M., Hill, R. L. & Walthall, C. (2005). Artificial neural networks for corn and soybean yield prediction. *Agricultural Systems*, *85*(1), 1-18.

Khaki, S. & Wang, L. (2019). Crop yield prediction using deep neural networks. *Frontiers in plant science*, *10*, 621.

Liakos, K. G., Busato, P., Moshou, D., Pearson, S. & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, *18*(8), 2674.

Lobell, D. B. & Burke, M. B. (2010). On the use of statistical models to predict crop yield responses to climate change. *Agricultural and forest meteorology*, *150*(11), 1443-1452.

Lobell, D. B., Thau, D., Seifert, C., Engle, E., & Little, B. (2015). A scalable satellite-based crop yield mapper. *Remote Sensing of Environment*, *164*, 324-333.

Lundberg, S. M. & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R. L. & Mouazen, A. M. (2016). Wheat yield prediction using machine learning and advanced sensing techniques. *Computers and electronics in agriculture*, *121*, 57-65.

van Ittersum, M. K., Leffelaar, P. A., Van Keulen, H., Kropff, M. J., Bastiaans, L. & Goudriaan, J. (2003). On approaches and applications of the Wageningen crop models. *European journal of agronomy*, *18*(3-4), 201-234.

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory. Spring, New York.*