

Prediction of Maize Yield Based on Soil Nutrients and Climate Variables

Jinhua Cheng, Wei Wang ✉

Institute of Life Sciences, Jiyang College of Zhejiang A&F University, Zhuji, 311800, Zhejiang, China

✉ Corresponding author: wei.wang@jicat.org

Computational Molecular Biology, 2026, Vol.16, No.2 doi: [10.5376/cmb.2026.16.0008](https://doi.org/10.5376/cmb.2026.16.0008)

Received: 02 Feb., 2026

Accepted: 08 Mar., 2026

Published: 21 Mar., 2026

Copyright © 2026 Cheng and Wang, This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preferred citation for this article:

Cheng J.H., and Wang W., 2026, Prediction of maize yield based on soil nutrients and climate variables, *Computational Molecular Biology*, 16(2): 98-113 (doi: [10.5376/cmb.2026.16.0008](https://doi.org/10.5376/cmb.2026.16.0008))

Abstract Maize yield prediction plays an essential role in ensuring food security and promoting sustainable agricultural management. This study explores a prediction framework based on soil nutrient characteristics and climate variables to improve the accuracy and reliability of maize yield estimation. Key soil indicators, including nitrogen, phosphorus, potassium, organic matter, and pH value, were combined with climate factors such as temperature, precipitation, and accumulated growing degree days. Multiple prediction models, including traditional statistical approaches, machine learning algorithms, and deep learning methods, were constructed and compared. The study further analyzed the interaction effects between soil and climate variables and evaluated model performance using indicators such as RMSE, MAE, and R^2 . A regional case study was conducted to verify the applicability and robustness of the proposed framework. The results demonstrate that integrating soil nutrient and climate data can significantly enhance maize yield prediction accuracy and provide valuable support for precision agriculture, crop management, and agricultural decision-making.

Keywords Maize yield prediction; Soil nutrients; Climate variables; Machine learning; Precision agriculture

1 Introduction

Global demand for maize is rising steadily as it underpins food, feed, and industrial supply chains, yet production is increasingly constrained by climate variability and degraded soils. Temperature extremes, altered rainfall, and declining soil fertility jointly threaten yield stability, especially in regions already facing food insecurity. Improving the accuracy of maize yield prediction by explicitly linking soil nutrients with key climate variables is therefore essential for optimizing fertilization, managing risk, and designing climate-smart production systems. Maize yields respond strongly to interactions between climate conditions and soil nutrient status. Studies in sub-Saharan Africa and China show that nitrogen (N), phosphorus (P), and potassium (K) inputs can buffer or amplify the impacts of changing CO₂, temperature, and rainfall on yield, and that soil indigenous nutrients strongly modulate yield losses under warming (Falconnier et al., 2020). Long-term experiments further indicate that soil fertility improvements (e.g., higher total and available N and P) enhance yield stability and sustainability, while climate warming tends to reduce yields where soil fertility is low. At the same time, nutrient management alone is insufficient; integrating soil, climate, and management information is needed to maintain productivity under ongoing climate change (Ocwa et al., 2023). In this context, a predictive framework that couples soil nutrient properties with climate variables can support more precise fertilizer recommendations, reduce environmental risks, and improve resilience of maize-based systems.

Internationally, two main directions have emerged. First, process-based crop models are used to simulate maize yield responses to climate scenarios and N management, revealing strong interactions between N inputs, soil N dynamics, and climate drivers in both low-input and intensive systems (Falconnier et al., 2020). Second, data-driven approaches, especially machine learning (ML) and deep learning (DL), increasingly predict crop yields from large datasets combining soil, climate, and management information. Systematic reviews show that temperature, rainfall, soil type, soil nutrients, and vegetation indices are among the most frequently used predictors, and that algorithms such as Random Forest (RF), Support Vector Machines, Artificial Neural Networks, CNNs, and LSTMs dominate recent work. For maize specifically, RF models trained on multi-year field trials in Ghana identified soil properties (e.g., organic carbon, total N, exchangeable bases) and maximum

temperature as the most important predictors of yield, surpassing purely climatic models and improving understanding of nutrient-climate interactions (Asamoah et al., 2024). Related studies using RF and other ML algorithms have shown that including both soil and weather variables substantially improves prediction of maize yield under zero N fertilization and in drought-stressed environments. These advances highlight the potential of combining soil nutrient information with climate variables in robust predictive frameworks, but also reveal gaps: many models rely on limited nutrient descriptors, treat climate and soil separately, or focus on short time periods and narrow environments.

Building on this progress, the present study focuses on prediction of maize yield based explicitly on soil nutrient status and climate variables, aiming to better capture their joint effects. The main research contents are: (1) construction of a comprehensive feature set describing soil nutrients (e.g., N, P, K, organic matter, pH and related properties) and key climate factors (temperature, precipitation, radiation, humidity) relevant to maize growth; (2) development and comparison of data-driven yield prediction models, with emphasis on ensemble methods such as Random Forest and other ML/DL techniques that have shown strong performance in crop yield prediction; and (3) quantitative analysis of variable importance and interaction patterns between soil nutrients and climate variables, to identify critical drivers of yield variation and potential leverage points for management. The technical route begins with data collection and preprocessing, including quality control and normalization of soil and climate data. Next, the dataset is split into training and testing subsets, and multiple candidate models are trained, tuned, and evaluated using metrics such as coefficient of determination (R^2) and root mean square error (RMSE), following best practices from recent ML yield-prediction studies. Finally, model interpretation techniques (e.g., variable importance analysis and partial response analysis) are applied to quantify how specific combinations of soil nutrients and climate variables influence predicted maize yield, providing both a practical prediction tool and theoretical insight for nutrient management and climate adaptation strategies.

Across diverse environments, maize yield is jointly controlled by soil nutrient status and climate conditions, and their interaction largely determines both productivity and stability. While process-based models and ML/DL approaches have advanced yield prediction, there remains a need for models that explicitly integrate detailed soil nutrient descriptors with key climate variables and provide interpretable guidance for management. This study addresses that gap by constructing and evaluating data-driven maize yield prediction models grounded in soil-climate interactions, aiming to support more precise fertilization, risk management, and climate-smart maize production.

2 Analysis of Factors Influencing Maize Yield

2.1 Mechanism of soil nutrients on maize growth

Maize yield is jointly controlled by soil nutrient supply and climate conditions throughout the growing season. Understanding how these drivers act individually and in combination is essential for reliable yield prediction and targeted management. Adequate N, P, and K fertilization strongly enhances maize growth traits such as plant height, leaf area, cob number, and grain weight, which together raise biomass accumulation and grain yield by large margins compared with unfertilized controls (Kaleri et al., 2026). Long-term NPK application improves key soil properties-including soil organic carbon and available N, P, and K-which in turn explain a larger share of yield variation than phenological factors in the North China Plain (Wang et al., 2024).

Nutrient deficiency, especially of nitrogen and phosphorus, markedly reduces yield and dry matter accumulation in maize-based systems (Sun et al., 2024). Under N, P, or K deficiency, maize root growth and activity are inhibited, and hundreds of genes related to nutrient transport, hormones, and transcription factors are differentially expressed, indicating complex molecular regulation of root adaptation to low nutrient supply (Nana et al., 2020).

2.2 Effects of climate factors on maize yield

Temperature, precipitation, drought, and vapor pressure deficit (VPD) strongly shape maize yield anomalies at regional to global scales. Temperature-related extremes generally show stronger associations with yield deviations than precipitation alone, although irrigation can partially buffer high-temperature damage (Figure 1) (Vogel et al., 2019). In Northeast China, compound drought and heat cause greater yield loss than either stress alone, with

warm-dry years producing the largest reductions and yield loss increasing with temperature and VPD but decreasing with precipitation (Li et al., 2021).

Beyond extremes, the balance between atmospheric evaporative demand and soil moisture is critical. Including interactions between VPD and root-zone soil moisture greatly improves statistical prediction of maize yield anomalies, and estimates that ignore soil moisture can overstate climate-induced yield damage by about a factor of two. Similar work in China shows that maize benefits only when atmospheric moisture demand and soil moisture remain in relative balance; accounting for soil moisture halves projected yield losses compared with using atmospheric demand alone (Zhao et al., 2023).

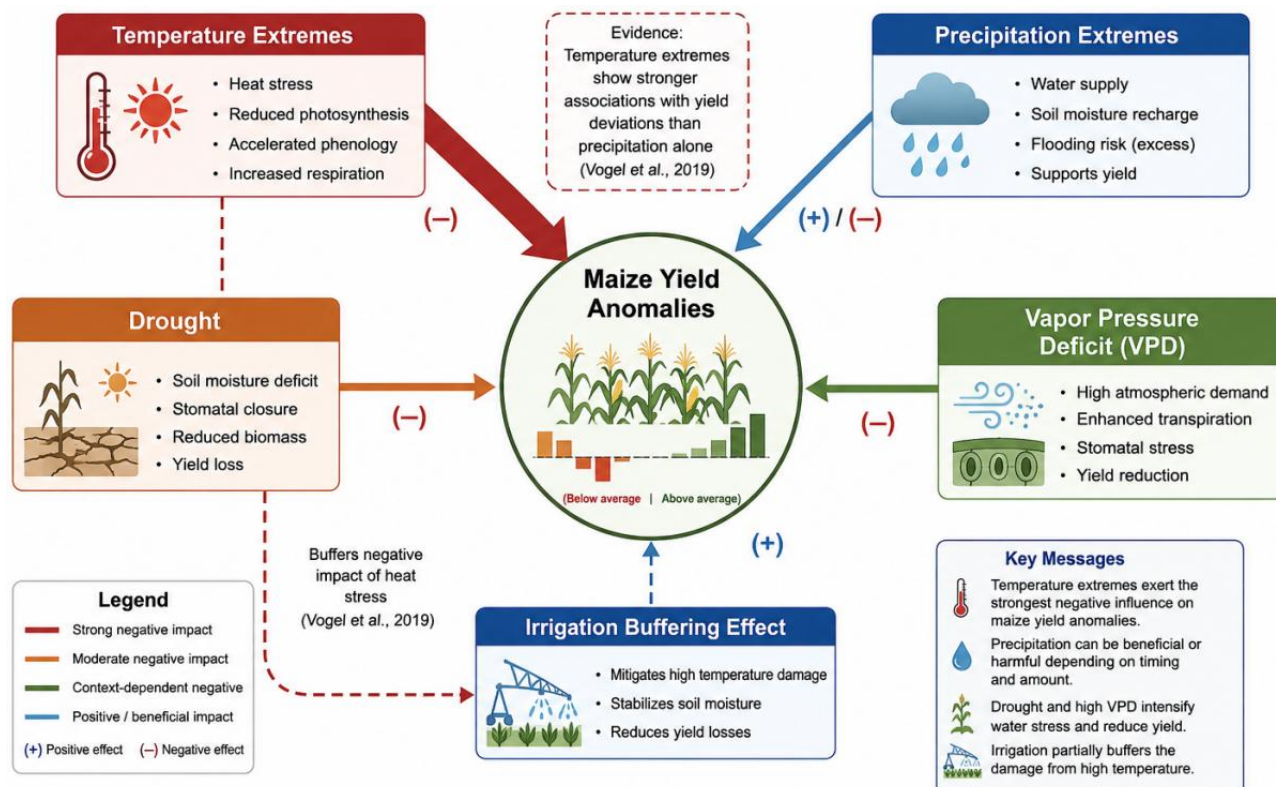


Figure 1 Climate extreme drivers of maize yield anomalies at regional to global scales

2.3 Synergistic mechanism of soil and climate factors

Soil fertility and climate interact to determine both average yield and its stability over time. Long-term experiments show that balanced NPK fertilization not only raises mean maize yield but also improves the stability of relative yield anomalies, while models that combine climate variables with nutrient status explain far more variation in yield anomalies than climate alone (Zhu et al., 2024). In diverse maize systems, soil moisture and temperature jointly drive yield damage, and predictions that include both components outperform those relying only on temperature and precipitation, underscoring the tight soil-climate coupling.

Nitrogen supply particularly modulates maize sensitivity to climate change. In low-input systems, higher N fertilization increases the crop's responsiveness to elevated CO₂, higher temperatures, and altered rainfall, making intensively managed maize more sensitive-and thus more climatically risky-than low-input maize (Falconnier et al., 2020). At larger scales, management intensification (including improved nutrients and technologies) accounts for most historical yield gains, but its benefits are increasingly constrained by warming and drought, meaning that future intensification must explicitly incorporate climate adaptation to sustain yield trends (Medina and Tian, 2023).

Maize yield depends on sufficient N, P, and K to build a productive canopy and reproductive sink, while climate factors-especially temperature extremes, drought, VPD, and soil moisture-govern year-to-year variability. Nutrient

management alters both climate sensitivity and yield stability, so predictive models and management strategies must jointly consider soil fertility and climate interactions rather than treating them in isolation.

3 Data Sources and Overview of the Study Area

3.1 Natural and agricultural conditions of the study area

The major maize-producing regions of northern and northeastern China are characterized by temperate monsoon climates with distinct growing seasons, where temperature, precipitation, and sunshine jointly determine maize climate suitability at different phenological stages (Wang et al., 2024). In the Northeast, relatively cooler temperatures and variable rainfall make precipitation a key limiting factor, while temperature plays a stronger role in the suitability index than in more southerly zones. In contrast, the Huang-Huai-Hai (3H) region has warmer average temperatures and generally higher comprehensive climate suitability, although spatial differences in precipitation and sunshine still create heterogeneous yield potentials. Across China's broader maize belt, temperature variability and climate perturbations can cause substantial yield losses, especially under warming, but these impacts are spatially heterogeneous (Chen et al., 2024).

Soil conditions in the maize belt range from high-soil organic carbon (SOC) soils in parts of the Northeast to more degraded or compacted soils in other regions, and these differences strongly affect yield responses to climate. High SOC, favorable texture, and adequate field capacity enhance buffering capacity against adverse temperature and moisture perturbations, stabilizing yields under climate variability (Feng et al., 2022). In contrast, soils with higher bulk density, coarser texture, or lower water-holding capacity tend to amplify yield losses under warming, underscoring the importance of soil improvement for resilient production. Regional tillage practices, such as deep ploughing or conservation tillage, also interact with local climate: in cooler sites, practices that improve early-season soil temperature and water availability promote maize emergence and growth, whereas in warmer, windier areas, systems that enhance water retention and aeration can be more beneficial (Qian et al., 2025).

3.2 Data sources and acquisition methods

Maize yield data and associated environmental variables can be obtained from long-term field trials, experimental stations, and statistical records, often at plot or county scales. Multi-year experiments in Northeast China and the North China Plain provide detailed measurements of yield, phenology, and management, suitable for evaluating soil-climate interactions and model performance. In some studies, plot-scale experiments under different fertilization or tillage systems supply yield and soil measurements across contrasting climate conditions, enabling analysis of management impacts on yield and soil properties (Meng et al., 2021; Qian et al., 2025). For broader regional coverage, station networks combining agronomic records with local weather observations support large-scale assessments of yield responses to climate variability and soil attributes.

Climate data are typically derived from ground-based automatic weather stations and gridded meteorological datasets, providing variables such as temperature, precipitation, radiation, humidity, and derived indices (e.g., heat degree days, consecutive dry days) during key growth stages (Dandrifosse et al., 2024; Wang et al., 2025). Remote sensing products supply complementary environmental information, including vegetation indices, land surface temperature, and solar-induced fluorescence that capture canopy status over time. Soil data come from field sampling, regional soil surveys, and derived soil property databases, covering SOC, texture, bulk density, water-holding capacity, and nutrient indicators. In advanced yield-prediction frameworks, these multi-source datasets—yield, weather, soil, and remote sensing—are integrated into unified databases for machine learning or crop model applications.

3.3 Data preprocessing and quality control

Prior to model construction, environmental and yield data require systematic preprocessing to ensure completeness and consistency. Weather station data are screened for missing values, range violations, and temporal or spatial inconsistencies, often using automated quality-control algorithms tailored to agricultural decision needs. Such systems flag implausible measurements—e.g., unrealistic temperature sequences, saturated relative humidity at too low values, or anomalous rainfall series—enabling early detection and correction or removal of erroneous records. For gridded or satellite-based climate products, temporal aggregation (e.g., daily to

monthly) and calculation of growing-season indices are performed to match crop growth stages and modeling time steps. Yield and management records are checked for outliers, coding errors, and inconsistent units across years and locations to avoid bias in training datasets (Archontoulis et al., 2020).

Remote sensing and soil datasets also undergo substantial preprocessing. For optical satellite data, procedures include cloud and shadow masking, compositing, and noise reduction to generate consistent vegetation index and land-surface-temperature time series suitable for yield prediction (Li et al., 2022). Novel image-cleaning techniques, such as quartile-based filtering of local pixel neighborhoods, can reduce sensor noise and atmospheric artifacts, improving the signal-to-noise ratio and enhancing model accuracy when combined with deep learning approaches. Soil property and nutrient data from field sampling or databases are harmonized across sources, interpolated or matched to field or grid units, and normalized or standardized for use in machine learning models that combine soil, climate, and management predictors (Diaz-Gonzalez et al., 2022). Overall, rigorous preprocessing and quality control across all data types are essential to ensure robust, interpretable relationships between soil nutrients, climate variables, and maize yield.

4 Construction and Selection of Feature Variables

4.1 Construction of soil nutrient indicator system

A scientific soil nutrient indicator system should reflect both the supply of key macronutrients and the broader edaphic conditions that control maize response. Long-term omission experiments identify available and total N, P, and K, soil organic carbon, C:N and N:P ratios as primary determinants of yield and nutrient use efficiency, showing that edaphic indicators explain more yield variation than phenological factors in maize systems (Wang et al., 2024). Meta-analysis in northern China further supports including soil organic matter, total N, and available P and K as core indicators, because these properties consistently increase under rational fertilization and are closely aligned with yield gains and water use efficiency (Jiang et al., 2024).

For predictive modeling, soil indicators must also capture spatial heterogeneity and nutrient limitations. Maize nutrient omission trials across 324 farmers' fields in the Eastern Indo-Gangetic Plains showed that soil pH was the most critical variable controlling relative N- and P-limited yields, while soil N and Zn strongly influenced Zn-limited yield (Figure 2) (Ahmed et al., 2024). Post-harvest soil test value prediction equations for N, P, and K demonstrate how pre-sowing soil tests, crop uptake, and fertilizer inputs can be combined to estimate dynamic soil nutrient status, supporting targeted fertilizer recommendations for subsequent crops (Abdel-Salam et al., 2024).

4.2 Extraction of climate variable features

Climate feature construction should represent both mean conditions and stress events during sensitive growth stages. Studies that assessed the relevance of climatic attributes for corn yield found that solar radiation, precipitation, vapor pressure, and maximum and minimum temperature are among the most influential variables, with radiation slightly exceeding precipitation in importance in Neotropical environments (Sierra-Forero et al., 2024). Regional analyses that combine multiple climate time series with yield records confirm that temperature- and water-related indicators together explain a large share of yield variability, especially when evaluated over the growing season (Luthra et al., 2024).

Careful temporal aggregation and transformation of climate variables can greatly improve prediction. Monthly vapor pressure deficit and precipitation expressed with spline functions produced the “best climate-only” model for rainfed corn, with high out-of-sample R^2 , and adding satellite vegetation indices further enhanced performance (Li et al., 2019). Similar work on climate-driven yield variability uses downscaled temperature, precipitation, and shortwave radiation, plus extreme-climate indices, to quantify how mean growing-season warming, radiation changes, and counts of hot or dry days affect maize yield projections (Chen et al., 2020).

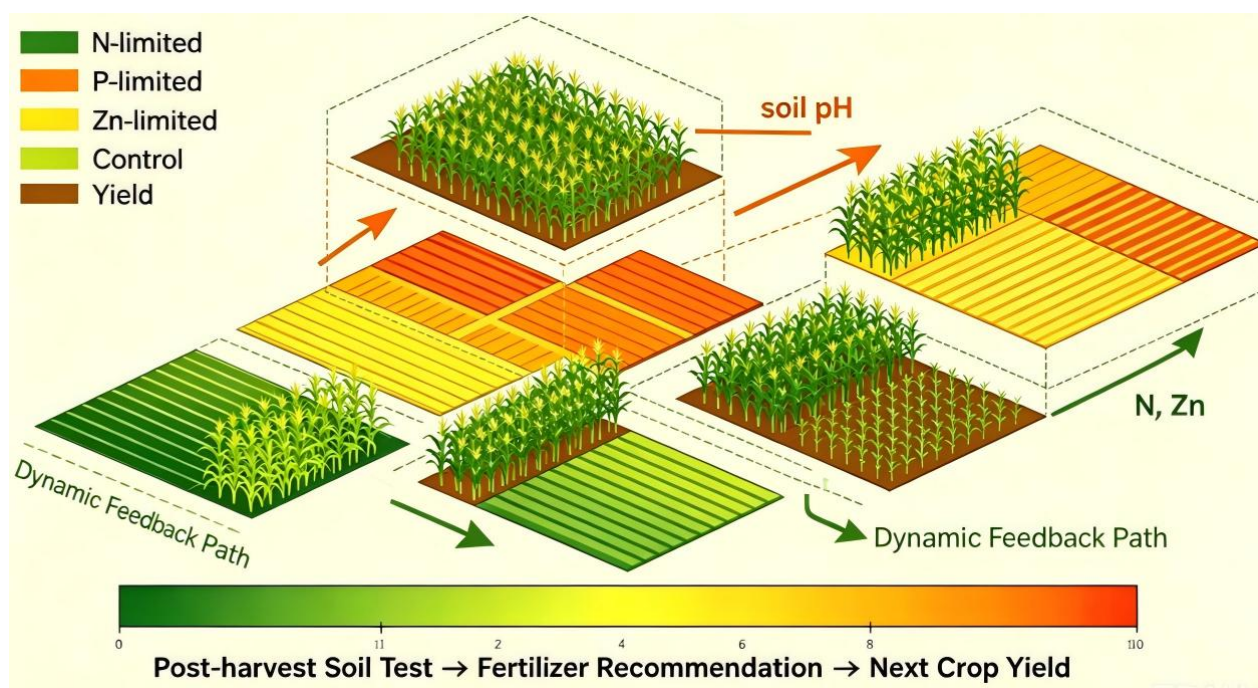


Figure 2 Spatial heterogeneity of soil nutrient limitations and their effects on maize yield

4.3 Feature selection and dimensionality reduction methods

High-dimensional soil-climate datasets require effective feature selection (FS) to avoid overfitting and reduce computational cost. Reviews of machine-learning yield models emphasize that optimal feature sets, obtained by FS, are essential because only a subset of soil, climate, and management variables truly drive prediction accuracy (Hara et al., 2021). In a dedicated framework for yield prediction, a Relief-based FS step was combined with linear discriminant analysis feature extraction, before applying machine-learning classifiers, which markedly improved accuracy over models using all raw variables (Gupta et al., 2022).

Comparative studies of dimensionality reduction for crop yield forecasting show that combining FS and feature extraction (FX) can outperform either alone. In rice yield models based on vegetation and temperature indices, a hybrid approach (FSX) integrating FS with principal component-type FX improved RMSE by up to 60% relative to using all features, and FSX-based models outperformed pure FS or FX in most regions (Pham et al., 2022). More recent works in crop yield prediction apply hybrid FS pipelines (e.g., correlation-based filters, ANOVA, ensemble FS) coupled with advanced learners such as XGBoost or optimized SVR, consistently reporting higher predictive accuracy and lower error once redundant and noisy predictors are removed.

5 Methods for Prediction Model Construction

5.1 Traditional statistical modeling methods

Traditional statistical methods for yield prediction are mainly based on linear or polynomial relationships between yield and a limited set of explanatory variables, often weather indices. Multiple linear regression and its variants have long been used as benchmarks when comparing newer machine learning approaches for maize and other crops, typically using growing-season temperature and precipitation plus a time trend to represent technological progress (Leng and Hall, 2020). Extensions such as quadratic, interaction, and polynomial regression have also been applied to maize and other cereals, and can achieve reasonable accuracy when relationships are approximately linear and the number of predictors is small (Shastry et al., 2017).

More recent work has introduced penalized regression techniques (LASSO, Elastic Net, ridge), which perform variable selection and effectively handle multicollinearity among many weather indices (Vashisth and Aravind, 2026). For maize in semi-arid New Delhi, Elastic Net outperformed stepwise multiple linear regression across vegetative, flowering, and grain-filling stages, with the lowest RMSE and normalized RMSE, highlighting the

value of shrinkage and regularization when many daily weather variables are used. Similar comparisons for rice show that penalized regressions can rival or exceed traditional stepwise regression, though they may still lag behind flexible non-linear models such as neural networks under highly complex climate-yield relationships (Satpathi et al., 2023).

5.2 Machine learning modeling methods

Machine learning (ML) methods such as Random Forest (RF), Support Vector Regression, and boosted trees have become central to crop yield prediction because they capture non-linear responses and interactions between soil, climate, and management variables without strict parametric assumptions. For maize, RF has been shown to outperform multiple linear regression at regional and global scales, reducing RMSE from 14-49% of mean yield with linear models to 6-14% with RF, and better reproducing spatial patterns of yield (Jeong et al., 2016). In the U.S. Midwest, a comparative study using Lasso, Support Vector Regressor, RF, and XGBoost with hundreds of environmental features found that XGBoost was the most accurate and stable algorithm for county-level maize yield prediction (Kang et al., 2020).

In some applications, ML models trained on relatively simple climate inputs also perform strongly. For Irish potato and maize in Rwanda, Random Forest using only rainfall and temperature achieved R^2 values of 0.875 and 0.817, respectively, outperforming polynomial regression and Support Vector Regressor and providing practically useful early-season predictions (Kuradusenge et al., 2023). ML has also been used to model silage maize yields from NDVI time-series; boosted regression trees and RF achieved correlations above 0.87, and were less sensitive to inconsistencies in satellite-derived vegetation profiles than conventional regressions (Aghighi et al., 2018). These studies underline the versatility of ML methods for integrating climate, soil, and remote-sensing predictors in maize yield models.

5.3 Deep learning and ensemble learning methods

Deep learning (DL) extends ML by learning complex, hierarchical representations from large, high-dimensional datasets composed of weather, soil, genotype, and remote sensing inputs. A deep neural network trained on thousands of maize hybrid trials across more than 2,000 locations substantially outperformed Lasso, shallow neural networks, and regression trees, reaching an RMSE close to 11-12% of average yield while also supporting feature selection to reduce input dimensionality with minimal accuracy loss (Khaki and Wang, 2019). However, DL does not always dominate: in a U.S. Midwest maize study, LSTM and CNN architectures did not surpass XGBoost, suggesting that tabular environmental datasets may not always benefit from image- or sequence-oriented deep architectures (Kang et al., 2020).

Ensemble learning combines multiple base learners to improve robustness and accuracy. For corn in the U.S. Corn Belt, CNN-DNN ensembles created via bagging and stacking outperformed ensembles of linear regression, Lasso, RF, XGBoost, and LightGBM, explaining about 77% of spatio-temporal yield variation with an RMSE of 866 kg/ha (Shahhosseini et al., 2021). Hybrid and ensemble DL frameworks that fuse convolutional, recurrent, and fully connected networks have also shown superior performance for crop yield prediction, with CNN-DNN or CNN-RNN-LSTM structures often exceeding single DL or ML models and achieving R^2 values near or above 0.85 in case studies (Oikonomidis et al., 2022). Deep ensemble approaches thus offer a promising route for integrating multi-source soil, climate, and remote-sensing data to achieve robust maize yield prediction under variable environments.

6 Model Training and Evaluation System

6.1 Dataset partitioning and validation strategies

A reasonable partition of the maize yield dataset is the basis for constructing reliable prediction models. In most supervised learning settings, data are divided into training, validation, and test subsets so that model fitting, hyperparameter tuning, and final performance assessment can be clearly separated and avoid information leakage (Bischl et al., 2021). When the number of yearly observations is small, directly reserving an independent test set

becomes difficult, and specialized cross-validation (CV) schemes such as leave-one-out (LOO) or nested CV are recommended to obtain unbiased generalization estimates (Dinh and Aires, 2022).

For crop yield prediction with strong spatial and temporal dependence, the choice of CV strategy affects both apparent skill and interpretability. Studies using simulated or observed yields show that random CV can give overly optimistic accuracy when neighboring samples are highly correlated, while spatial or cluster-based CV provides more realistic estimates on held-out regions (Radočaj et al., 2025). Nested CV or nested leave-two-out schemes further separate inner folds for model selection from outer folds for performance estimation, preventing overly complex models from being chosen and improving transferability across years and locations (Sweet et al., 2023).

6.2 Model parameter optimization methods

Hyperparameters of machine learning models, such as the number of trees in random forests or learning rates in gradient boosting, strongly influence predictive performance and must be tuned systematically rather than by ad-hoc trial-and-error (Bischl et al., 2021). Classical search strategies include grid search and random search, which evaluate candidate configurations on resampling-based performance estimates, but they become inefficient as the hyperparameter space grows.

More advanced approaches treat hyperparameter tuning as a black-box optimization problem and use probabilistic surrogate models. Bayesian optimization with Gaussian processes or related surrogates iteratively proposes promising configurations based on past evaluations and has been shown to find better settings than random search under comparable budgets (Wu et al., 2019). In crop yield estimation, Bayesian optimization frameworks applied to tree-based models such as LightGBM achieve high coefficients of determination and low mean squared error across several agricultural datasets, demonstrating the gains from automated hyperparameter optimization. Random forest-specific tuning via model-based optimization (e.g., tuning mtry, node size, sample size) can further increase accuracy over default settings while controlling runtime (Probst et al., 2018).

6.3 Model evaluation indicator system

Because maize yield prediction is a regression problem, a comprehensive indicator system is needed to evaluate both accuracy and explanatory power. Error-based metrics such as root mean square error (RMSE), mean absolute error (MAE), and related deviations are widely used in crop model evaluation because they directly characterize the magnitude of prediction errors in yield units (Yang et al., 2014). RMSE is particularly sensitive to large errors and is appropriate when error distributions are approximately Gaussian, whereas MAE provides a more robust and interpretable measure of average error and is less influenced by outliers (Chai and Draxler, 2014).

To complement absolute error measures, goodness-of-fit and efficiency statistics assess how much of the observed variance is explained by the model. The coefficient of determination (R^2) is often preferred as a standard metric in regression because it relates performance to the variance of ground-truth yields and is more informative than stand-alone error magnitudes in many applications (Chicco et al., 2021). In process-based crop modeling, additional indices such as modeling efficiency (EF) and the index of agreement (d) are used alongside RMSE and MAE to provide a balanced view of model bias, dispersion, and agreement with observations (Yang et al., 2014). For maize yield prediction models based on soil nutrients and climate variables, combining R^2 (or EF) with RMSE and MAE yields a robust evaluation framework that captures both accuracy and reliability across different environments.

7 Case Study: Empirical Analysis of Regional Maize Yield Prediction

7.1 Study area and sample construction

In many recent maize yield prediction studies, the study area is defined to capture both environmental gradients and management diversity so that models generalize beyond a single field or season. For example, plot-scale work integrates multi-year trials under contrasting fertilizer systems, combining climate, soil, and satellite data to represent heterogeneous growing conditions across years and treatments (Meng et al., 2021). Similar multi-farm designs in Western Australia aggregate yield monitor data from thousands of hectares over several seasons, then

collocate each observation with soil, terrain, and weather variables to form a dense spatio-temporal sample set (Filippi et al., 2019).

Large-area studies, such as county-level maize analyses in the US Midwest or regional work in Northeast China, construct samples by merging official yield statistics with gridded or station-based climate data, soil maps, and multi-source satellite products (Figure 3) (Kang et al., 2020; Li et al., 2022). In Ghana, plot-level samples from hundreds of maize field trials are georeferenced and linked to 0-30 cm soil properties, climate variables during the planting season, and management practices, enabling model training across wide environmental and agronomic ranges (Asamoah et al., 2024).

Corn Yield Prediction Sample Construction Process

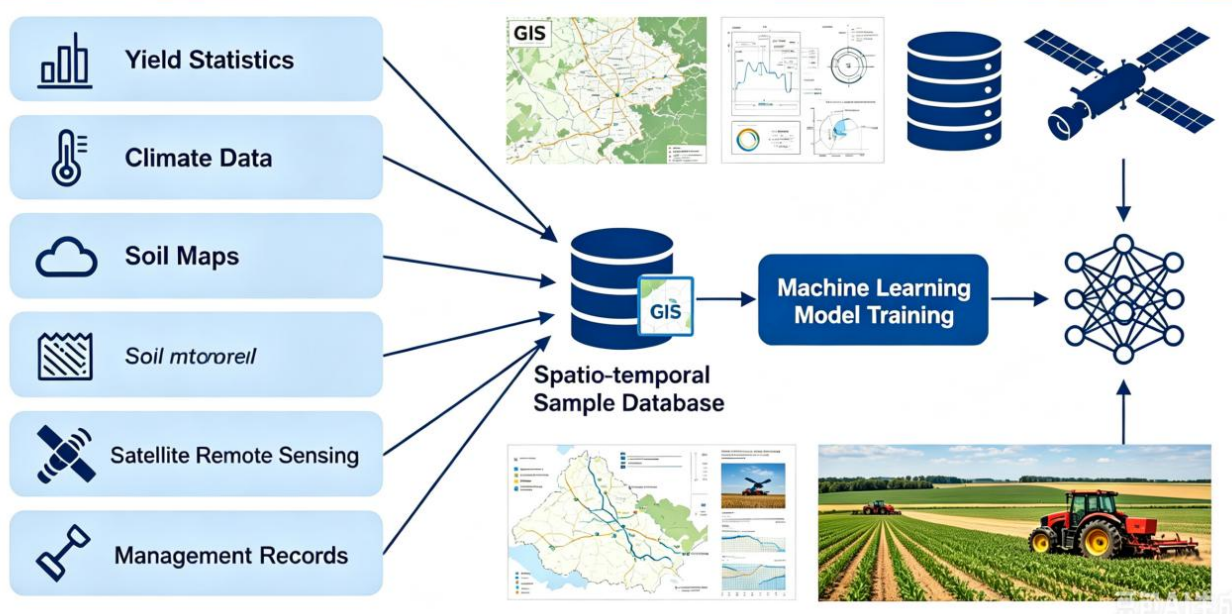


Figure 3 Workflow for integrating multi-source environmental and agricultural datasets into maize yield prediction samples

7.2 Comparative Analysis Of Multi-Model Prediction Results

Comparative studies consistently show that model performance depends strongly on algorithm choice and input richness. At the plot scale, combining vegetation indices, climate, soil, and fertilizer data, Random Forest and Adaptive Boosting clearly outperform linear regression, SVM, GPR, and KNN, with R^2 often above 0.85 and lowest RMSE values (Meng et al., 2021). In a Hungarian field using detailed spatio-temporal soil and micro-relief measurements, XGBoost surpassed neural and kernel methods, reaching test accuracies above 95%, while lattice-based smoothing further improved predictive AUC (Nyéki et al., 2021).

At regional scales, ensemble or tree-based machine learning models generally outperform both traditional regression and deep learning architectures. In the US Midwest, XGBoost provided the most accurate and stable county-level maize forecasts when hundreds of environmental features were used, while LSTM and CNN did not show clear advantages (Kang et al., 2020). Across Northeast China, an ensemble of several ML methods improved yield prediction over individual linear and ML models when integrating environmental and multi-sensor satellite data, explaining more than 70% of maize yield variability (Li et al., 2022).

7.3 Result validation and agricultural application analysis

Robust validation is essential to ensure that multi-model predictions have practical value. Studies highlight that naïve random data splits can substantially overestimate predictive skill, especially when the goal is true forecasting rather than interpolation within a season (Morales and Villalobos, 2023). More rigorous schemes, such as nested k-fold cross-validation across years and fields, or leave-one-field/leave-one-year-out designs, better

reflect operational performance and were used, for instance, in multi-farm machine-learning models and in Ghanaian RF models for maize yield and agronomic efficiency (Filippi et al., 2019; Asamoah et al., 2024).

When rigorously validated, yield prediction models support several agricultural applications. Plot-scale maize models that accurately forecast yield under different fertilizer systems enable assessment of input strategies and refinement of site-specific recommendations before harvest (Meng et al., 2021). Large-area models that integrate climate, soil, and satellite indicators have been used for early-season yield forecasting, outperforming official forecasts and providing actionable information for logistics, market planning, and food-security assessments (Li et al., 2022). Such applications demonstrate how reliable maize yield prediction, grounded in soil-climate interactions, can inform precision fertilization, risk management, and regional policy decisions.

8 Results Analysis and Discussion

8.1 Contribution analysis of soil nutrient variables

Feature-importance and interpretable ML studies highlight that specific soil nutrients can dominate maize yield responses, even in data-rich settings. In a data-intensive farm management trial, Random Forest analysis showed that urea application was consistently the most critical variable for explaining spatial yield variation, with soil phosphorus, pH, clay content, sodium and plant population also among the leading contributors in different seasons (Maseko et al., 2024). This indicates that both applied N and inherent soil fertility properties jointly control yield in high-resolution, within-field prediction. Similar work in precision agriculture, using RF and other models on over 145,000 corn and soybean yield observations, found that soil test P, K, Zn, soil organic matter and cation exchange capacity were key predictors, underscoring the strong explanatory power of nutrient and related soil indicators for yield variation at sub-field scales (Burdett and Wellen, 2022).

Under nutrient-limited conditions, omission trials combined with AutoML provide a more explicit decomposition of nutrient contributions. In 324 nutrient omission plot trials across ten agroecological zones in the Eastern Indo-Gangetic Plains, stack-ensemble and deep learning models predicted relative nutrient-limited yields with low RMSE, and permutation importance identified soil pH as the dominant variable controlling N- and P-limited yields (Ahmed et al., 2024). The same analysis showed that soil N and Zn strongly influenced Zn-limited yield, while spatial trends in K-limited yield emerged along an east-west gradient, revealing distinct fertility constraints for different nutrients. These findings suggest that soil nutrient variables-especially applied N, soil P, Zn, pH and texture-related properties-provide high marginal gains in predictive power and are indispensable components of maize yield models based on soil-climate interactions.

8.2 Influence weight analysis of climate variables

Across diverse modeling frameworks, climate variables frequently emerge as the largest single contributors to interannual maize yield variability. A global meta-analysis using 68 simulation studies for wheat, maize and rice showed that maximum temperature and precipitation significantly affected yield responses, with yields declining by 4.21% per 1 °C increase in maximum temperature but increasing by 0.43% per 1% rise in precipitation (Qin et al., 2023). This quantitative gradient highlights the high negative weight of heat stress and the compensating effect of adequate rainfall in crop-climate response functions. At the global scale, mixed-effects models updating projected yield responses under CMIP6 scenarios indicate that temperature-related stress is a dominant driver of future maize yield losses, with projected global maize declines around 22% by late century under high emissions if adaptation is limited (Li et al., 2025).

Machine-learning-based attribution provides more detailed rankings of individual climate indicators. A hybrid GGCM-Random Forest framework for China's maize belt found that chilling days, drought indicators and crop pests/diseases were the main factors influencing projected maize yield changes, with relative importance quantified via RF partial-dependence analysis (Li et al., 2023). In a separate process-based and ML study on wheat under future climate scenarios, precipitation explained most yield variability in mid-century high-emission conditions, whereas maximum temperature became the dominant limiting factor under later, more strongly warmed scenarios (El-Mahroug et al., 2025). For site-specific maize prediction with spatio-temporal XGBoost models, precipitation during the juvenile growth phase (May) was identified as the single most important factor

over five years, followed by soil pH, clay content, electrical conductivity and NDVI, again emphasizing the high influence weight of water-related variables alongside key soil properties.

8.3 Discussion on model applicability and uncertainty

The applicability of soil-nutrient- and climate-based yield models depends critically on how uncertainty is handled across space, time and scenario conditions. A recent meta-analysis of crop yield responses to projected climate change combined mixed-effects modeling with block bootstrapping to partition uncertainty arising from model structure, climate projections (CMIP6) and emissions pathways, showing that simple pooled OLS tends to underestimate yield losses and under-represent uncertainty ranges (Li et al., 2025). Similarly, a crop-model and ML ensemble for maize and soybean across China demonstrated that coupling GGCMs with Random Forest greatly improved correlation (r up to 0.77 for maize) and reduced normalized RMSE, while variance decomposition revealed that the dominant uncertainty source shifted from crop models in the baseline GGCM runs to global climate models and then scenarios as projections extended further into the century (Li et al., 2023). These results imply that model applicability under future climates requires explicit accounting for structural, climate and scenario uncertainties rather than relying on single-model projections.

Transferability across domains and scales introduces additional uncertainty dimensions for data-driven yield models. Domain-adaptation work on maize in the US Corn Belt, using DANN, KLIEP and RTNN, found that models trained in temperate regions with medium-high growing degree days and moderate vapor pressure deficit generalized well, whereas strong dependence on vegetation indices (GCI) reduced transferability when source and target domains had limited overlap (Priyatikanto et al., 2023). Independent evaluations of cross-validation strategies in UAV-based yield prediction further showed that random CV can substantially overestimate performance when models are applied outside their training spatial domain, whereas spatial or leave-one-field-out CV and simpler, regularized models gave more realistic extrapolation accuracy (Habibi et al., 2024). Together with county-scale ensemble studies that link large prediction errors to low cropland ratios and extreme weather events (Sajid et al., 2022), these findings stress that robust maize yield prediction demands careful validation design, domain-aware training, and transparent uncertainty quantification before models are applied for management or policy decisions in new regions or under novel climate conditions.

9 Conclusions and Future Research Directions

Existing studies confirm that integrating soil nutrients, soil physical properties, and climate variables can explain a substantial share of maize yield variability across diverse agroecological zones. Soil indicators such as nitrogen fertilizer rate, soil organic carbon, pH, bulk density, and exchangeable bases consistently emerge among the most influential predictors, often exceeding the importance of individual climate variables for yield prediction in tropical and semi-arid environments. At the same time, temperature, rainfall, and related weather indices remain key drivers of interannual variation, especially when combined with management and genotype information in large datasets. From a modeling perspective, tree-based and boosting algorithms (Random Forest, XGBoost, Gradient Boosting) generally outperform linear methods and many deep architectures for maize yield prediction using soil-climate feature sets. Meta-modeling of process-based simulations and large empirical trial datasets shows that these methods can achieve relative errors around 10-15% when sufficient training samples and well-designed features are available. Systematic reviews across maize and other crops further indicate that these algorithms are among the most frequently adopted and robust options, particularly when coupled with feature engineering and multimodal data integration.

High-accuracy soil-climate yield models provide actionable information for fertilizer management and nutrient efficiency. In Ghana, Random Forest and XGBoost models trained on long-term maize trials successfully predicted both yield and agronomic efficiency, highlighting nitrogen rate, rainfall, and key soil properties as dominant management levers. Such models support the design of site-specific recommendations that can raise productivity while reducing the environmental costs of blanket fertilizer application. Similar ML-process-model hybrids using APSIM outputs demonstrate that meta-models can rapidly explore genotype-environment-management scenarios for pre-season planning. At larger scales, integrating soil maps,

meteorological series, and satellite indicators enables early-season forecasts that outperform conventional statistical baselines and even some official forecasts. County-level yield prediction in the U.S. Midwest has shown that XGBoost models using hundreds of environmental features can provide reliable maize forecasts several months before harvest, improving on models based only on basic weather or historical yields. Reviews of precision agriculture emphasize that such predictive systems contribute to resource optimization, risk management, and food-security planning by linking sensing technologies, big data platforms, and advanced analytics into operational decision support tools.

Despite these advances, several limitations constrain the reliability and transferability of current soil-climate yield models. Studies comparing algorithms against simple baselines show that, under realistic forecasting setups using ordered train-test splits, ML models sometimes offer only modest gains over farm-level average yields, especially when weather forecast errors are ignored. Systematic reviews also highlight persistent challenges with obtaining high-quality, harmonized datasets on soil nutrients, management, and high-resolution yields, which can limit model generalization across regions and seasons. In addition, many models are trained and validated under random data partitioning, leading to over-optimistic performance estimates for true out-of-sample prediction. Future research directions point toward hybrid, transferable, and explainable frameworks. Hybrid models that couple process-based crop simulators with ML or deep learning have improved accuracy and reduced uncertainty in semi-arid maize systems, particularly when fusing remote sensing, climate, and soil information. Domain adaptation and transfer-learning approaches, including partial adversarial networks, are beginning to address domain shifts between ecological zones and could substantially improve cross-regional maize yield prediction. Reviews stress the need for standardized data protocols, interpretable architectures (e.g., SHAP- or XAI-enhanced models), and scalable, crop-agnostic pipelines so that soil nutrient and climate-based yield prediction can be robustly embedded in precision agriculture and sustainability strategies.

Acknowledgments

We would like to thank the anonymous reviewers for their detailed review of the draft. Their specific feedback helped us correct the logical loopholes in our arguments.

Conflict of Interest Disclosure

The authors affirm that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Abdel-Salam M., Kumar N., and Mahajan S., 2024, A proposed framework for crop yield prediction using hybrid feature selection approach and optimized machine learning, *Neural Computing and Applications*, 36: 20723-20750.
<https://doi.org/10.1007/s00521-024-10226-x>
- Aghighi H., Azadbakht M., Ashourloo D., Shahrabi H., and Radiom S., 2018, Machine learning regression techniques for the silage maize yield prediction using time-series images of Landsat 8 OLI, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11: 4563-4577.
<https://doi.org/10.1109/jstars.2018.2823361>
- Ahmed Z., Krupnik T., Timsina J., Islam S., Hossain K., Kurishi A., Emran S., Harun-Ar-Rashid M., McDonald A., and Gathala M., 2024, Prediction of spatial heterogeneity in nutrient-limited sub-tropical maize yield: implications for precision management in the eastern indo-gangetic plains, *Artificial Intelligence in Agriculture*, 12: 1-15.
<https://doi.org/10.1016/j.aiia.2024.08.001>
- Archontoulis S., Castellano M., Licht M., Nichols V., Baum M., Huber I., Martinez-Feria R., Puntel L., Ordóñez R., Iqbal J., Wright E., Dietzel R., Helmers M., Vanlooche A., Liebman M., Hatfield J., Herzmann D., Córdova S., Edmonds P., Togliatti K., Kessler A., Danalatos G., Pasley H., Pederson C., and Lamkey K., 2020, Predicting crop yields and soil-plant nitrogen dynamics in the US Corn Belt, *Crop Science*, 60: 721-738.
<https://doi.org/10.1002/csc2.20039>
- Asamoah E., Heuvelink G., Chairi I., Bindraban P., and Logah V., 2024, Random forest machine learning for maize yield and agronomic efficiency prediction in Ghana, *Heliyon*, 10: e37065.
<https://doi.org/10.1016/j.heliyon.2024.e37065>
- Bischl B., Binder M., Lang M., Pielok T., Richter J., Coors S., Thomas J., Ullmann T., Becker M., Boulesteix A., Deng D., and Lindauer M., 2021, Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 13(2): e1484.
<https://doi.org/10.1002/widm.1484>

- Burdett H., and Wellen C., 2022, Statistical and machine learning methods for crop yield prediction in the context of precision agriculture, *Precision Agriculture*, 23: 1553-1574.
<https://doi.org/10.1007/s11119-022-09897-0>
- Chai T., and Draxler R., 2014, Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature, *Geoscientific Model Development*, 7: 1247-1250.
<https://doi.org/10.5194/gmd-7-1247-2014>
- Chen F., Xu X., Chen S., Wang Z., Wang B., Zhang Y., Zhang C., Feng P., and Hu K., 2024, Soil buffering capacity enhances maize yield resilience amidst climate perturbations, *Agricultural Systems*, 222: 103870.
<https://doi.org/10.1016/j.agsy.2024.103870>
- Chen X., Wang L., Niu Z., Zhang M., Li C., and Li J., 2020, The effects of projected climate change and extreme climate on maize and rice in the Yangtze River Basin, China, *Agricultural and Forest Meteorology*, 282-283: 107867.
<https://doi.org/10.1016/j.agrformet.2019.107867>
- Chicco D., Warrens M., and Jurman G., 2021, The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation, *PeerJ Computer Science*, 7: e623.
<https://doi.org/10.7717/peerj-cs.623>
- Dandriofosse S., Jago A., Huart J., Michaud V., Planchon V., and Rosillon D., 2024, Automatic quality control of weather data for timely decisions in agriculture, *Smart Agricultural Technology*, 8: 100445.
<https://doi.org/10.1016/j.atech.2024.100445>
- Diaz-Gonzalez F., Vuelvas J., Correa C., Vallejo V., and Patiño D., 2022, Machine learning and remote sensing techniques applied to estimate soil indicators - Review, *Ecological Indicators*, 135: 108517.
<https://doi.org/10.1016/j.ecolind.2021.108517>
- Dinh T., and Aires F., 2022, Nested leave-two-out cross-validation for the optimal crop yield model selection, *Geoscientific Model Development*, 15: 3519-3536.
<https://doi.org/10.5194/gmd-15-3519-2022>
- El-Mahroug S., Suleiman A., Zoubi M., Al-Omari S., Abu-Afifeh Q., Al-Jawaldeh H., Alta'any Y., Al-Nawaiseh T., Obeidat N., Alsoud S., Alshoshan A., Al-Shibli F., and Ta'any R., 2025, Predictive modeling of climate-driven crop yield variability using DSSAT towards sustainable agriculture, *AgriEngineering*, 7(5): 156.
<https://doi.org/10.3390/agriengineering7050156>
- Falconnier G., Corbeels M., Boote K., Affholder F., Adam M., MacCarthy D., Ruane A., Nendel C., Whitbread A., Justes É., Ahuja L., Akinseye F., Alou I., Amouzou K., Anapalli S., Baron C., Basso B., Baudron F., Bertuzzi P., Challinor A., Chen Y., Deryng D., Elsayed M., Faye B., Gaiser T., Galdos M., Gayler S., Gérardaux E., Giner M., Grant B., Hoogenboom G., Ibrahim E., Kamali B., Kersebaum K., Kim S., Laan M., Leroux L., Lizaso J., Maestrini B., Meier E., Mequanint F., Ndoli A., Porter C., Priesack E., Ripoche D., Sida T., Singh U., Smith W., Srivastava A., Sinha S., Tao F., Thorburn P., Timlin D., Traoré B., Twine T., and Webber H., 2020, Modelling climate change impacts on maize yields under low nitrogen input conditions in sub-Saharan Africa, *Global Change Biology*, 26: 5942-5964.
<https://doi.org/10.1111/gcb.15261>
- Feng P., Wang B., Harrison M., Wang J., Liu K., Huang M., Liu D., Yu Q., and Hu K., 2022, Soil properties resulting in superior maize yields upon climate warming, *Agronomy for Sustainable Development*, 42(5): 81.
<https://doi.org/10.1007/s13593-022-00818-z>
- Filippi P., Jones E., Wimalathunge N., Somarathna P., Pozza L., Ugbaje S., Jephcott T., Paterson S., Whelan B., and Bishop T., 2019, An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning, *Precision Agriculture*, 20(5): 1015-1029.
<https://doi.org/10.1007/s11119-018-09628-4>
- Gupta S., Geetha A., Sankaran K., Zamani A., Ritonga M., Raj R., Ray S., and Mohammed H., 2022, Machine learning- and feature selection-enabled framework for accurate crop yield prediction, *Journal of Food Quality*, 2022: 6293985.
<https://doi.org/10.1155/2022/6293985>
- Habibi L., Matsui T., and Tanaka T., 2024, Critical evaluation of the effects of a cross-validation strategy and machine learning optimization on the prediction accuracy and transferability of a soybean yield prediction model using UAV-based remote sensing, *Journal of Agriculture and Food Research*, 18: 101096.
<https://doi.org/10.1016/j.jafr.2024.101096>
- Hara P., Piekutowska M., and Niedbała G., 2021, Selection of independent variables for crop yield prediction using artificial neural network models with remote sensing data, *Land*, 10(6): 609.
<https://doi.org/10.3390/land10060609>
- Jeong J., Resop J., Mueller N., Fleisher D., Yun K., Butler E., Timlin D., Shim K., Gerber J., Reddy V., and Kim S., 2016, Random forests for global and regional crop yield predictions, *PLoS ONE*, 11(6): e0156571.
<https://doi.org/10.1371/journal.pone.0156571>
- Jiang M., Dong C., Bian W., Zhang W., and Wang Y., 2024, Effects of different fertilization practices on maize yield, soil nutrients, soil moisture, and water use efficiency in northern China based on a meta-analysis, *Scientific Reports*, 14: 57031.
<https://doi.org/10.1038/s41598-024-57031-z>

- Kaleri A., Khanzada B., Rajput W., Bijarani A., Shafqat A., Arain A., Mirbahar S., Jokhio N., Majeedano A., and Majeedano S., 2026, Combined effects of nitrogen, phosphorus, and potassium on maize growth, development, and yield, *Jammu Kashmir Journal of Agriculture*, 5(3): 297-305.
<https://doi.org/10.56810/jkjagri.005.03.0297>
- Kang Y., Ozdogan M., Zhu X., Ye Z., Hain C., and 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.
<https://doi.org/10.1088/1748-9326/ab7df9>
- Khaki S., and Wang L., 2019, Crop yield prediction using deep neural networks, *Frontiers in Plant Science*, 10: 621.
<https://doi.org/10.3389/fpls.2019.00621>
- Kim K., and Lee B., 2023, Effects of climate change and drought tolerance on maize growth, *Plants*, 12(20): 3548.
<https://doi.org/10.3390/plants12203548>
- Kuradusenge M., Hitimana E., Hanyurwimfura D., Rukundo P., Mtonga K., Mukasine A., Uwitonze C., Ngabonziza J., and Uwamahoro A., 2023, Crop yield prediction using machine learning models: Case of Irish potato and maize, *Agriculture*, 13(1): 225.
<https://doi.org/10.3390/agriculture13010225>
- Leng G., and Hall J., 2020, Predicting spatial and temporal variability in crop yields: An inter-comparison of machine learning, regression and process-based models, *Environmental Research Letters*, 15(4): 044027.
<https://doi.org/10.1088/1748-9326/ab7b24>
- Li C., Camac J., Robinson A., and Kompas T., 2025, Predicting changes in agricultural yields under climate change scenarios and their implications for global food security, *Scientific Reports*, 15: 87047.
<https://doi.org/10.1038/s41598-025-87047-y>
- Li E., Zhao J., Pullens J., and Yang X., 2021, The compound effects of drought and high temperature stresses will be the main constraints on maize yield in Northeast China, *Science of the Total Environment*, 812: 152461.
<https://doi.org/10.1016/j.scitotenv.2021.152461>
- Li L., Zhang Y., Wang B., Feng P., He Q., Shi Y., Liu K., Harrison M., Liu D., Yao N., Li Y., He J., Feng H., Siddique K., and Yu Q., 2023, Integrating machine learning and environmental variables to constrain uncertainty in crop yield change projections under climate change, *European Journal of Agronomy*, 151: 126917.
<https://doi.org/10.1016/j.eja.2023.126917>
- Li Y., Guan K., Yu A., Peng B., Zhao L., Li B., and Peng J., 2019, Toward building a transparent statistical model for improving crop yield prediction: Modeling rainfed corn in the U.S., *Field Crops Research*, 234: 55-65.
<https://doi.org/10.1016/j.fcr.2019.02.005>
- Li Z., Ding L., and Xu D., 2022, Exploring the potential role of environmental and multi-source satellite data in crop yield prediction across Northeast China, *Science of the Total Environment*, 806: 152880.
<https://doi.org/10.1016/j.scitotenv.2021.152880>
- Luthra N., Srivastava A., Shahi U., Singh V., Dey P., and Singh A., 2024, Prediction of post-harvest soil nutrient status through multiple linear regression for targeted yield of hybrid maize, *Indian Journal of Agronomy*, 68(4): 547-553.
<https://doi.org/10.59797/ija.v68i4.5471>
- Maseko S., Van Der Laan M., Tesfamariam E., Delpont M., and Otterman H., 2024, Evaluating machine learning models and identifying key factors influencing spatial maize yield predictions in data intensive farm management, *European Journal of Agronomy*, 160: 127193.
<https://doi.org/10.1016/j.eja.2024.127193>
- Matiu M., Ankerst D., and Menzel A., 2017, Interactions between temperature and drought in global and regional crop yield variability during 1961-2014, *PLoS One*, 12(5): e0178339.
<https://doi.org/10.1371/journal.pone.0178339>
- Medina H., and Tian D., 2023, Synergistic contributions of climate and management intensifications to maize yield trends from 1961 to 2017, *Environmental Research Letters*, 18(3): 034021.
<https://doi.org/10.1088/1748-9326/acb27f>
- Meng L., Liu H., Ustin S., and Zhang X., 2021, Predicting maize yield at the plot scale of different fertilizer systems by multi-source data and machine learning methods, *Remote Sensing*, 13(18): 3760.
<https://doi.org/10.3390/rs13183760>
- Morales A., and Villalobos F., 2023, Using machine learning for crop yield prediction in the past or the future, *Frontiers in Plant Science*, 14: 1128388.
<https://doi.org/10.3389/fpls.2023.1128388>
- Nyéki A., Kerepesi C., Daróczy B., Benczúr A., Milics G., Nagy J., Harsányi E., Kovács A., and Neményi M., 2021, Application of spatio-temporal data in site-specific maize yield prediction with machine learning methods, *Precision Agriculture*, 22: 1397-1415.
<https://doi.org/10.1007/s11119-021-09833-8>
- Ocwa A., Harsányi E., Széles A., Holb I., Szabó S., Rátonyi T., and Mohammed S., 2023, A bibliographic review of climate change and fertilization as the main drivers of maize yield: Implications for food security, *Agriculture and Food Security*, 12(1): 19.
<https://doi.org/10.1186/s40066-023-00419-3>
- Oikonomidis A., Catal C., and Kassahun A., 2022, Hybrid deep learning-based models for crop yield prediction, *Applied Artificial Intelligence*, 36(1): 2031823.
<https://doi.org/10.1080/08839514.2022.2031823>

- Pham H., Awange J., and Kuhn M., 2022, Evaluation of three feature dimension reduction techniques for machine learning-based crop yield prediction models, *Sensors*, 22(17): 6609.
<https://doi.org/10.3390/s22176609>
- Priyatikanto R., Lu Y., Dash J., and Sheffield J., 2023, Improving generalisability and transferability of machine-learning-based maize yield prediction model through domain adaptation, *SSRN Electronic Journal*, 1: 1-29.
<https://doi.org/10.2139/ssrn.4122021>
- Probst P., Wright M., and Boulesteix A., 2018, Hyperparameters and tuning strategies for random forest, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(3): e1301.
<https://doi.org/10.1002/widm.1301>
- Qian Y., Zhang Z., Jiang F., Wang J., Dong F., Liu J., and Peng X., 2025, Impacts of tillage treatments on soil physical properties and maize growth at two sites under different climatic conditions in black soil region of Northeast China, *Soil and Tillage Research*, 257: 106471.
<https://doi.org/10.1016/j.still.2025.106471>
- Qin M., Zheng E., Hou D., Meng X., Meng F., Gao Y., Chen P., Qi Z., and Xu T., 2023, Response of wheat, maize, and rice to changes in temperature, precipitation, CO₂ concentration, and uncertainty based on crop simulation approaches, *Plants*, 12(14): 2709.
<https://doi.org/10.3390/plants12142709>
- Radočaj D., Plaščak I., and Jurišić M., 2025, A comparative assessment of regular and spatial cross-validation in subfield machine learning prediction of maize yield from Sentinel-2 phenology, *Eng*, 6(10): 270.
<https://doi.org/10.3390/eng6100270>
- Satpathi A., Setiya P., Das B., Nain A., Jha P., Singh S., and Singh S., 2023, Comparative analysis of statistical and machine learning techniques for rice yield forecasting for Chhattisgarh, India, *Sustainability*, 15(3): 2786.
<https://doi.org/10.3390/su15032786>
- Shahhosseini M., Hu G., Khaki S., and Archontoulis S., 2021, Corn yield prediction with ensemble CNN-DNN, *Frontiers in Plant Science*, 12: 709008.
<https://doi.org/10.3389/fpls.2021.709008>
- Shastry A., Sanjay H., and Bhanusree E., 2017, Prediction of crop yield using regression techniques, *International Journal of Computing*, 6(5): 1-5.
- Sierra-Forero B., Barón-Velandia J., and Vanegas-Ayala S., 2024, Assessment of the relevance of features associated with corn crop yield prediction in Colombia, a country in the Neotropical zone, *International Journal of Information Technology*, 16: 2129-2138.
<https://doi.org/10.1007/s41870-024-01762-9>
- Sun Z., Yang R., Wang J., Zhou P., Gong Y., Gao F., and Wang C., 2024, Effects of nutrient deficiency on crop yield and soil nutrients under winter wheat-summer maize rotation system in the North China Plain, *Agronomy*, 14(11): 2690.
<https://doi.org/10.3390/agronomy14112690>
- Sweet L., Müller C., Anand M., and Zscheischler J., 2023, Cross-validation strategy impacts the performance and interpretation of machine learning models, *Artificial Intelligence for the Earth Systems*, 2(4): e230026.
<https://doi.org/10.1175/aies-d-23-0026.1>
- Vashisth A., and Aravind K., 2026, Maize yield estimation at different growth stage using weather variables by LASSO, elastic net and stepwise multiple linear regression techniques, *Scientific Reports*, 16: 34239.
<https://doi.org/10.1038/s41598-025-34239-1>
- Vogel E., Donat M., Alexander L., Meinshausen M., Ray D., Karoly D., Meinshausen N., and Frieler K., 2019, The effects of climate extremes on global agricultural yields, *Environmental Research Letters*, 14(5): 054010.
<https://doi.org/10.1088/1748-9326/ab154b>
- Wang N., Ai Z., Zhang Q., Leng P., Qiao Y., Li Z., Tian C., Cheng H., Chen G., and Li F., 2024, Impacts of nitrogen (N), phosphorus (P), and potassium (K) fertilizers on maize yields, nutrient use efficiency, and soil nutrient balance: Insights from a long-term diverse NPK omission experiment in the North China Plain, *Field Crops Research*, 317: 109616.
<https://doi.org/10.1016/j.fcr.2024.109616>
- Wang X., Li X., Lou Y., You S., and Zhao H., 2024, Refined evaluation of climate suitability of maize at various growth stages in major maize-producing areas in the North of China, *Agronomy*, 14(2): 344.
<https://doi.org/10.3390/agronomy14020344>
- Wang Y., Shen Y., Yu S., Zhang X., and Xiao D., 2025, Climate extremes are critical to maize yield and will be severer in North China, *Climate Risk Management*, 47: 100710.
<https://doi.org/10.1016/j.crm.2025.100710>
- Wu J., Chen X., Zhang H., Xiong L., Lei H., and Deng S., 2019, Hyperparameter optimization for machine learning models based on Bayesian optimization, *Journal of Electronic Science and Technology*, 17(1): 26-40.
<https://doi.org/10.11989/jest.1674-862x.80904120>
- Yang J., Yang J., Liu S., and Hoogenboom G., 2014, An evaluation of the statistical methods for testing the performance of crop models with observed data, *Agricultural Systems*, 127: 81-89.
<https://doi.org/10.1016/j.agsy.2014.01.008>
- Zhao F., Wang G., Li S., Hagan D., and Ullah W., 2023, The combined effects of VPD and soil moisture on historical maize yield and prediction in China, *Frontiers in Environmental Science*, 11: 1117184.
<https://doi.org/10.3389/fenvs.2023.1117184>



BioSci Publisher®

Zhu W., Rezaei E., Sun Z., Wang J., and Siebert S., 2024, Soil-climate interactions enhance understanding of long-term crop yield stability, *European Journal of Agronomy*, 160: 127386.

<https://doi.org/10.1016/j.eja.2024.127386>



BioSci Publisher®

Disclaimer/Publisher's Note

The statements, opinions, and data contained in all publications are solely those of the individual authors and contributors and do not represent the views of the publishing house and/or its editors. The publisher and/or its editors disclaim all responsibility for any harm or damage to persons or property that may result from the application of ideas, methods, instructions, or products discussed in the content. Publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
