Hybrid Explainable Artificial Intelligence and Evolutionary Optimization for Climate-Resilient Wheat and Maize Yield Prediction: A Systematic Review

Dr. Manpreet Kaur[1]*, Prof. (Dr.) Subhendu Kumar Pani[2]

[1]Post Doctoral Researcher, Lincoln University College, Malaysia
[2]Krupajal Engineering College, Biju Patnaik university of Technology, Odisha, India
*Email ID: pdf.manpreetkaur@lincoln.edu.my

ABSTRACT

Background: The prediction of crop yield, accurate and interpretable, is becoming more and more relevant to the quest of making agriculture climate resilient. The yield of wheat (Triticum aestivum L.) and maize (Zea mays L.) is influenced by the interaction of nonlinearity between climate change, soil moisture, crop growth, remote sensing signals and management. Machine-learning and deep-learning models can capture such complexity, but they are not very interpretable, and their extrapolation is unstable, and have been plagued by model-selection problems, limiting their use under changing climates.

Objective: To assess, in a systematic way, the wheat and maize yield prediction models that include explainable artificial intelligence (XAI) and/or evolutionary optimization, focusing on their predictive accuracy and transparency, optimization performance, robustness and generalizability under climatic and environmental variability.

Methods: The review was conducted according to the guidelines of the PRISMA 2020. A literature search was conducted in PubMed/MEDLINE, Scopus, Web of Science Core Collection, ScienceDirect, Emerald Insight and Google Scholar (including additional searching), up to 18 August 2026. A total of 155 empirical English-language studies were included, published between January 2016 and August 2026, that assessed wheat and/or maize yield as a result of applying machine learning/deep learning, with XAI, interpretable modelling, evolutionary optimization or a directly relevant hybrid approach. A narrative synthesis was conducted since study designs and metrics were diverse. Domains adopted from PROBAST+AI were used to assess the quality and applicability.

Results. A total of 2338 records were found. Following removal of 798 duplicates, 1540 titles/abstracts were screened with 189 full text reports assessed and 14 studies were included in the qualitative synthesis. The most popular XAI technique was SHapley Additive exPlanations (SHAP), and the most popular evolutionary optimizers were those based on genetic algorithms and NSGA-III. While XAI was consistently improving transparency, evolutionary optimization was also sometimes better than a good conventional tuning in feature selection or model configuration, but not always better than the conventional tuning. More evidence existed for external climate generalizability than there was for predictive accuracy.

Conclusions. The use of hybrid XAI and evolutionary optimization is a promising approach for wheat and maize yield prediction, but an R^2 or a low prediction error level is not an indicator of climate resilience. Stronger evidence necessitates temporally and geographically independent validation and evaluation under climatic extremes, uncertainty reporting, comparison with well-tuned non-evolutionary baselines, and assessment of the stability of XAI explanations in the face of environmental change, in an agronomically plausible sense.

KEYWORDS: Crop Yield Prediction, Wheat, Maize, Explainable Artificial Intelligence, Machine Learning, SHAP, Genetic Algorithm, Evolutionary Optimization, Climate Resilience, Remote Sensing.

INTRODUCTION

The need for accurate and timely crop-yield prediction is growing in importance for food-security planning, farm management, agricultural insurance, breeding, irrigation and fertilizer scheduling, market forecasting, grain storage and reactions to climatic shocks. The productivity of wheat and maize is very sensitive to weather and climate, and is a major cereal crop. The evidence on the impact of warming on wheat and maize yields is found throughout all regions of the world, suggesting that without adequate adaptation, warming may decrease average productivity of these crops (Zhao et al., 2017; Ortiz-Bobea et al., 2021).

Yield formation is a multivariate, time dependent process. The interactions between temperature, precipitation, solar radiation, soil moisture, soil physicochemical properties, the availability of nutrients, genotype, phenology, planting date, irrigation, nutrient application, and other management decisions are non-linear. The significance of these variables varies with growth stage, site, year, cultivar and production system. As a result, a correlation gained in one climatic zone or region might not hold in another zone or region.

The use of remote sensing data has increased the amount of information available to model yields. The use of time series satellite data in combination with district-level yield statistics has proven to be useful for historical wheat yield mapping in Uttar Pradesh, India (Baghel & Sharma, 2022). Time-series MODIS imagery and meteorological and soil data have also been used to enhance regional maize yield estimation and provided interpretation of the contribution of the predictors (Lyu et al., 2024). More recently, there have been opportunities to apply reproducible resources to subnational crop yield forecasting across countries and data modalities, e.g. CY-Bench (Kallenberg et al., 2026).

Although traditional statistical models are still useful due to their relative simplicity, process-based crop models offer biologically based representations of crop growth and environmental responses. Yet, statistical methods can be limited for the cases of high-dimensional nonlinear relationships, while process-based models can be limited for the need for significant calibration and reliable physiological and management data. Machine-learning (ML) and deep-learning (DL) models have thus gained more and more significance for crop yield forecasting, such as random forests (RF), support vector regression, gradient boosting, XGBoost, LightGBM, artificial neural networks, convolutional neural networks (CNNs), and recurrent architectures (van Klompenburg et al., 2020).

A good predictive performance on a held-out set does not necessarily mean that a model has learned transferable and agronomically meaningful climate-yield relationships. Lischeid et al. (2022) stated that ML as a tool for crop-yield modelling is a valuable tool, but cannot be a replacement for scientific understanding. Hu et al. (2023) also demonstrated the capacity of black-box models to capture the historical observations but to learn an erroneous climate-yield correlation and react erratically to new climate conditions. Therefore, the predictive accuracy, interpretability, and climate robustness of the model should be regarded as three separate facets of the quality of the model.

Explainable artificial intelligence

Explainable AI (XAI) aims to make AI prediction systems more comprehensible by uncovering the factors and insights affecting the model's results. Positive and negative contributions of predictors are quantified with SHapley Additive exPlanations (SHAP) (Friedman et al., 2018) while Local Interpretable Model-agnostic Explanations (LIME) (Ribeiro et al., 2016) approximate a complex model locally by using a simpler interpretable representation. Other methods involve (1) permutation importance, (2) partial-dependence analysis, (3) individual conditional expectation, (4) accumulated local effects, (5) counterfactual explanations, and (6) intrinsically interpretable models.

In agriculture, XAI can provide insights into whether the temperature, precipitation, soil moisture, radiation, vegetation indices, crop height, soil properties, nutrients or management variables are most important for prediction, and whether their influence varies over different parts of the crop development cycle or region (Ryo, 2022). In Indian context, recent research has explored how the SHAP method and other explanation techniques have been leveraged to make predictions about crop yield (Dubey et al., 2025; Yenkikar et al., 2025), and a broader review has reported the burgeoning of XAI in the field of sustainable agriculture (Rajbongshi et al., 2026).

However, caution is still required for interpretation. A post-hoc explanation is dependent upon the model and the data that the model has been learning from. Not causal or distribution shift guarantee. This is especially critical for high stakes decisions and future climate extrapolation (Rudin, 2019).

Evolutionary optimization and hybrid modelling

The concept of evolutionary optimization and hybrid modelling were introduced in 1.2. Another approach to developing a model is evolutionary and population-based optimization. Feature selection, hyperparameter optimization, neural network configuration, construction of ensemble, or optimization of management using genetic algorithms (GA), particle swarm optimization (PSO), differential evolution, grey wolf optimization, and multi-objective algorithm such as NSGA-II and NSGA-III. Optimisation in agriculture has been a longstanding challenge of using heuristic approaches (Memmah et al., 2015), and today, optimization is widely used in agricultural engineering using bio- and nature-inspired algorithms (Maraveas et al., 2023).

Addressing different weaknesses with XAI and evolutionary optimization. Optimization is an approach that searches for an efficient model configuration or a subset of features that will yield the best prediction, while XAI explains the prediction that the model produces. The models they can create by integrating them might therefore be accurate, parsimonious and interpretable. For wheat-yield modelling, for instance, Jahan et al. (2026) did a combination of mutual information, recursive feature elimination, NSGA-III, stacked LightGBM/deep-neural-network prediction, and SHAP.

Hybridization can also be used to link data-driven models and process-based knowledge. In this context, Luo et al. (2024) tested a hybrid approach combining ML and a dynamical model to incorporate remote-sensing information for in-season forecasting of maize under drought conditions. To enhance the predictions of winter-wheat yield in a semi-arid environment, Oulaid et al. (2024) employed the use of satellite imagery, hemispherical photography and PSO-assisted parameterization. The following are examples of methods that have been developed to move past the single algorithm to a more integrated system for prediction.

Rationale and review gap

Previous review work has compiled crop yield forecasting using machine learning (van Klompenburg et al., 2020), interpretable ML in agriculture (Ryo, 2022), nature-inspired algorithms for agriculture (Maraveas et al., 2023) and XAI applications in sustainable agriculture (Rajbongshi et al., 2026). But these strands tend to be looked at individually. There is a need for a focused synthesis at the nexus of wheat and maize yield prediction, XAI, evolutionary optimization and climate resilience.

The main question of the method is not only which algorithm has the least error. A climate-resilient prediction system should also have clear reasoning, a suitable optimisation gain, temporal and spatial transferability, resilience with drought/climatic extremes and plausible behaviour in an environment outside the distribution of the training data. This review hence combines predictive, explanatory, optimization and generalization result as opposed to considering them as separate research problems.

Review question and objectives

Review question: How do AI-based wheat and maize yield-prediction models using explainable AI and/or evolutionary optimization algorithms, as compared to conventional and non-hybrid approaches, perform in terms of predictive accuracy, transparency, optimization efficiency, and generalizability within the range of climate variability and climate change between 2016 and 2026 across different cropping systems?

Objectives:

1. To assess the prediction accuracy of wheat and maize yield prediction using hybrid AI models under different climatic and environmental conditions and their explainability.
2. To evaluate the impact of evolutionary optimization on being able to select features, accuracy of the model, computational efficiency, robustness, and generalizability.
3. To understand methodological gaps and evidence limitations that impact wheat and maize hybrid AI predictive models.

METHODS

Review design and reporting framework

The systematic review was performed and reported following the guidance given by PRISMA 2020 (Page et al., 2021). The review was aimed at empirical evidence on XAI, interpretable modelling, evolutionary optimization and pertinent hybrid approaches for AI in the prediction of yield of wheat and maize in the context of climate vulnerability.

Eligibility criteria

Studies were accepted when they: (a) examined wheat and/or maize under climatic, environmental, or soil variability, or management variability; (b) used an ML/DL model with XAI, interpretable modelling, evolutionary optimization, or related hybrid approaches; (c) reported predictive accuracy, explainability, optimization, robustness, or generalizability; (d) presented empirical results; and (e) were published in English between 2016 and 2026 (August). Comparators may comprise conventional ML/DL, non-optimized models, non-explainable models, statistical

models, process-based crop models or any other baseline models. The primary evidence synthesis did not include reviews, editorials or purely theoretical papers, non-yield applications, or studies not related to wheat or maize. When not relevant, earlier or review level publications were retained only as background literature.

Information sources

The final search included PubMed/MEDLINE, Scopus, Web of Science Core Collection, ScienceDirect, Emerald Insight and Google Scholar up to 18 August 2026. The number of potentially relevant results from the first 500 as entered in Google Scholar was screened. Supplementary searching consisted of reference-list checking and other appropriate searching methods. These information sources are the same as those reported in the Results as the final PRISMA database counts.

Search strategy

All searches took place with a combination of crop, yield, AI, explainability, optimization and climate/environment terms, using boolean operators. The syntax of the database was modified to support each platform, while preserving the conceptual structure. A typical query string was:

> ("wheat" OR "maize" OR "corn") AND ("yield prediction" OR "yield forecasting" OR "yield estimation") AND ("machine learning" OR "deep learning" OR "artificial intelligence") AND ("SHAP" OR "LIME" OR "explainable AI" OR "interpretable machine learning" OR "genetic algorithm" OR "particle swarm" OR "differential evolution" OR "grey wolf optimizer" OR "NSGA-II" OR "NSGA-III") AND ("climate" OR "weather" OR "drought" OR "temperature" OR "remote sensing" OR "environmental")

Study selection

Duplicates were eliminated and titles/abstracts screened by two reviewers independently, followed by full text assessment of eligibility based on predefined criteria. Conflict was settled by consensus or by consulting with a third reader, as the case may be. Any reasons for a full text exclusion were noted and summarised in the PRISMA flow diagram.

Data extraction

The following information was captured using a standard extraction framework: publication year, crop, country/context, spatial scale, size of the dataset, length of the time series, climate and environmental predictors, remote-sensing inputs, ML/DL architecture, XAI method, optimizer, comparator, validation design, R^2, RMSE, MAE, robustness/generalization evidence, and key findings. Specific focus was on whether performance was assessed in a temporal holdout, geographic transfer, drought, climatic extremes, or other type of environmental distribution shift.

Quality and risk-of-bias appraisal

Since the available evidence was a collection of studies on prediction models, the domains of quality, risk of bias and applicability were assessed according to the agriculture adapted domains developed in PROBAST+AI (Moons et al., 2025). Judgements were grouped into Low, Unclear or High. For agricultural prediction, the importance of data representativeness, predictor availability, validation data reliability, overfitting control, independence of the validation data, and external temporal/spatial evaluation was emphasized.

Data synthesis

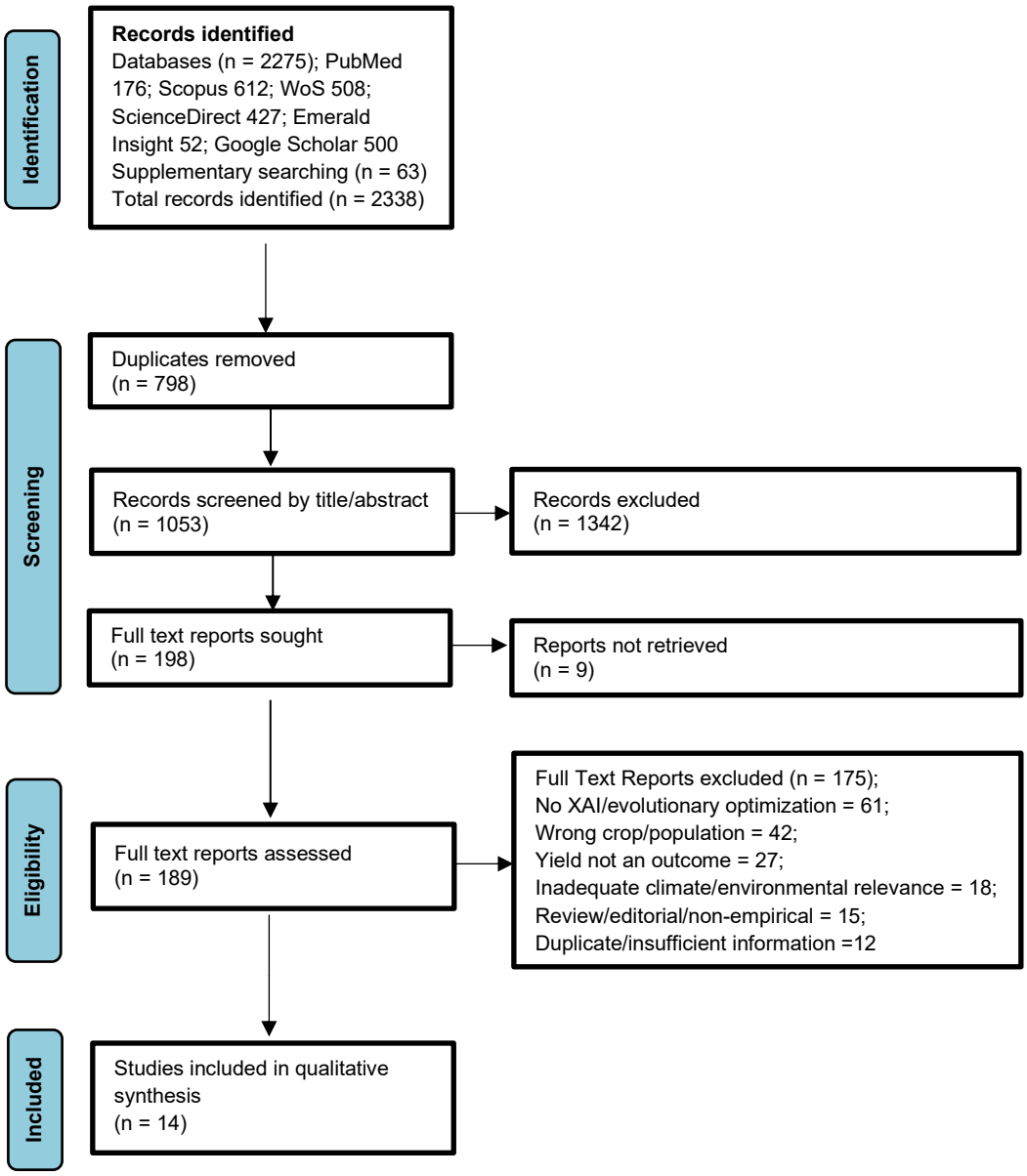
This was expected to be a great deal of variability in regards to the crop, the scale of the study, the data sources, the predictor sets, the model architecture, the optimization method, the validation design, and the performance metrics. The major analysis method used was narrative synthesis. Studies were classified according to crop, XAI technique, optimizer, data source, the context to which the study was related to climate stress, and the generalization design. Because of the differences in outcome definitions and validation designs, a pooled meta-analysis was not performed.

RESULTS

Study selection

A total of 2,338 records were found: 2,275 in the six databases and 63 by supplementary searching. A total of 176 records from PubMed, 612 records from Scopus, 508 records from Web of Science, 427 records from ScienceDirect, 52 records from Emerald Insight, and the first 500 potentially relevant Google Scholar records were included in the database. One thousand, five hundred and forty (1,540) titles and abstracts were screened, yielding 1,342 records that were excluded after the removal of 798 duplicates. All 198 reports were requested for retrieval; of which, nine were not retrieved, and 189 reports were available for eligibility assessment. Another 175 reports were also removed, including no XAI/evolutionary optimization (n = 61), wrong crop/population (n = 42), yield not an outcome (n = 27), inadequate climate/environmental relevance (n = 18), review/editorial/non-empirical publication (n = 15) and duplicate/insufficient information (n = 12). Fourteen studies were included in the qualitative synthesis.

Figure 1. PRISMA 2020 flow diagram for study selection



Characteristics of included studies

All 14 of the studies included included wheat and maize production systems in Asia, Europe, and North America, and data included observations at the field and UAV scale, as well as multi-year county and multi-state data. The interpretability approaches were clearly dominated by SHAP and model-agnostic importance methods, while GA and NSGA-II were clearly the best evolutionary optimizers. The evidence that is included is summarized in Table 1.

Table 1. Characteristics and principal findings of the 14 included studies

Study	Country / crop	Design / data extent	Hybrid AI / XAI / optimization	Comparator	Period	Principal finding

Study	Country / crop	Design / data extent	Hybrid AI / XAI / optimization	Comparator	Period	Principal finding
Shahhosseini et al. (2021)	USA; maize	County-level retrospective modelling; 10,016 observations from 293 counties	APSIM + LightGBM / RF / XGBoost; feature - importance analysis	Stand-alone ML and APSIM	1984 - 2018	APSIM-derived variables reduced RMSE by about 7%-20%; soil-water and drought-stress information was influential.
Srivastava et al. (2022)	Germany; winter wheat	Regional longitudinal modelling; 271 counties	1D-CNN + SHAP / force plots	Eight ML/DL baselines	1999 - 2019	CNN reduced RMSE by about 7%-14% and MAE by 3%-15% versus the strongest baseline; SHAP identified temporally specific weather and soil drivers.
Nayak et al. (2022)	Northwest India; wheat	On-farm observational modelling; 6,181 field observations across two seasons	Fine-tuned ML + permutation importance and interaction analysis	Ridge, LASSO, CART, KNN, SVM, boosting, RF	Two growing seasons	RF performed best (reported R^2 approximately 0.44-0.63); interpretable analysis identified management and environmental drivers and interactions.

Study	Country / crop	Design / data extent	Hybrid AI / XAI / optimization	Comparator	Period	Principal finding
Hu et al. (2023)	USA; Ohio maize + synthetic experiments	Long-term climate-yield modelling and synthetic experiments	Interpretable Bayesian ensemble model	ElasticNet, NN, MARS, SVM, RF, XGBoost	Long-term historical series	Interpretable Bayesian modelling better recovered underlying climate-yield relationships; RF and XGBoost were less reliable under novel climate extrapolation.
Bai et al. (2024)	China; North China Plain wheat	Crop-model/ML climate-extreme assessment	APSIM + RF / LGBM with GA feature selection	APSIM alone; non-GA feature selection	Historical and future climate scenarios	APSIM+RF-GA explained about 93% of observed yield variability and improved accuracy by about 33% over APSIM alone; heat and frost extremes were important.
Lyu et al. (2024)	China; maize	Regional multi-year time-series MODIS, meteorological and soil modelling	Multi-source ML + interpretability analysis	Alternative ML/input combinations	2010 - 2020	The comprehensive model reached $R^2 = 0.76$; integrating multi-source data improved estimation and revealed substantial soil and environmental contributions.

Study	Country / crop	Design / data extent	Hybrid AI / XAI / optimization	Comparator	Period	Principal finding
Azrai et al. (2024)	Indonesia; maize under drought	Multi-environment field experiment; 35 genotypes and 525 plot observations	GA-optimized RF, SVM and KNN	Grid-search ensemble and conventional models	Normal and drought environments	GA-RF achieved R^2 = 0.91 for grain yield, while the best ensemble reached R^2 = 0.92, showing strong but not universal superiority of evolutionary optimization.
Kumar et al. (2025)	USA; field-scale corn	UAV multispectral modelling	Ensemble ML + SHAP, PDP, ICE, permutation importance	GLM, KNN, PCR, RF, SVM, BRNN	Field-season observations	RF-based ensemble reached R^2 = 0.90 versus R^2 = 0.67 for the strongest individual model; XAI clarified predictor contributions and residual behaviour.

Study	Country / crop	Design / data extent	Hybrid AI / XAI / optimization	Comparator	Period	Principal finding
Khodjaev et al. (2025)	Germany; wheat	UAV field study; three plots covering about 30 ha	Optimized RF / XGBoost and other ML + SHAP	Default ML and linear regression	2020 season	Optimization improved performance relative to untuned/linear alternatives; SHAP highlighted red-edge vegetation information and crop height, but the one-season design limited external inference.
Rufaioğlu et al. (2025)	Türkiye; durum wheat	Growth-stage sensor modelling under semi-arid conditions	RF / boosting / LightGBM / XGBoost + SHAP	Alternative ML models and Zadoks stages	Four Zadoks stages	ZD31 was the strongest prediction stage; XGBoost reached about $R^2 = 0.81$, with NDVI and INSEY among the most influential predictors.

Study	Country / crop	Design / data extent	Hybrid AI / XAI / optimization	Comparator	Period	Principal finding
Lokeshwari et al. (2026)	India; irrigated / rainfed wheat	Field-scale germplasm study; 3,350 entries at two locations	GA-optimized DNN + feature-importance analysis	RF, LASSO, SVR	2020 - 2021 season	GA-DNN outperformed conventional ML; proximal-sensing variables including NDVI showed substantial predictive value across irrigated and rainfed conditions.
Jahan et al. (2026)	Iran; irrigated wheat	Regional longitudinal modelling; 17 counties and 47 variables	MI + RFE + NSGA-III + stacked LightGBM / DNN + SHAP	Alternative feature subsets/models	2004 - 2023	The optimized 10-feature model achieved independent-test $R^2 = 0.44$; SHAP showed regional and climate-soil-nutrition interactions.
Smith et al. (2026)	USA; rainfed maize	Multi-state commercial yield-monitor modelling; 134 crop-site-years across nine states	ML + SHAP	Alternative ML configurations	Multi-year	Models achieved strong predictive skill; SHAP identified weather, soil and terrain constraints across commercial production environments.

Study	Country / crop	Design / data extent	Hybrid AI / XAI / optimization	Comparator	Period	Principal finding
Lu et al. (2026)	Eastern China; winter wheat	County-level spatiotemporal modelling	TabPFN + SHAP with SIF, vegetation and climate data	Conventional ML models	Multi-year	Combined inputs achieved R^2 up to 0.81 and retained useful performance for unseen years and geographic areas; SHAP revealed stress-dependent changes in predictor importance.

Quality and risk-of-bias assessment

There were 4 studies with a relatively low overall Risk of Bias rating, 9 studies with an Unclear overall Risk of Bias rating and 1 study with a High overall Risk of Bias rating. Most common concerns were limited external validation, especially in cases of a single season, experimental site, or random split of the dataset for evaluation. More robust designs included withheld years, test sets independent across geographic regions, drought years or anomalous years explicitly.

Synthesis of findings

Explainable AI and model transparency

They have all improved the transparency of otherwise complex yield-prediction system, specifically for SHAP and related post-hoc interpretation methods. Srivastava et al. (2022) demonstrated that SHAP not only revealed influential variables but also when during wheat development those variables became influential. In corn, Kumar et al. (2025) used SHAP, permutation importance, PDP, and ICE to complement each other in interpreting corn-yield predictions, and Rufaioğlu et al. (2025) showed how to interpret PDP and ICE at different stages of durum wheat development.

Important predictors changed in terms of scale and production context. Field and UAV scale models drew attention to spectral indices and crop height while regional and commercial field models focused on the weather, soil, terrain and wider environmental controls. Multi-source satellite, meteorological and soil information were found to aid in the estimation of maize and assist in the interpretation of the respective contribution of the different information sources by Lyu et al. (2024).

But reliability in climate extrapolation was not achieved by explainability. Hu et al. (2023) showed that models with an acceptable historical fit may be able to learn climate-yield relationships that are misleading. Therefore, external climatic robustness and transparency and predictive accuracy must be understood as distinct.

Evolutionary optimization

The superiority of evolutionary optimization over conventional optimization was not consistent, and in general it improved the configuration and/or the feature selection. Bai et al. (2024) reported strong performance of APSIM+RF-GA framework for climate-extreme evaluation and Lokeshwari et al. (2026) concluded that GA-optimized DNN performed better than traditional ML methods for wheat prediction for irrigated and rainfed environments.

The grid search ensemble (R^2 = 0.92) provided an important counter example to Azrai et al. (2024) for maize grain yield with GA-RF (R^2 = 0.91). This is not a drawback of the evolutionary search but rather a reminder that the optimizer value should be determined by experiments against a strong conventional competitor.

Combined optimization and explainability

The conceptually most similar kind of review to the topic of the review was the combination of optimization and explanation in one modelling pipeline. Jahan et al. (2026) integrated mutual information, recursive feature elimination, NSGA-III, stacked LightGBM/DNN prediction and SHAP. This complex architecture still did not translate into a high out-of-sample test result, with the independent-test R^2 remaining at 0.44, which proves that complex architecture does not necessarily result in high out-of-sample performance for environments that are heterogeneous.

Additionally, Optimization + SHAP was found to yield high within-study performance in the wheat modelling work done by Khodjaev et al. (2025), but only allowed inferences on short-term climate robustness due to the one-season design of the study. These studies collectively support the evaluation of the benefits of optimization, explanation quality, and external generalization separately.

Climate resilience and generalizability

In comparison, evidence of climate resilience was less comprehensive than evidence of predictive accuracy. Shahhosseini et al. (2021) added variables that represent the process of soil water and drought stress, and simulated an extreme dry year. The work of Bai et al. (2024) explicitly included heat and frost extremes and Azrai et al. (2024) assessed maize under normal and drought stressed conditions.

The most robust generalizations were the studies which directly tested models against years that have not been used in the training, with geographical regions or with unusual environmental circumstances. Lu et al. (2026) were able to report good performance under non-similar conditions in space and time, while Hu et al. (2023) were able to showcase the opposite problem: typical black-box algorithms become unreliable when extrapolated outside the range of historical climate situations. Therefore, the split between a random train/test set is not sufficient to conclude climate resilience.

Overall synthesis

Overall, hybrid methods boosted at least one of the four criteria in the 14 studies included in this work (predictive accuracy, interpretability, feature selection, optimization efficiency and robustness), but there was no single at algorithmic architecture that worked well across the various crops and environments. XAI was most well-supported for identifying model drivers, and optimization for tuning or feature reduction. The most variable response was for external generalizability under climate change. The evidence thus shows that a high R^2 value or a low RMSE is not always a good indication of climate resilience.

DISCUSSION

Principal findings

The aim of this systematic review was to assess whether AI-based wheat and maize yield-prediction models with XAI and/or evolutionary optimization offer benefits in terms of predictive accuracy, transparency, optimization performance, robustness and generalizability in the face of climate variability and climate change. The effectiveness of hybrid approaches varied widely across the 14 studies reviewed, although generally this strategy enhanced at least one of these dimensions.

Explainable AI exhibited the greatest score to increase transparency and identify influential yield drivers. In many cases, evolutionary optimization of features or model parameters resulted in an improvement, but not always a superior improvement over traditional tuning. Most importantly, evidence of climate robustness or external generalizability from strong predictive performance within the dataset was not consistent.

Interpretation of XAI evidence

The studies reviewed here have shown the ability of explanation tools to unlock ‘black boxes' of models and to reveal crop-, stage-, and scale-specific drivers. This is agronomically useful because the questions the county-level climate model answers are different than the questions the UAV field experiment answers, and the questions the multi-state commercial-field model answers are different. Expecting the same feature-importance rankings depends on answering the same question.

However, it should not be a substitute for causal inference, at the same time. For both agriculture and data analysis, Ryo (2022) stressed on the importance of interpretability, whereas Rajbongshi et al. (2026) highlighted the growing presence of XAI in sustainable agriculture. However, Hu et al. (2023) and others, like Rudin (2019), have shown that an explanation can be correct and capture a model whose learnt association is shaky and/or misguided. Hence, the stability of explanations should be tested in future studies over years, sites, algorithms and climate regimes.

Interpretation of optimization evidence

The evolutionary optimization proved to be very useful when the spaces of the predictors were large, correlated or multi-objective. The results of GA-based workflows in Bai et al. (2024) and Lokeshwari et al. (2026) demonstrated significant improvements over their respective baselines, and NSGA-III in Jahan et al. (2026) facilitated multi-objective feature reduction. The results are consistent with the overall literature on nature-inspired optimization in agriculture-engineering (Maraveas et al., 2023).

It is important to note that the comparison in Azrai et al. (2024) indicates the disadvantages of relying on evolutionary techniques automatically - a well-tuned ensemble is marginally better than GA-RF. Future comparisons need to be with the same data partitions, the same nested tuning procedures and to include not only accuracy but feature count, computational cost, convergence behaviour and external validation performance. Data from remote sensing, process modeling and multi-source data are utilized.

Remote sensing, process models, and multi-source data

A new bridge between observation of crop state by remote sensing and crop state prediction by data was found. Historical wheat-yield mapping was carried out by Baghel and Sharma (2022) in Uttar Pradesh and maize mapping was done by Lyu et al. (2024) through integrating MODIS imagery with meteorological and soil parameters. Not only can satellite imagery be used as a direct ML predictor, but according to Oulaid et al. (2024) it can also be used to support the process-model parameterization.

However, hybrid crop-model-ML methods could be particularly beneficial in the context of climate stress, as they can incorporate variables that are not explicitly observed but are relevant to soil-water balance, phenology, and physiological response. This was demonstrated by Luo et al. (2024) for in-season maize prediction under drought, and by Shahhosseini et al. (2021) for the use of crop-model output for maize-yield prediction in the U.S. Corn Belt.

Climate robustness: interpolation is not extrapolation

The most significant cross-study result is the difference between interpolation and extrapolation. Typically, a holdout is random, testing observations that are statistically similar to the training observations. A new type of climate-resilient decision support, however, could need predictions under combinations of environmental conditions that have not been experienced before. Hu et al. (2023) directly showed that a strong historical fit does not necessarily imply good recovery of underlying climate-yield relationships.

More convincing evidence of resilience was provided by studies in which years were withheld, geographic blocks were excluded, drought environments were excluded, or by extreme climate tests. This indicates that future research on agricultural AI should consider leave-one-year-out, leave-one-region-out, climate-extreme holdouts, and explicit out-of-distribution evaluation as key validation strategies as opposed to optional analyses.

Benchmarking and reproducibility

The difficulty in cross-study comparison is driven by the fact that there are a number of differences in data sources, spatial scales, crops, validation procedures and metrics. This issue can be mitigated with standardized datasets. CY-Bench is a recent subnational crop yield benchmark that integrates yield statistics, meteorological, remote-sensing and soil data, across multiple countries (Kallenberg et al., 2026).

Benchmarking should allow for more stringent comparisons of optimization algorithms, explanation stability, transfer between various countries and results in response to climate shifts. Comprehensive search/tuning spaces, random seeds, data partition, code, model uncertainty and detailed preprocessing procedures would be further reported for achieving reproducibility.

Implications for Indian agriculture

The evidence is relevant to an Indian context. Baghel and Sharma (2022) explored wheat yield mapping by satellite techniques in the state of Uttar Pradesh; Nayak et al. (2022) applied interpretable ML to analyze wheat yield variation in NW Indian farms; Dubey et al. (2025) studied wheat yield prediction using satellite-based explanation by SHAP; Lokeshwari et al. (2026) used GA-optimized deep learning for the prediction of wheat yield in both irrigated and rainfed condition; and Yenkikar et al. (2025) developed a hybrid XAI wheat yield framework. These studies suggest that local decision systems can be developed based on three components: satellite observations, meteorological and soil parameters, and field measurement data, with the addition of XAI and optimized ML, which have great potential for India. But, the agroecological diversity of India necessitates cross region and multi-year validation for operational deployment.

STRENGTHS AND LIMITATIONS

In this review one of the strengths is the integrated aspect of prediction, interpretability and optimization in the context of climate robustness instead of considering these outcomes separately. It also features evidence from proximal sensing, UAV data, remote sensing, crop-model output, farm records and regional statistics, enabling the review to differentiate between success at farm level and generalization of broader environmental conditions.

There are a number of caveats to interpretation. First, the studies included were very heterogeneous in terms of crop context, spatial scale, sets of predictors, algorithms, optimization strategies, validation design and reported metrics which prevented a defensible pooled meta-analysis. Second, just four studies were deemed at relatively low overall risk of bias, whereas nine were deemed Unclear and one High, mostly due to limited external validation. Third, a significantly lower number of studies explicitly tested novel climates and out-of-distribution conditions. Fourth, the outputs of post-hoc XAI do not prove cause and effect. Lastly, the field is growing quickly and results could change with the availability of more extensive multi-region datasets and improved climate-shift validation.

Future research

- Focus multi-year, multi-location, drought, heat and climate extreme validations.
- Whenever possible, employ leave-one-year-out, leave-one-region-out and out-of-distribution evaluation.
- Compare evolutionary optimizers to other strong baselines (Bayesian optimization, random search and grid search) on identical validation schemes.
- In addition to prediction error, also report optimization costs, feature reduction, uncertainty and computational requirements.
- Compare the stability of SHAP/LIME and other explanation outputs across years, locations, model and random seeds.
- Combine remote sensing, soil, management, climate, phenology and process-model data while avoiding leakage and collinearity.
- Utilize the existing standardized benchmark datasets and share code/data partitions to enhance reproducibility.

- Differentiate predictive explanation from causal inference, and leverage agronomic knowledge to interpret results of XAI.

CONCLUSION

The use of hybrid AI methodologies with explainable AI and evolutionary optimization can enhance wheat and maize prediction of yield, but the potential benefits vary depending on the outcome. XAI is the most uniform contributor of model transparency and identification of influential yield drivers. Feature selection or model configuration is often optimized using evolutionary techniques, which may yield better results than traditional tuning or feature combination techniques, but is not necessarily superior.

The robustness under climate variability and climate change is relatively scarce and relies heavily on the temporal, spatial, drought or climate extreme validation. The predictive accuracy alone is thus not enough to determine that a yield-prediction model is climate resilient. The most promising future systems will require the ability to be accurate, have clear explanation, to be optimized appropriately, to perform uncertainty characterization, and to perform well on unseen environmental conditions.

REFERENCES

1. Azrai, M., Aqil, M., Andayani, N. N., Efendi, R., Suarni, Suwardi, Jihad, M., Zainuddin, B., Salim, Bahtiar, Muliadi, A., Yasin, M., Hannan, M. F. I., Rahman, & Syam, A. (2024). Optimizing ensembles machine learning, genetic algorithms, and multivariate modeling for enhanced prediction of maize yield and stress tolerance index. *Frontiers in Sustainable Food Systems*, *8*, 1334421. https://doi.org/10.3389/fsufs.2024.1334421
2. Baghel, R., & Sharma, P. (2022). Historical wheat yield mapping using time-series satellite data and district-wise yield statistics over Uttar Pradesh state, India. *Remote Sensing Applications: Society and Environment*, *27*, 100808. https://doi.org/10.1016/j.rsase.2022.100808
3. Bai, H., Xiao, D., Tang, J., & Liu, D. L. (2024). Evaluation of wheat yield in North China Plain under extreme climate by coupling crop model with machine learning. *Computers and Electronics in Agriculture*, *217*, 108651. https://doi.org/10.1016/j.compag.2024.108651
4. Dubey, Y., Sakhare, A., Tasare, A., Kakad, S., & Umate, R. (2025). Explainable model for agricultural crop yield prediction in Indian conditions with SHAP analysis. *International Journal of Electronics and Communication Engineering*, *12*(1), 236-244. https://doi.org/10.14445/23488549/IJECE-V12I1P118
5. Hu, T., Zhang, X., Bohrer, G., Liu, Y., Zhou, Y., Martin, J., Li, Y., & Zhao, K. (2023). Crop yield prediction via explainable AI and interpretable machine learning: Dangers of black box models for evaluating climate change impacts on crop yield. *Agricultural and Forest Meteorology*, *336*, 109458. https://doi.org/10.1016/j.agrformet.2023.109458
6. Jahan, M., Bannayan, M., Nassiri-Mahallati, M., & Yaghoubi, F. (2026). A novel hybrid NSGA-III and machine learning framework for modeling wheat yield variability using climatic, edaphic, and nutritional drivers. *Scientific Reports*, *16*, 20855. https://doi.org/10.1038/s41598-026-48918-0
7. Kallenberg, M., Paudel, D., Ofori-Ampofo, S., Baja, H., van Bree, R., Potze, A., Poudel, P., Saleh, A., Anderson, W., von Bloh, M., Castellano, A., Ennaji, O., Hamed, R., Laudien, R., Lee, D., Luna, I.,

Masiliūnas, D., Meroni, M., Mutuku, J. M., ... Athanasiadis, I. N. (2026). CY-Bench: A comprehensive benchmark dataset for sub-national crop yield forecasting. *Earth System Science Data*, *18*, 3997-4018. https://doi.org/10.5194/essd-18-3997-2026

8. Khodjaev, S., Bobojonov, I., Kuhn, L., & Glauben, T. (2025). Optimizing machine learning models for wheat yield estimation using a comprehensive UAV dataset. *Modeling Earth Systems and Environment*, *11*, 15. https://doi.org/10.1007/s40808-024-02188-9
9. Kumar, C., Dhillon, J., Huang, Y., & Reddy, K. N. (2025). Explainable machine learning models for corn yield prediction using UAV multispectral data. *Computers and Electronics in Agriculture*, *231*, 109990. https://doi.org/10.1016/j.compag.2025.109990
10. Lischeid, G., Webber, H., Sommer, M., Nendel, C., & Ewert, F. (2022). Machine learning in crop yield modelling: A powerful tool, but no surrogate for science. *Agricultural and Forest Meteorology*, *312*, 108698. https://doi.org/10.1016/j.agrformet.2021.108698
11. Lokeshwari, M., Jha, G. K., Praveenkumar, A., Kumari, J., Navathe, S., Singh, G. P., & Prasad, P. V. V. (2026). A novel deep learning framework for field-scale wheat yield prediction. *Theoretical and Applied Genetics*, *139*(3), 79. https://doi.org/10.1007/s00122-026-05166-0
12. Lu, B., Yu, K., Qiu, L., Li, H., Wang, H., Zhang, W., Liu, X., Shan, J., & Li, N. (2026). Predicting county-level winter wheat yield in Eastern China using multi-source spatiotemporal data: An explainable machine learning approach. *Agricultural and Forest Meteorology*, *386*, 111236. https://doi.org/10.1016/j.agrformet.2026.111236
13. Luo, Y., Wang, H., Cao, J., Li, J., Tian, Q., Leng, G., & Niyogi, D. (2024). Evaluation of machine learning-dynamical hybrid method incorporating remote sensing data for in-season maize yield prediction under drought. *Precision Agriculture*, *25*, 1982-2006. https://doi.org/10.1007/s11119-024-10149-6
14. Lyu, Y., Wang, P., Bai, X., Li, X., Ye, X., Hu, Y., & Zhang, J. (2024). Machine learning techniques and interpretability for maize yield estimation using time-series images of MODIS and multi-source data. *Computers and Electronics in Agriculture*, *222*, 109063. https://doi.org/10.1016/j.compag.2024.109063
15. Maraveas, C., Asteris, P. G., Arvanitis, K. G., Bartzanas, T., & Loukatos, D. (2023). Application of bio and nature-inspired algorithms in agricultural engineering. *Archives of Computational Methods in Engineering*, *30*, 1979-2012. https://doi.org/10.1007/s11831-022-09857-x
16. Memmah, M.-M., Lescourret, F., Yao, X., & Lavigne, C. (2015). Metaheuristics for agricultural land use optimization: A review. *Agronomy for Sustainable Development*, *35*, 975-998. https://doi.org/10.1007/s13593-015-0303-4
17. Moons, K. G. M., Damen, J. A. A., Kaul, T., Hooft, L., Andaur Navarro, C., Dhiman, P., Beam, A. L., Van Calster, B., Celi, L. A., Denaxas, S., Denniston, A. K., Ghassemi, M., Heinze, G., Kengne, A. P., Maier-Hein, L., Liu, X., Logullo, P., McCradden, M. D., Liu, N., ... van Smeden, M. (2025). PROBAST+AI: An updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ*, *388*, e082505. https://doi.org/10.1136/bmj-2024-082505
18. Nayak, H. S., Silva, J. V., Parihar, C. M., Krupnik, T. J., Sena, D. R., Kakraliya, S. K., Jat, H. S., Sidhu, H. S., Sharma, P. C., Jat, M. L., & Sapkota, T. B. (2022). Interpretable machine learning methods to explain on-farm yield variability of high productivity wheat in Northwest India. *Field Crops Research*, *287*, 108640. https://doi.org/10.1016/j.fcr.2022.108640

19. Ortiz-Bobea, A., Ault, T. R., Carrillo, C. M., Chambers, R. G., & Lobell, D. B. (2021). Anthropogenic climate change has slowed global agricultural productivity growth. *Nature Climate Change*, *11*, 306-312. https://doi.org/10.1038/s41558-021-01000-1
20. Oulaid, B., Milne, A. E., Waine, T., El Alami, R., Rafiqi, M., & Corstanje, R. (2024). Stepwise model parametrisation using satellite imagery and hemispherical photography: Tuning AquaCrop sensitive parameters for improved winter wheat yield predictions in semi-arid regions. *Field Crops Research*, *309*, 109327. https://doi.org/10.1016/j.fcr.2024.109327
21. Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ*, *372*, n160. https://doi.org/10.1136/bmj.n160
22. Rajbongshi, A., Johora, F. T., Hossain, A., Sarker, M. S., Rahman, M. H., Rahman, M. W., Alotaibi, F. T., & Moni, M. A. (2026). Leveraging explainable AI for sustainable agriculture: A comprehensive review of recent advances. *Artificial Intelligence Review*, *59*, 105. https://doi.org/10.1007/s10462-025-11459-5
23. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135-1144). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939778
24. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*, 206-215. https://doi.org/10.1038/s42256-019-0048-x
25. Rufaioğlu, S. B., Bilgili, A. V., Savaşlı, E., Özberk, İ., Aydemir, S., Ismael, A. M., Kaya, Y., & Matos-Carvalho, J. P. (2025). Sensor-based yield prediction in durum wheat under semi-arid conditions using machine learning across Zadoks growth stages. *Remote Sensing*, *17*(14), 2416. https://doi.org/10.3390/rs17142416
26. Ryo, M. (2022). Explainable artificial intelligence and interpretable machine learning for agricultural data analysis. *Artificial Intelligence in Agriculture*, *6*, 257-265. https://doi.org/10.1016/j.aiia.2022.11.003
27. Shahhosseini, M., Hu, G., Huber, I., & Archontoulis, S. V. (2021). Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. *Scientific Reports*, *11*, 1606. https://doi.org/10.1038/s41598-020-80820-1
28. Smith, H. W., Heffernan, C. J., Ashworth, A. J., Nalley, L. L., Bullock, D. S., Tullis, J., & Owens, P. R. (2026). Harvesting insights: Interpretable machine learning to understand environmental drivers of U.S. maize and soybean yield. *Scientific Reports*, *16*, 8994. https://doi.org/10.1038/s41598-026-38724-z
29. Srivastava, A. K., Safaei, N., Khaki, S., Lopez, G., Zeng, W., Ewert, F., Gaiser, T., & Rahimi, J. (2022). Winter wheat yield prediction using convolutional neural networks from environmental and phenological data. *Scientific Reports*, *12*, 3215. https://doi.org/10.1038/s41598-022-06249-w
30. van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, *177*, 105709. https://doi.org/10.1016/j.compag.2020.105709

31. Yenkikar, A., Mishra, V. P., Bali, M., & Ara, T. (2025). An explainable AI-based hybrid machine learning model for interpretability and enhanced crop yield prediction. *MethodsX*, *15*, 103442. https://doi.org/10.1016/j.mex.2025.103442
32. Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., Huang, M., Yao, Y., Bassu, S., Ciais, P., Durand, J.-L., Elliott, J., Ewert, F., Janssens, I. A., Li, T., Lin, E., Liu, Q., Martre, P., Müller, C., ... Asseng, S. (2017). Temperature increase reduces global yields of major crops in four independent estimates. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(35), 9326-9331. https://doi.org/10.1073/pnas.1701762114