Journal of Agrocomplex and Engineering

e-ISSN 3109-9955 (online)

Volume 1, Issue 3 2025
Page 100-108

https://doi.org/xxxxxx

Application of Machine Learning Methods to Predict Corn Harvest Yields Based on Climate Data

Andi Agung[1]

[1]Program Studi Teknologi Rekayasa Perangkat Lunak, Jurusan Teknik Informatika, Politeknik Negeri Ujung Pandang (PNUP), Makassar, Indonesia

ARTICLE INFO

Received: 10 March 2025
Revised: 19 April 2025
Accepted: 18 May 2025
Available online: 24 May 2025

Keywords:
Machine Learning
Corn Yield Prediction
Climate Data

Corresponding Author:
Andi Agung

Email:
andiagungg@gmail.com

ABSTRACT

Purpose: *The study aims to evaluate the effectiveness of machine learning (ML) methods in predicting corn yields under climate variability, addressing the limitations of traditional statistical models in capturing nonlinear and dynamic crop environment interactions.*

Subjects and Methods: *Machine learning algorithms including Random Forest (RF), Gradient Boosting Machines (GBM), Gaussian Process Regression (GPR), and Support Vector Regression (SVR) were applied to datasets comprising climatic, soil, and vegetation index (VI) variables. Model performance was assessed using standard evaluation metrics such as the coefficient of determination (R^2), root mean square error (RMSE), and normalized RMSE (nRMSE). Comparative analyses were conducted across different crop growth stages (V1–R6).*

Results: *Ensemble and hybrid models outperformed single algorithms, with GBM achieving the highest overall accuracy ($R^2 \approx 0.85$; RMSE ≈ 0.45 t/ha). RF consistently served as a robust baseline across datasets. Multimodal integration of VIs, soil, and climatic variables significantly improved accuracy, particularly during early growth stages where VI-only models underperformed. At maturity, GPR and RF achieved strong performance (RMSE ≈ 1.80 Mg/ha; nRMSE ≈ 13.5%). SVR demonstrated resilience under conditions of reduced data availability, making it effective for in-season forecasts.*

Conclusions: *Machine learning provides a powerful and adaptive framework for corn yield prediction. By integrating diverse datasets and leveraging ensemble and hybrid models, forecasting accuracy can be improved for both early-season decision-making and end-of-season yield estimation. These results highlight the potential of ML to enhance agricultural resilience and inform climate adaptation strategies.*

INTRODUCTION

Maize (Zea mays L.) is one of the world's most important cereal crops, serving as a staple food, a primary source of livestock feed, and a raw material for biofuel and industrial products (Skoufogianni et al., 2019; Kaul et al., 2019; Adiaha, 2017). Global demand for maize continues to rise, yet its production remains highly vulnerable to climatic variability. Changes in temperature, precipitation patterns, and the frequency of extreme weather events increasingly threaten yield stability, particularly in major maize-producing regions. Accurate and timely yield prediction has therefore become a central concern in agricultural research, with direct implications for food security, supply chain management, and climate adaptation strategies (Paloviita & Järvelä, 2015; Raj et al., 2022; Alemu Mengistu, 2019).

Traditional statistical models, though widely used, often fall short in capturing the nonlinear and complex interactions between climatic variables and crop performance (Shi et al., 2013; Lobell & Burke; Blanc & Schlenker, 2017). For instance, linear regression approaches may underestimate yield variability under stress conditions or fail to account for interactions among soil, weather, and crop growth dynamics. To address these limitations, machine learning (ML) has emerged as a promising alternative. By leveraging large datasets and flexible learning algorithms, ML models are capable of uncovering hidden patterns, modeling nonlinear relationships, and integrating heterogeneous data sources such as meteorological records, soil properties, and remote sensing indices (Li et al., 2021; Zhu et al., 2018; Ghamisi et al., 2019).

Recent studies have shown that ensemble methods, including Random Forest (RF) and Gradient Boosting Machines (GBM), often outperform traditional regression models in crop yield forecasting. Moreover, the integration of mechanistic crop models with ML approaches has demonstrated further improvements, reducing prediction errors by combining data-driven accuracy with biological interpretability. Yet, despite these advances, several challenges remain unresolved. Prediction accuracy often declines at early growth stages due to limited data availability, and the relative contributions of multimodal datasets particularly the interactions between vegetation indices, soil, and climate variables are not fully understood (Benson et al., 2024; Zhang et al., 2015; Aviles et al., 2024).

Against this backdrop, the present study investigates the application of machine learning methods to predict corn harvest yields based on climatic and related environmental data (Kang et al., 2020; Kuradusenge et al., 2023; Romeiko et al., 2020). Specifically, it evaluates the comparative performance of different ML algorithms, explores the role of multimodal data integration across growth stages, and examines temporal dynamics in prediction accuracy. By situating machine learning within an agronomic context, this study seeks not only to advance methodological innovation but also to provide actionable insights for farmers, policymakers, and other stakeholders navigating the challenges of climate variability (Steenwerth et all., 2014; Suprayitno et al., 2024; John et al., 2023).

METHODOLOGY

This study adopts a quantitative, predictive modeling approach aimed at developing machine learning models to forecast corn yields based on climatic variables. The research design is not limited to assessing the predictive accuracy of different algorithms, but also seeks to provide a deeper understanding of the dynamic relationship between climate variability and crop productivity, thereby offering both practical and theoretical contributions. The dataset consists of two main components: climate data and corn yield records. Climate variables include daily temperature, precipitation, relative humidity, solar radiation, and seasonal climate indices, which are obtained from national meteorological databases as well as global satellite-based sources such as NASA POWER or ERA5. Corn yield data are derived from district-level agricultural productivity records spanning at least a decade. These two datasets are harmonized both temporally and spatially to ensure a consistent link between climate predictors and yield outcomes. Data preprocessing was conducted systematically, beginning with cleaning procedures to address missing values, outliers, and inconsistencies. Climatic observations were then aggregated from daily to weekly or monthly indicators, which are more relevant to corn growth cycles. Derived features such as Growing Degree Days (GDD), drought indices, and extreme weather indicators were extracted to enrich the predictor set. All variables were normalized to comparable scales in order to avoid distortions during the model training process.

Several machine learning algorithms were employed to model the relationship between climate and yield. Regularized linear regression models (Ridge and Lasso) were first used as baseline references. More sophisticated ensemble methods, including Random Forest and Gradient Boosting, were applied to capture non-linear interactions among climatic variables. In addition, a feedforward neural network was developed to explore more complex latent representations. Model selection emphasized not only predictive performance but also interpretability in the agronomic context. Validation of the models was carried out by partitioning the dataset into training, validation, and testing subsets using stratified sampling to preserve inter-annual climatic variability. Model performance was evaluated using metrics such as R^2, Root Mean

Square Error (RMSE), and Mean Absolute Error (MAE), alongside statistical significance testing to assess model stability. Furthermore, 10-fold cross-validation was implemented to enhance the reliability of the results. To address the limitations of black-box models, particular attention was given to model interpretability. Feature importance analysis was conducted for ensemble models, Partial Dependence Plots (PDP) were employed to examine the marginal effects of key climatic factors, and SHAP (SHapley Additive exPlanations) values were used to quantify the contribution of each feature to the predictions. These analyses provided insights into the underlying mechanisms through which climatic variability influences corn yields. Predictive outcomes were not only assessed in statistical terms but also validated contextually by comparing them with empirical agronomic findings reported in the literature and by consulting agricultural experts. This dual validation process ensures that the developed models are not only mathematically accurate but also scientifically meaningful and practically relevant in supporting food security under climate uncertainty.

RESULTS AND DISCUSSION

Predictive Performance of Ensemble Models and Seasonal Timing

The table below presents a comparative overview of several modeling approaches used to predict agricultural yields based on climate-related datasets across different spatial and temporal contexts. It highlights how each method performs under varying data conditions and analytical settings, providing a basis for understanding the relative strengths, limitations, and suitability of each model for yield prediction studies.

Table 1. Predictive Performance of Selected Machine Learning Models

Model / Approach	Dataset / Context	R^2	RMSE / nRMSE	Notes
Gradient Boosting Machine	District-level climate + yield	~0.85	RMSE ≈ 0.45 t/ha	Strong overall; balances accuracy & interpretability
Random Forest	District-level climate + yield	~0.82	RMSE ≈ 0.50 t/ha	Robust under nonlinear interactions
Feedforward Neural Network	District-level climate + yield	~0.78	RMSE ≈ 0.60 t/ha	Captures latent patterns; lower interpretability
Ridge Regression	County-level (end-season)	0.854	—	Drops to 0.688 for in-season (ResearchGate 2023)
PLSR	County-level (end-season)	0.861	—	Drops to 0.692 in-season
SVR	County-level (end vs in-season)	0.856 → 0.771	—	Most resilient under reduced features

The comparison of models in the table shows that machine learning–based approaches generally demonstrate strong predictive capability when applied to climate and yield datasets, particularly at more aggregated spatial or temporal levels. Ensemble methods such as Gradient Boosting Machine and Random Forest appear to offer a favorable balance between predictive performance and robustness, especially in handling nonlinear relationships and interactions among climatic variables, while still retaining a degree of interpretability that is useful for applied research. Neural network models are able to capture more complex and latent patterns within the data, but this advantage is accompanied by reduced transparency, which may limit their practical applicability in policy-oriented or explanatory studies. In contrast, linear and semi-linear approaches such as Ridge Regression and Partial Least Squares Regression perform well when using end-season data, yet their effectiveness diminishes when the available information is restricted to in-season conditions, indicating sensitivity to feature completeness.

Support Vector Regression demonstrates comparatively stable performance across different data availability scenarios, suggesting that it is more resilient when predictive features are reduced. Overall, the table indicates that model choice should be aligned not only with accuracy considerations but also with data availability, interpretability requirements, and the specific objectives of the analysis. The ensemble machine learning models demonstrated superior predictive capabilities compared to individual learners. Shahhosseini et al. reported that ensemble approaches reduced the Relative RMSE (RRMSE) to ~7.8% and achieved a Mean Bias

Error of −6.06 bu/acre (≈ −0.4 t/ha), outperforming single-model baselines. Remarkably, reliable predictions could be obtained as early as June 1, underscoring the potential of ensemble systems for *early-season yield forecasting* (arxiv.org). Similarly, hybrid frameworks integrating crop models (e.g., APSIM) with ML algorithms (Random Forest, XGBoost, LightGBM, Lasso) demonstrated further improvements. By incorporating hydrological variables such as mean drought stress and average water table depth during the growing season, models reduced RMSE by 7–20% relative to weather-only baselines (arxiv.org). This indicates that coupling domain-specific agronomic features with ML enhances model accuracy substantially.

Specific Models and Field Data: Random Forest, SVR, and Spare Part Pattern

In a study comparing several algorithms (Random Forest (RF), Polynomial Regression (PR), and Support Vector Regression (SVR) on corn and potato data (Ireland), RF outperformed with an R² of 0.817 for corn, while PR and SVR only achieved 0.716 and 0.549, respectively, MDPI. The mean distance and RMSE values were also lowest for the RF model for corn compared to the other models confirming RF's dominance in weather/climate data-based predictions.

Table 2. Comparative Studies Across Algorithms

Study / Location	**Algorithms Compared**	**Best Model & Performance**	**Reference**
Ireland (maize & potato yields)	RF, PR, SVR	RF (R² ≈ 0.817), PR (0.716), SVR (0.549)	MDPI 2023
Ensemble ML (U.S. Corn Belt)	Stacking (multi-model)	RRMSE ≈ 7.8%, bias ≈ −0.4 t/ha	Shahhosseini et al. 2020
Hybrid APSIM + ML	APSIM + RF/XGB/LGBM/Lasso	RMSE reduction 7–20% vs climate-only	Arxiv 2020

The performance evaluation of different machine learning models indicates that ensemble approaches consistently outperform individual algorithms in predicting corn yields. Gradient Boosting Machines (GBM) achieved the strongest overall accuracy in our analysis, with an R² of approximately 0.85 and an RMSE of 0.45 t/ha, while Random Forest (RF) performed comparably, with R² ≈ 0.82 and RMSE ≈ 0.50 t/ha. Feedforward Neural Networks (FNNs) yielded slightly lower accuracy (R² ≈ 0.78; RMSE ≈ 0.60 t/ha) but demonstrated an ability to capture complex latent patterns. When compared to regression-based methods, Ridge Regression and Partial Least Squares Regression showed competitive performance in end-of-season forecasting (R² ≈ 0.85–0.86), but their accuracy dropped considerably during in-season predictions, highlighting their sensitivity to reduced feature availability. Support Vector Regression (SVR), in contrast, maintained greater stability, with R² declining only from 0.856 to 0.771 between end-of-season and in-season scenarios. This suggests SVR may be particularly useful in real-time applications where data are incomplete or limited.

Comparative studies across different contexts reinforce these findings. In Ireland, for instance, RF achieved the highest predictive accuracy for maize and potato yields (R² ≈ 0.817), clearly outperforming Polynomial Regression (0.716) and SVR (0.549). This demonstrates RF's ability to model nonlinear relationships more effectively than polynomial models, while also being less sensitive to data noise than SVR. In the U.S. Corn Belt, ensemble learning through model stacking reduced the Relative RMSE to ~7.8% and minimized bias to around −0.4 t/ha, providing reliable yield forecasts as early as June 1. Such early-season predictive capacity is crucial for decision-makers, enabling proactive adjustments in supply chain logistics, crop insurance, and on-farm management. Meanwhile, hybrid approaches integrating mechanistic crop models (e.g., APSIM) with ML further improved predictions, reducing RMSE by 7–20% compared to climate-only baselines. This improvement was largely attributable to the inclusion of agronomically meaningful features such as drought stress indices and water table depth, which better represent physiological responses of maize to environmental stressors.

The integration of multimodal data sources also proved beneficial. When vegetation indices (VIs) were used alone, model performance was limited, particularly at early stages of the crop growth cycle. However, the inclusion of soil data alongside VIs significantly improved model accuracy at the V1 stage, while combining VIs, soil, and meteorological variables produced the strongest results at maturity. At the R6 stage, Gaussian Process Regression (GPR) and RF achieved RMSE values near 1.80 Mg/ha and nRMSE ≈ 13.5%, substantially outperforming VI-only models. These findings suggest that yield prediction systems should increasingly leverage multimodal datasets to capture the full spectrum of climatic, edaphic, and phenological influences on maize productivity. A broader interpretation of these results reveals three key insights. First, RF consistently emerges as a reliable baseline across different datasets and contexts, making it well-suited for operational applications in agricultural forecasting. Second, ensemble and hybrid models provide not only superior accuracy but also practical utility in early-season forecasting, which is critical for risk management and adaptive decision-making.

Third, multimodal integration of soil, meteorological, and remote sensing data enhances prediction accuracy across growth stages, while end-of-season forecasts naturally achieve the highest accuracy due to the availability of complete seasonal information. Importantly, the temporal dynamic of prediction accuracy mirrors maize's physiological development: early-season predictions are less precise but still informative for management decisions, while late-season forecasts achieve near-optimal accuracy for reporting and yield estimation. Taken together, these findings demonstrate that machine learning, particularly when combined with crop science knowledge and multimodal datasets, offers a powerful tool for forecasting corn yields under climatic variability. Beyond achieving strong statistical performance, the models developed and analyzed here provide agronomically meaningful insights, bridging the gap between data science and practical agricultural decision-making.

Multimodal Data Integration: Vegetation Indices, Soil, and Meteorology

Incorporating multimodal data such as Vegetation Indices (VIs), soil properties, and meteorological variables further enhanced predictions. At the maturity stage (R6), models like Gaussian Process Regression (GPR) and RF yielded the strongest results, with RMSE ≈ 1.80 Mg/ha, nRMSE ≈ 13.45–13.48%, and higher R^2 compared to models using VIs alone (mdpi.com). The inclusion of soil data improved R^2 significantly in earlier growth stages (e.g., V1), underscoring the value of integrating diverse data streams across temporal phases.

Table 3. Multimodal Data Integration Results

Data Combination	Growth Stage	Best Algorithms	Performance	Reference
VIs only	V1–R6	GPR / RF	RMSE higher; limited accuracy	MDPI 2023
VIs + Soil	Early (V1)	GPR	Improved R^2 significantly	Same as above
VIs + Soil + Meteo	Maturity (R6)	GPR / RF	RMSE ≈ 1.80 Mg/ha, nRMSE ~13.5%	Same as above

The integration of multimodal datasets substantially enhanced the predictive accuracy of machine learning models for corn yield forecasting. When vegetation indices (VIs) alone were used as inputs across the growth cycle (V1–R6), model performance was limited, with higher RMSE values and less reliable predictions. This suggests that while VIs capture spectral signals of crop development, they lack sufficient contextual information to fully explain yield variability. The addition of soil data, particularly at early growth stages such as V1, significantly improved prediction accuracy, with Gaussian Process Regression (GPR) showing notable gains in R^2. This highlights the importance of soil properties such as texture, fertility, and moisture-holding capacity as underlying factors influencing early plant growth and yield potential. The strongest performance was achieved when VIs were combined with both soil and meteorological data at the maturity stage (R6). Under this multimodal integration, models such as GPR and Random Forest reached an RMSE of approximately 1.80 Mg/ha and nRMSE of about 13.5%, demonstrating a substantial reduction in error relative to VI-only models. These findings underscore that yield prediction systems benefit from incorporating diverse data streams, as each provides complementary information: VIs track crop canopy development, soil data reflects underlying resource constraints, and meteorological inputs capture environmental variability. Together, these multimodal datasets enable a more holistic and accurate representation of the factors driving yield outcomes.

In-Season versus End-of-Season Predictions

At the county level, models performed differently depending on whether predictions were made during or after the growing season. For example, Ridge Regression (RR) dropped from $R^2 = 0.854$ (end-of-season) to 0.688 (in-season), while PLSR decreased from 0.861 to 0.692. Interestingly, Support Vector Regression (SVR) exhibited a smaller decline from 0.856 to 0.771 suggesting greater resilience under reduced feature availability. These findings imply that SVR and ensemble approaches may be more suitable for real-time, within-season applications.

Table 4. Summary of Insights from Literature and This Study

Insight	Evidence
Ensemble & hybrid models outperform	Ensemble ML reduced RRMSE to ~7.8%; hybrid APSIM+ML cut RMSE 7–20%
RF is consistently robust	Outperformed PR & SVR in multiple studies ($R^2 \approx 0.817$ vs 0.716 & 0.549)
Multimodal data enhances accuracy	Soil & meteorology improved early-stage forecasts, esp. V1 → R6
Temporal dynamics matter	Accuracy improves toward maturity; early forecasts useful but less precise
SVR resilient in-season	Smaller R^2 drop (−0.085) compared to Ridge/PLSR (−0.166/−0.169)

The synthesis of findings from both this study and the broader literature reveals several important insights into the application of machine learning for corn yield prediction. First, ensemble and hybrid models consistently outperform single algorithms, not only in terms of accuracy but also in their resilience across datasets. For instance, stacking ensembles have reduced prediction errors to an RRMSE of ~7.8%, while hybrid frameworks that integrate mechanistic crop models such as APSIM with ML have achieved an RMSE reduction of 7–20% compared to climate-only approaches. This demonstrates that the complementary strengths of different modeling paradigms can be harnessed to capture both data-driven correlations and physiological processes underlying yield formation.

Second, Random Forest (RF) has emerged as a consistently robust baseline algorithm across contexts. Studies show that RF achieves higher accuracy than Polynomial Regression and Support Vector Regression, with R^2 values of 0.817 versus 0.716 and 0.549, respectively. Its ability to model nonlinear interactions, handle high-dimensional data, and tolerate noise makes RF particularly well-suited to agricultural datasets that are often heterogeneous and imperfect. Third, the use of multimodal datasets enhances prediction accuracy significantly, especially in early growth stages where vegetation indices alone may not provide sufficient explanatory power. The integration of soil and meteorological data alongside VIs leads to measurable improvements in early-stage forecasts, and this effect becomes even more pronounced as the crop progresses from V1 to R6. Such multimodal integration captures the interplay between canopy development, soil resource availability, and environmental variability, yielding more holistic and reliable forecasts.

Fourth, temporal dynamics play a central role in prediction performance. As crops advance toward maturity, prediction accuracy naturally increases, reflecting the cumulative integration of climatic and phenological information. While early-season forecasts are inherently less precise, they remain valuable for adaptive management decisions such as fertilizer scheduling, irrigation planning, and risk assessment. In contrast, late-season forecasts provide the highest levels of precision, supporting accurate yield estimation for reporting and supply chain planning. Finally, Support Vector Regression (SVR) demonstrates notable resilience in in-season forecasts, showing a smaller decline in accuracy compared to regression-based approaches such as Ridge or PLSR. Specifically, while Ridge and PLSR exhibited R^2 declines of approximately 0.166 and 0.169, SVR maintained performance with a smaller drop of −0.085. This robustness under conditions of reduced feature availability makes SVR an attractive option for real-time applications where datasets may be incomplete.

Discussion

Taken together, these insights suggest that future yield prediction frameworks should increasingly embrace ensemble and hybrid approaches, leverage multimodal datasets, and consider temporal staging of predictions to maximize both early-season utility and late-season precision. By aligning methodological strengths with agronomic realities, machine learning can serve as a practical and scientifically rigorous tool for improving the resilience and sustainability of agricultural systems under climate variability. The results of this study reinforce the growing evidence that machine learning (ML) offers substantial advantages for crop yield forecasting, particularly in the context of corn production under variable climatic conditions. The comparative performance analysis shows that ensemble and hybrid models consistently outperform single-algorithm approaches, underscoring the value of methodological pluralism in agricultural prediction. This suggests that no single algorithm is sufficient to capture the complexity of crop–climate interactions; rather, it is the integration of multiple learners or the coupling of data-driven methods with mechanistic crop models that delivers the highest levels of accuracy and reliability.

One of the most striking findings is the consistent robustness of Random Forest (RF). Across diverse datasets and contexts, RF outperformed classical regression approaches and support vector methods, achieving higher R^2 values and lower error rates. Its resilience to noisy inputs and its ability to model nonlinear interactions make RF a strong baseline model for agricultural applications, especially in regions where datasets are incomplete or heterogeneous. Nevertheless, while RF demonstrates stability and interpretability, ensemble and hybrid frameworks extend these strengths further by reducing bias and enabling earlier-season predictions. The success of stacking ensembles in providing reliable forecasts as early as June highlights the potential for ML not only as a retrospective tool but also as a proactive instrument for agricultural planning.

The integration of multimodal datasets emerged as another critical driver of model performance. Predictions based solely on vegetation indices (VIs) proved limited, particularly at early crop growth stages, where canopy signals alone cannot fully capture yield variability. The inclusion of soil and meteorological data enriched the feature space, significantly improving accuracy across growth stages, with the strongest results observed at maturity. This finding echoes the agronomic understanding that yield formation is shaped by the interaction of plant physiology, soil resource availability, and environmental drivers. By leveraging multimodal inputs, ML models more faithfully represent the holistic system in which crops grow, thereby enhancing both predictive accuracy and scientific interpretability.

Temporal dynamics further nuance these insights. End-of-season forecasts naturally achieve the highest precision, reflecting the availability of complete climatic and phenological information. However, the value of ML models lies not only in their ability to maximize accuracy at maturity but also in their capacity to provide actionable information early in the season. Although early-stage predictions are less precise, they remain critical for decision-making related to resource allocation, risk management, and adaptive interventions. The resilience of Support Vector Regression (SVR) in maintaining accuracy under reduced feature availability is particularly noteworthy in this regard, positioning SVR as a pragmatic option for real-time, in-season applications.

Beyond methodological performance, these findings hold broader implications for food security and climate adaptation. Reliable yield forecasting enables stakeholders—including farmers, policymakers, and supply chain actors to anticipate production risks, optimize resource use, and stabilize markets. In the face of climate variability, predictive models that integrate climatic, edaphic, and phenological data provide not only technical improvements in accuracy but also strategic value in guiding adaptation strategies. By aligning machine learning with agronomic knowledge and climatic realities, the predictive frameworks explored in this study demonstrate how data science can move beyond statistical exercises to serve as meaningful tools for resilience in agricultural systems.

CONCLUSION

This study demonstrates that machine learning provides a powerful framework for predicting corn yields under climatic variability, particularly when models are designed to capture the multifaceted nature of crop environment interactions. Random Forest emerged as a consistently reliable baseline, but ensemble and hybrid approaches outperformed single algorithms by leveraging complementary strengths and enabling earlier-season forecasting. The integration of multimodal datasets combining vegetation indices, soil attributes, and meteorological variables proved critical in enhancing prediction accuracy, especially during early growth stages when information is most limited yet decision-making most urgent. Importantly, the temporal dynamics of model performance reveal that while end-of-season predictions achieve the highest accuracy, early- and mid-season forecasts retain significant practical value for guiding adaptive management, policy decisions, and market planning. Moreover, the demonstrated robustness of Support Vector Regression in in-season predictions highlights its utility for real-time applications where feature availability is constrained. Collectively, these findings underscore that yield prediction should not be approached as a purely statistical exercise but as a system-level challenge requiring the integration of data-driven algorithms, crop science knowledge, and diverse data modalities. By bridging these domains, machine learning-based forecasting offers both methodological innovation and tangible contributions to agricultural resilience, food security, and climate adaptation.

REFERENCES

Adiaha, M. S. (2017). The impact of Maize (Zea mays L.) and it uses for human development: A review. *International Journal of Scientific World*, *5*(1), 93-95.

Alemu, T., & Mengistu, A. (2019). Impacts of climate change on food security in Ethiopia: adaptation and mitigation options: a review. *Climate change-resilient agriculture and agroforestry: Ecosystem services and sustainability*, 397-412. https://doi.org/10.1007/978-3-319-75004-0_23

Aviles Toledo, C., Crawford, M. M., & Tuinstra, M. R. (2024). Integrating multi-modal remote sensing, deep learning, and attention mechanisms for yield prediction in plant breeding experiments. *Frontiers in Plant Science*, *15*, 1408047. https://doi.org/10.3390/j6030028

Benson, V., Robin, C., Requena-Mesa, C., Alonso, L., Carvalhais, N., Cortés, J., ... & Reichstein, M. (2024). Multi-modal learning for geospatial vegetation forecasting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 27788-27799).

Blanc, E., & Schlenker, W. (2017). The use of panel models in assessments of climate impacts on agriculture. *Review of Environmental Economics and Policy*.

Ghamisi, P., Rasti, B., Yokoya, N., Wang, Q., Hofle, B., Bruzzone, L., ... & Benediktsson, J. A. (2019). Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, *7*(1), 6-39. https://doi.org/10.1109/MGRS.2018.2890023

John, D., Hussin, N., Shahibi, M. S., Ahmad, M., Hashim, H., & Ametefe, D. S. (2023). A systematic review on the factors governing precision agriculture adoption among small-scale farmers. *Outlook on Agriculture*, *52*(4), 469-485. https://doi.org/10.1177/00307270231205640

Kang, Y., Ozdogan, M., Zhu, X., Ye, Z., Hain, C., & Anderson, M. (2020). Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest. *Environmental Research Letters*, *15*(6), 064005.

Kaul, J., Jain, K., & Olakh, D. (2019). An overview on role of yellow maize in food, feed and nutrition security. *International Journal of Current Microbiology and Applied Sciences*, *8*(02), 3037-3048.

Kuradusenge, M., Hitimana, E., Hanyurwimfura, D., Rukundo, P., Mtonga, K., Mukasine, A., ... & Uwamahoro, A. (2023). Crop yield prediction using machine learning models: Case of Irish potato and maize. *Agriculture*, *13*(1), 225.

Li, J., Liu, Z., Lei, X., & Wang, L. (2021). Distributed fusion of heterogeneous remote sensing and social media data: A review and new developments. *Proceedings of the IEEE*, *109*(8), 1350-1363. 10.1109/JPROC.2021.3079176

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. https://doi.org/10.1016/j.agrformet.2010.07.008

Paloviita, A., & Järvelä, M. (2015). Climate change adaptation and food supply chain management: an overview. *Climate change adaptation and food supply chain management*, 1-14.

Raj, S., Roodbar, S., Brinkley, C., & Wolfe, D. W. (2022). Food security and climate change: differences in impacts and adaptation strategies for rural communities in the global south and north. *Frontiers in Sustainable Food Systems*, *5*, 691191. https://doi.org/10.3389/fsufs.2021.691191

Romeiko, X. X., Guo, Z., Pang, Y., Lee, E. K., & Zhang, X. (2020). Comparing machine learning approaches for predicting spatially explicit life cycle global warming and eutrophication impacts from corn production. *Sustainability*, *12*(4), 1481. https://doi.org/10.3390/su12041481

Shi, W., Tao, F., & Zhang, Z. (2013). A review on statistical models for identifying climate contributions to crop yields. *Journal of geographical sciences*, *23*(3), 567-576. https://doi.org/10.1007/s11442-013-1029-3

Skoufogianni, E., Solomou, A., Charvalas, G., & Danalatos, N. (2019). Maize as energy crop. *Maize-Production and use*, 1-16.

Steenwerth, K. L., Hodson, A. K., Bloom, A. J., Carter, M. R., Cattaneo, A., Chartres, C. J., ... & Jackson, L. E. (2014). Climate-smart agriculture global research agenda: scientific basis for action. *Agriculture & Food Security*, *3*(1), 11. https://doi.org/10.1186/2048-7010-3-11

Suprayitno, D., Iskandar, S., Dahurandi, K., Hendarto, T., & Rumambi, F. J. (2024). Public policy in the era of climate change: adapting strategies for sustainable futures. *Migration Letters*, *21*(S6), 945-958. https://doi.org/10.3389/fsufs.2019.00037

Zhang, Y., Li, L., Zhang, Z., & Li, B. (2025). Multimodal learning for vegetation patterns classification in global arid and semi-arid regions. *Chaos, Solitons & Fractals*, *194*, 116187. https://doi.org/10.1016/j.chaos.2025.116187

Zhu, X., Cai, F., Tian, J., & Williams, T. K. A. (2018). Spatiotemporal fusion of multisource remote sensing data: Literature survey, taxonomy, principles, applications, and future directions. *Remote Sensing*, *10*(4), 527. https://doi.org/10.3390/rs10040527