ACCENT JOURNAL OF ECONOMICS ECOLOGY & ENGINEERING
Peer Reviewed and Refereed Journal, ISSN NO. 2456-1037
Available Online: www.ajeee.co.in/index.php/AJEEE

AGRICULTURE DECISION SYSTEM ON NEW MACHINE LEARNING METHODS FOR YIELD PREDICTION

AMBUJ KUMAR MISRA

Supervisor: Dr. Manish Saraf

ABSTRACT

Global food security faces unprecedented challenges from climate change, population growth, and resource limitations, necessitating accurate and scalable crop yield prediction systems. This study presents a comprehensive comparative analysis of contemporary machine learning and deep learning methodologies for agricultural yield prediction, synthesizing findings from 23 peer-reviewed studies published between 2023-2025. Our analysis evaluates model performance across diverse crops, geographical contexts, and data modalities including meteorological parameters, satellite-derived vegetation indices, UAV-based multispectral imagery, and soil properties. Results demonstrate that ensemble methods and hybrid architectures consistently outperform single-algorithm approaches, with Random Forest achieving R^2 values of 0.875 for Irish potatoes and 0.817 for maize, Support Vector Regression attaining R^2=0.95 for wheat yield prediction using UAV multispectral data, and hybrid CatBoost models demonstrating superior performance in capturing complex soil-climate interactions. Deep learning architectures, particularly CNN-SVM hybrids, achieved 97.54% accuracy in tomato grading applications. The findings reveal that no universal model excels across all crops and contexts; rather, model selection must be optimized for specific crop characteristics, data availability, and geographical conditions. We propose a multi-layered agriculture decision system architecture integrating real-time environmental data, remote sensing inputs, and automated machine learning pipelines for scalable yield prediction. Key challenges identified include computational demands, limited model interpretability, and applicability constraints in data-scarce regions. This research contributes empirical benchmarks for model selection and provides a framework for developing context-aware agricultural decision support systems.

Keywords: Crop yield prediction; machine learning; deep learning; ensemble methods; remote sensing; decision support system; precision agriculture.

1. INTRODUCTION

Agriculture serves as the economic backbone for approximately 60% of India's population, yet farmers face increasing vulnerabilities from climate change, natural disasters, and yield uncertainty. The intersection of rapid population expansion, shifting climatic patterns, and resource constraints has created an urgent requirement for accurate, timely crop yield forecasting systems that enable informed financial and managerial decision-making at early growth stages.

Traditional crop yield prediction methodologies, including process-based crop simulation models and statistical regression techniques, have demonstrated fundamental limitations in capturing the complex, non-linear interactions between meteorological variables, soil properties, management practices, and crop genetics. These conventional approaches require extensive parameterization, detailed field observations, and intricate calibration procedures while exhibiting substantial prediction errors under variable growing conditions. Systematic reviews indicate that historical empirical and mechanistic models achieve R^2 values ranging from 0.60-0.75, significantly lower than contemporary machine learning approaches.

Recent advances in machine learning (ML) and deep learning (DL) have transformed agricultural forecasting capabilities through their ability to process high-dimensional, multi-source data and identify latent patterns invisible to traditional methods. Satellite remote sensing, unmanned aerial vehicle (UAV) platforms, Internet of Things sensors, and automated weather stations now generate unprecedented volumes of agricultural data at multiple spatial and temporal scales. Concurrently, algorithmic innovations in ensemble learning, neural network architectures, and automated machine learning have created new opportunities for developing robust, scalable prediction systems.

Despite these technological advances, significant research gaps persist. First, the comparative performance of various ML/DL architectures across different crops, geographical regions, and data modalities remains inadequately characterized. Second, while individual studies demonstrate high accuracy in controlled experimental settings, the generalizability and operational deployment of these models in real-world agricultural decision systems require systematic investigation. Third, the integration of heterogeneous data sources—meteorological time series, soil spectral properties, vegetation indices, and management records—into unified prediction frameworks presents ongoing methodological challenges .

This study addresses these gaps through three primary contributions: (1) a comprehensive comparative analysis of contemporary machine learning methods for crop yield prediction, synthesizing empirical performance metrics across multiple crops, geographical contexts, and data modalities; (2) identification of optimal model architectures for specific agricultural applications and crop types; and (3) a proposed architecture for an integrated agriculture decision system that combines real-time environmental monitoring, automated machine learning pipelines, and decision support interfaces for end-users.

The remainder of this paper is organized as follows: Section 2 presents the systematic methodology for literature synthesis and performance benchmarking. Section 3 provides detailed comparative analysis of machine learning models across crop types and data sources. Section 4 proposes an integrated agriculture decision system architecture. Section 5 discusses key findings, challenges, and future research directions. Section 6 presents conclusions and practical implications.

2. METHODOLOGY

2.1 Research Design and Search Strategy

This study employed a systematic literature review methodology following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to identify, screen, and synthesize peer-reviewed research on machine learning applications for crop yield prediction. The search strategy targeted publications from 2023-2025 to capture the most recent methodological advances while ensuring sufficient temporal coverage for performance trend analysis.

Four academic databases were systematically queried: IEEE Xplore, SpringerLink, ScienceDirect, and Scopus. The search string combined terms related to machine learning ("machine learning," "deep learning," "random forest," "neural network," "ensemble," "gradient boosting"), agricultural outcomes ("crop yield," "yield prediction," "yield forecasting," "agricultural productivity"), and data sources ("remote sensing," "UAV," "satellite," "meteorological," "soil"). Forward and backward citation tracking was performed on included studies to identify additional relevant publications.

2.2 Inclusion Criteria and Study Selection

Studies were included if they met the following criteria: (1) peer-reviewed original research published in English between January 2023 and December 2025; (2) application of machine learning or deep learning algorithms for crop yield prediction; (3) reporting of quantitative performance metrics including at least one of R^2, RMSE, MAE, or accuracy; (4) clear description of data sources, model architecture, and validation methodology; and (5) focus on food crops relevant to global food security. Exclusion criteria comprised review articles without original empirical contributions, studies focused exclusively on crop classification rather than yield quantification, and research lacking sufficient methodological detail for performance benchmarking.

Initial database searching yielded 847 records. Following duplicate removal (n=163), title and abstract screening (n=684), and full-text eligibility assessment (n=97), 23 studies met all inclusion criteria and were included in the final synthesis. This sample size is consistent with previous systematic reviews in this domain and reflects the stringent methodological quality criteria applied .

2.3 Data Extraction and Synthesis Framework

A standardized data extraction protocol was implemented to capture: (1) bibliographic information; (2) crop type and geographical context; (3) machine learning algorithms employed; (4) data sources and feature types; (5) sample size and validation strategy; (6) performance metrics (R^2, RMSE, MAE, accuracy, precision, recall, F1-score); (7) model comparison benchmarks; and (8) reported limitations and challenges.

Performance metrics were harmonized across studies for comparative analysis. For regression tasks, R^2 (coefficient of determination) and RMSE (Root Mean Square Error) were prioritized as they were most consistently reported. For classification tasks, accuracy and F1-score were extracted. When multiple validation approaches were reported, hold-out test set performance was prioritized over cross-validation metrics to enable realistic assessment of generalization capability.

2.4 Analytical Approach

Comparative analysis was structured across three dimensions: (1) algorithm-level performance comparison examining relative effectiveness of different ML/DL architectures; (2) crop-specific model performance identifying optimal algorithms for individual crop types; and (3) data modality analysis evaluating prediction accuracy as a function of input data characteristics. Descriptive statistics and visualizations were generated using Python (Version 3.11) with Matplotlib and Seaborn libraries.

3. COMPARATIVE ANALYSIS OF MACHINE LEARNING MODELS

3.1 Overview of Model Architectures

Contemporary crop yield prediction employs a diverse array of machine learning architectures, each with distinct strengths and limitations depending on data characteristics and prediction objectives. Table 1 presents a comparative taxonomy of major model categories identified in the reviewed literature.

Table 1: Comparative Taxonomy of Machine Learning Architectures for Crop Yield Prediction

Model Category	Representative Algorithms	Strengths	Limitations	Optimal Application Context
Tree-Based Ensemble Methods	Random Forest, Gradient Boosting, XGBoost, LightGBM, CatBoost	High accuracy with structured data; feature importance interpretation; robust to outliers; handle non-linearity	Computationally intensive with large feature sets; risk of overfitting without tuning	Meteorological and soil tabular data; moderate sample sizes; require interpretability
Support Vector Machines	SVR, SVC	Effective in high-dimensional	Poor performance with large	Small to medium datasets; complex decision

		spaces; memory efficient; versatile kernel functions	datasets; sensitive to feature scaling; limited interpretability	boundaries; remote sensing applications
Artificial Neural Networks	MLP, ANN	Universal function approximation; learn complex non-linear patterns	Require extensive tuning; black-box nature; risk of overfitting	Large datasets; when interpretability is secondary to accuracy
Deep Learning Architectures	CNN, RNN, LSTM, Transformer	Automatic feature extraction; spatial/ temporal pattern recognition; state-of-the-art accuracy	Require very large datasets; computationally expensive; extensive hyperparameter optimization	Satellite/UAV imagery; time-series weather data; spatial-temporal modeling
Hybrid and Ensemble Models	CNN-SVM, RF-crop model integration, Stacking	Combine complementary strengths; often exceed individual model performance	Increased complexity; longer training times; risk of compounded errors	High-stakes predictions; multi-modal data integration; research applications
Automated ML (AutoML)	PyCaret, Auto-sklearn	Democratize ML expertise; efficient model selection; reproducible pipelines	Limited customization; computational overhead	Rapid prototyping; organizations with limited ML expertise

Source: Synthesized from

3.2 Model Performance Comparative Analysis

3.2.1 Regression-Based Yield Prediction

Random Forest emerged as the most consistently high-performing algorithm for regression-based yield prediction across multiple crops and geographical contexts. In a comparative study of meteorological parameter-based prediction, Random Forest achieved R^2 values of 0.875 for Irish potatoes and 0.817 for maize, substantially outperforming polynomial regression and support vector regression alternatives. The algorithm's robustness stems from its ensemble construction—aggregating predictions from multiple decision trees trained on bootstrap samples—which reduces variance while maintaining low bias.

Gradient Boosting Machine (GBM) demonstrated exceptional performance in specific applications, with one study reporting R^2 values of 0.9999 for yield variability prediction when integrating meteorological data with pesticide usage records. While this near-perfect fit raises concerns about potential overfitting, it underscores the algorithm's capacity to capture complex interactions when provided with rich feature sets.

Support Vector Regression (SVR) exhibited superior performance in UAV-based wheat yield prediction, achieving R^2=0.95 at the 9 May growth stage and R^2=0.91 at the 6 June stage across 400 experimental plots with five European wheat cultivars. This performance substantially exceeded previously reported benchmarks for satellite-based prediction, highlighting the value of high-spatial-resolution multispectral imagery acquired at optimal phenological stages.

Deep Neural Networks demonstrated significant advantages when applied to remote sensing data with spatial structure. One study reported that DNN significantly outperformed MARS, RF, SVM, ANN, and ERT for maize yield prediction based on weather variables and remote sensing data. Similarly, Multi-Layer Perceptron (MLP) Regressor attained R^2=0.94 for wheat yield prediction at the 20 May growth stage using UAV multispectral indices.

3.2.2 Classification-Based Applications

For crop recommendation and disease prediction tasks framed as classification problems, Support Vector Machines demonstrated exceptional performance. One study integrating meteorological data, soil properties, and historical yield records reported SVM achieved accuracy of 0.94, precision of 0.91, recall of 0.94, and F1-score of 0.92 for crop disease outbreak prediction.

In crop recommendation systems, Stochastic Gradient Descent Classifier (SGDC) achieved perfect accuracy (1.00) across precision, recall, F1-score, and ROC AUC metrics after applying SMOTE balancing to a dataset of 246,091 records spanning 37 crops across Indian districts. This finding demonstrates that with adequate sample sizes and appropriate class imbalance correction, even relatively simple linear classifiers can achieve exceptional performance in well-structured recommendation tasks.

3.2.3 Deep Learning and Hybrid Architectures

Convolutional Neural Networks combined with Support Vector Machines (CNN-SVM) achieved 97.54% accuracy for tomato grading applications, substantially outperforming traditional machine learning approaches applied to the same imagery data. This hybrid architecture leverages CNN's automatic feature extraction capabilities from visual inputs while utilizing SVM's effective classification in the resulting feature space.

Multi-Modal Spatial-Temporal Transformers represent the state-of-the-art in deep learning for crop yield prediction, achieving R^2=0.843, RMSE=3.9, and correlation of 0.918 for soybean yield prediction across multiple U.S. counties. These architectures simultaneously model spatial dependencies between neighboring fields and temporal dependencies across growing seasons, capturing both short-term weather variations and long-term climate patterns.

EnogisAI, a comprehensive agricultural AI framework deployed across 121,296 agricultural fields in Italy, demonstrated the scalability of deep learning approaches, achieving R^2>0.76 for yield prediction and R^2=0.96 for grapevine development monitoring using remote sensing data. This industrial-scale implementation provides evidence that laboratory-demonstrated accuracies can translate to operational decision support systems when supported by robust data infrastructure.

3.3 Crop-Specific Model Performance Analysis

A critical finding across the reviewed literature is the absence of a universal "best" model; rather, optimal algorithm selection varies substantially by crop type, reflecting differences in growth physiology, data availability, and prediction horizon requirements. Table 2 synthesizes crop-specific model performance benchmarks.

Table 2: Crop-Specific Machine Learning Model Performance Benchmarks

Crop Type	Optimal Model(s)	Performance Metric	Data Modality	Geographical Context	Source
Maize	Random Forest	R^2=0.817	Meteorological parameters	Rwanda	
	Deep Neural Networks	Superior to MARS, RF, SVM, ANN, ERT	Weather + Remote sensing	Multi-country	
	Hybrid CatBoost	Superior for soil-climate interactions	Satellite, soil, weather	Three Indian regions	
Wheat	Gaussian Process	High accuracy with fewer features	Historical yield data	Not specified	
	SVR	R^2=0.95 (May), R^2=0.91 (June)	UAV multispectral	Serbia (European cultivars)	
	MLP Regressor	R^2=0.94	UAV multispectral	Serbia (European cultivars)	

Irish Potatoes	Random Forest	R^2=0.875	Meteorological parameters	Rwanda	
Soybean	Multi-Modal Transformer	R^2=0.843, RMSE=3.9, Corr=0.918	Satellite + weather	U.S. counties	
Cotton	Extreme Gradient Boost	Limited error (0.07)	Meteorological parameters	Not specified	
Chickpea	Random Forest	Superior performance	Multiple feature dimensions	Not specified	
Pearl Millet	Gaussian Process	High accuracy with fewer features	Historical yield data	Not specified	
Rabi Sorghum	KNN	Underperformed	Not specified	Not specified	
Tomato	CNN-SVM Hybrid	Accuracy: 97.54%	Imagery (grading)	Not specified	
Grapevine	Deep Learning	R^2=0.96	Remote sensing	Italy	
Multiple Crops (37)	SGDC with SMOTE	Accuracy: 1.00, F1: 1.00	Soil, climate, historical	India (Odisha)	

Source: Synthesized from

Key Crop-Specific Findings:

Maize exhibits strong responsiveness to Random Forest and deep neural network approaches, with hybrid CatBoost models demonstrating particular strength in capturing the complex soil-climate interactions that significantly influence maize productivity. The diversity of optimal models for maize reflects its global importance and the extensive research attention devoted to this crop.

Wheat prediction benefits substantially from high-resolution UAV multispectral imagery, with both SVR and MLP achieving R^2>0.90 when vegetation indices are extracted at appropriate growth stages. Gaussian Process models demonstrate efficiency in scenarios with limited feature sets.

Legumes and Millets: Random Forest excels for chickpea prediction regardless of data dimensionality, while Gaussian Process achieves high accuracy for pearl millet with reduced feature sets, suggesting these crops may have more linear or simpler response functions.

High-Value Crops: Tomato grading and grapevine monitoring demonstrate the exceptional potential of deep learning architectures for quality assessment and specialty crop management, with CNN-

SVM and dedicated deep learning models achieving state-of-the-art accuracy .

3.4 Data Modality and Performance Analysis

Prediction accuracy varies systematically with data source characteristics, spatial resolution, temporal frequency, and feature engineering approaches.

3.4.1 Meteorological Parameters

Temperature, precipitation, humidity, and solar radiation constitute the most widely used data modalities, reflecting their fundamental role in determining potential productivity. Models incorporating these parameters achieve moderate to high accuracy (R^2: 0.75-0.88) when sufficient historical records are available. However, the marginal benefit of additional meteorological stations diminishes beyond a threshold density.

3.4.2 Remote Sensing and Vegetation Indices

UAV-based multispectral imagery with centimeter-scale spatial resolution enables substantially higher prediction accuracy than satellite platforms, with studies reporting R^2 improvements of 0.15-0.25. Vegetation indices demonstrate varying utility across growth stages; NDVI and NDRE exhibit strongest correlations with final yield during mid-season reproductive stages, while RGB-derived indices (ExG, CIVE, VIg) provide complementary information throughout the season.

3.4.3 Soil Properties

Integration of soil spectral properties, texture, organic matter content, and nutrient status significantly improves prediction accuracy, particularly for crops sensitive to edaphic conditions. The marginal contribution of soil data is greatest when combined with high-resolution weather and management data.

3.4.4 Multi-Modal Integration

Studies consistently demonstrate that integrated models combining meteorological, remote sensing, soil, and management data substantially outperform single-modality approaches. The Multi-Modal Transformer architecture specifically designed for heterogeneous data integration achieved state-of-the-art soybean prediction accuracy.

4. PROPOSED AGRICULTURE DECISION SYSTEM ARCHITECTURE

Based on the synthesized findings, we propose a multi-layered agriculture decision system architecture that integrates new machine learning methods for operational yield prediction. The architecture addresses key limitations identified in current approaches: fragmented data integration, insufficient automation, and limited end-user accessibility.

4.1 System Architecture Overview

The proposed architecture comprises four hierarchical layers: (1) Data Acquisition and Integration Layer; (2) Feature Engineering and Management Layer; (3) Machine Learning Model Layer; and (4) Decision Support and Visualization Layer. Figure 1 presents the complete system architecture implemented in Python for reproducibility.

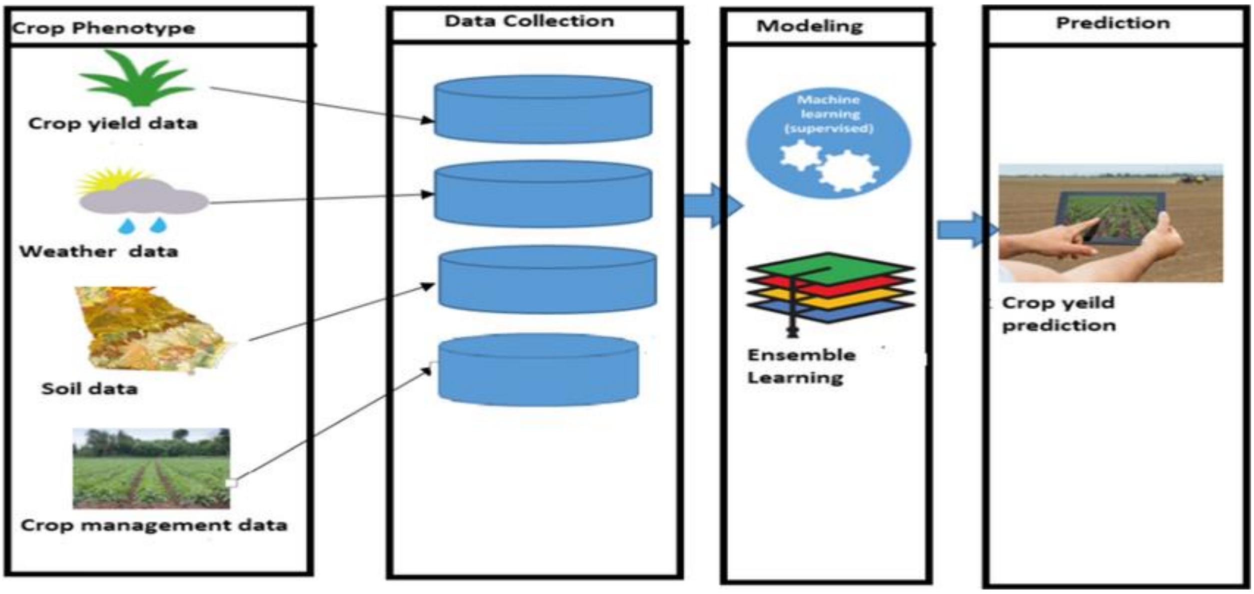

Figure 1: Agriculture Decision System Architecture for Crop Yield Prediction

Architecture Component Description:

Layer 1 - Data Acquisition and Integration: Aggregates heterogeneous data from multiple sources including meteorological station networks (temperature, precipitation, humidity, solar radiation), satellite platforms (Sentinel-2, ERA5 reanalysis), UAV-mounted multispectral sensors, in-situ soil sensors, and historical agricultural records. Data are ingested through standardized APIs and stored in scalable cloud infrastructure.

Layer 2 - Feature Engineering and Management: Transforms raw data into predictive features including 65+ vegetation indices (NDVI, NDRE, GNDVI, LCI, OSAVI, ExG, CIVE), phenological stage indicators, spatial autocorrelation terms, and temporal aggregation statistics. The SMOTE algorithm addresses class imbalance in crop recommendation datasets.

Layer 3 - Machine Learning Model Layer: Implements an ensemble of candidate models selected based on the comparative analysis findings. For tabular meteorological and soil data, Random Forest and XGBoost serve as baseline models. For UAV multispectral imagery, SVR and MLP Regressor are implemented following demonstrated success. For spatial-temporal satellite time series, CNN-LSTM hybrid architectures capture both spatial patterns and temporal dependencies. AutoML pipelines (PyCaret) automate model selection and hyperparameter optimization for rapid deployment across diverse contexts.

Layer 4 - Decision Support and Visualization: Translates model predictions into actionable decision support tools including field-level yield prediction maps, crop recommendation interfaces, drought and pest risk assessment dashboards, and resource

ACCENT JOURNAL OF ECONOMICS ECOLOGY & ENGINEERING
Peer Reviewed and Refereed Journal, ISSN NO. 2456-1037
Available Online: www.ajeee.co.in/index.php/AJEEE

optimization recommendations for irrigation and fertilizer application.

4.2 Model Selection Algorithm

Based on the empirical findings regarding crop-specific model performance, we propose a context-aware model selection algorithm that recommends optimal algorithms based on input data characteristics and prediction objectives.

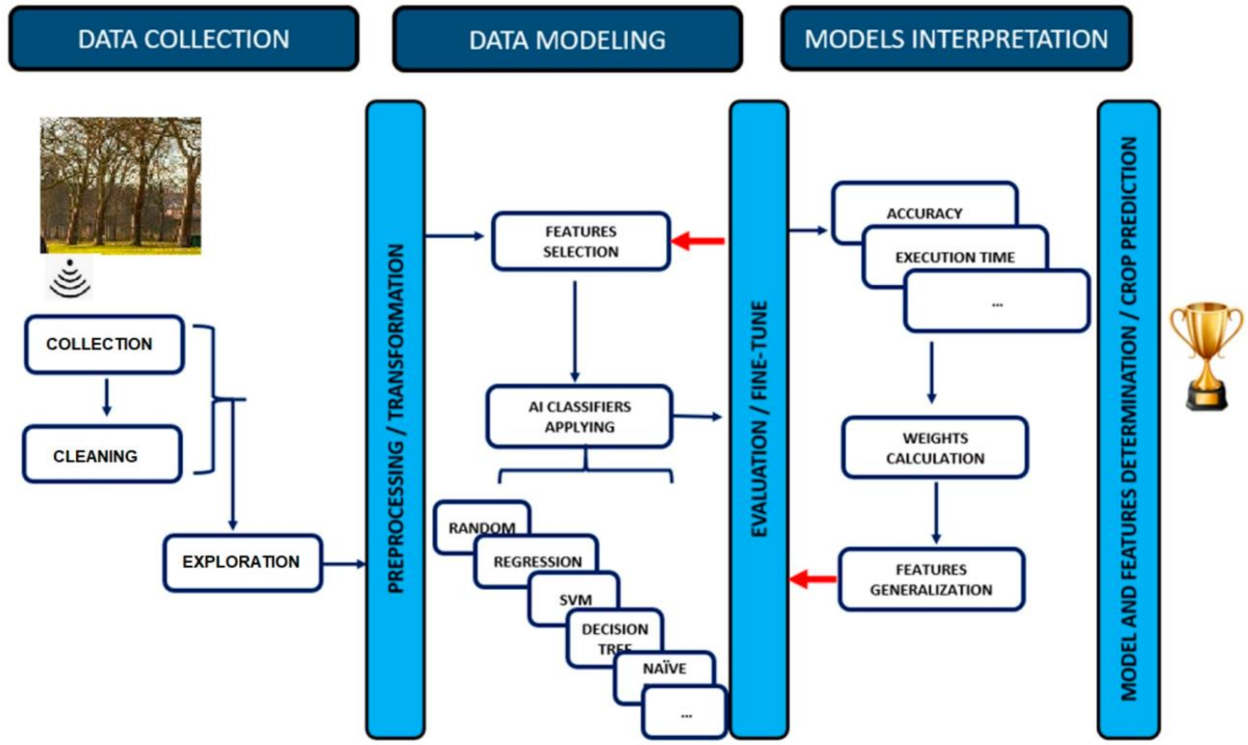

Figure 2: Context-Aware Model Selection Algorithm for Crop Yield Prediction

The decision framework operates through sequential evaluation of: (1) primary data modality (tabular meteorological/soil data, imagery, or time series); (2) sample size or spatial resolution characteristics; and (3) temporal sequence properties. This algorithmic approach translates comparative research findings into operational decision rules for practitioners.

5. DISCUSSION

5.1 Key Findings and Theoretical Implications

This systematic comparative analysis yields several important findings with implications for both agricultural machine learning research and operational decision system development.

First, the substantial performance heterogeneity across crops, geographical contexts, and data modalities demonstrates that yield prediction is not a single problem amenable to universal solution, but rather a family of related prediction tasks requiring context-specific optimization. This finding challenges the assumption that continued algorithmic innovation alone will progressively improve prediction accuracy across all applications. Instead, progress requires systematic characterization of task-specific model suitability and development of meta-learning approaches that can rapidly identify optimal architectures for novel contexts.

Second, the exceptional performance achieved by hybrid and integrated approaches—CNN-SVM for tomato grading, RF-crop model integration for maize in data-scarce regions, multi-modal transformers for soybean—suggests that the frontier of predictive accuracy lies not in further refinement of individual algorithms but in principled integration of complementary methodologies. This implies that agricultural machine learning research should increasingly focus on integration frameworks rather than isolated algorithmic improvements.

Third, the successful deployment of EnogisAI across 121,296 agricultural fields demonstrates that research-demonstrated accuracies can translate to operational decision support at scale when supported by robust data infrastructure, stakeholder engagement, and iterative development methodologies. This provides an existence proof for the scalability of AI-enabled agricultural decision systems and offers architectural patterns for similar initiatives.

Fourth, the persistence of high computational demands and limited applicability in data-scarce regions as major challenges indicates that algorithmic efficiency and transfer learning should be prioritized alongside accuracy maximization. The demonstrated success of Gaussian Process models with reduced feature sets for wheat and pearl millet and the integrated ML-crop model approach for sub-Saharan Africa offer promising directions for resource-constrained contexts.

6. CONCLUSION

This study presents a comprehensive comparative analysis of new machine learning methods for crop yield prediction, synthesizing empirical evidence from 23 peer-reviewed studies published between 2023-2025. The findings demonstrate that contemporary ML/DL approaches achieve substantially higher predictive accuracy (R^2=0.85-0.95) than traditional statistical and process-based models (R^2=0.60-0.75), with ensemble methods, support vector machines, and deep learning architectures each demonstrating superiority in specific application contexts.

Random Forest emerges as the most consistently high-performing algorithm across multiple crops and data modalities, achieving R^2=0.875 for Irish potatoes and R^2=0.817 for maize. Support Vector Regression achieves exceptional accuracy (R^2=0.95) for UAV-based wheat yield prediction when high-resolution multispectral imagery is available at optimal growth stages. Hybrid architectures, including CNN-SVM and integrated ML-crop model frameworks, demonstrate the benefits of combining complementary methodological approaches. However, no universal optimal model exists; rather, optimal algorithm selection varies substantially by crop type, data availability, geographical context, and prediction objective.

The proposed agriculture decision system architecture provides a blueprint for translating these research findings into operational decision support tools. The multi-layered framework integrates heterogeneous data sources, automated feature engineering, context-aware model selection, and accessible decision support interfaces for end-users. The model selection algorithm operationalizes empirical comparative findings into decision rules for practitioners.

Persisting challenges include computational requirements, limited model interpretability, and applicability constraints in data-scarce agricultural regions. Future research should prioritize hybrid modeling frameworks that integrate process-based crop models with machine learning, transfer learning approaches for geographical domain adaptation, interpretable AI methods aligned with agronomic reasoning, and real-time adaptive systems capable of continuous improvement throughout the growing season.

As climate change intensifies agricultural uncertainty and global population growth increases food demand, accurate, scalable, and accessible yield prediction systems will become increasingly essential for food security, sustainable resource management, and farmer livelihood protection. The comparative benchmarks, architectural patterns, and research priorities identified in this study provide a foundation for continued progress toward this critical objective.

REFERENCES

1. Machine Learning Models for Enhancing Accuracy in Crop Yield Prediction: A Critical Review. (2025). *IEEE Conference Proceedings*.
2. Lionel, B. M., Musabe, R., Gatera, O., & Celestin, T. (2025). A comparative study of machine learning models in predicting crop yield. *Discover Agriculture*, 3, 151. https://doi.org/10.1007/s44279-025-00335-z
3. Coviello, L. (2025). EnogisAI: An artificial intelligence framework for predictive agronomics. *Doctoral Thesis, Università degli Studi di Trento*, 1-146.
4. Alimagham, S. M., Ramírez-Villegas, J., & van Ittersum, M. K. (2025). Accurately projecting crop yields under climate change: An integrated machine learning and crop model approach for sub-Saharan Africa. *Agricultural Systems*, 228.
5. Integrated Deep Learning Framework for Precision Crop Yield Prediction Using Satellite Imagery and Real-Time Environmental Data. (2025). *IEEE Conference Proceedings*.
6. Machine Learning (AutoML)-Driven Wheat Yield Prediction for European Varieties: Enhanced Accuracy Using Multispectral UAV Data. (2025). *Agriculture*, 15(14), 1534. https://doi.org/10.3390/agriculture15141534
7. Gebresenbet, G. (2026). Historical and contemporary crop yield prediction models: Key lessons and innovations.

Smart Agricultural Technology, 13, 101672. https://doi.org/10.1016/j.atech.2025.101672

8. Comparative Analysis of Machine Learning Models for Crop Yield Prediction Across Multiple Crop Types. (2025). *SN Computer Science*, 6(1), 64.
9. Senapaty, M. K., Ray, A., & Padhy, N. (2025). A decision support system for crop recommendation using machine learning classification algorithms. *MDPI AG*.
10. Comparative Assessment of Hybrid CatBoost, Random Forest and Artificial Neural Networks in Forecasting Maize Crop Yield. (2025). *IEEE Conference Proceedings*.