

Predictive Artificial Intelligence for Multi-Crop Yield Optimization in Bangladesh under Climate Variability

ONJON SUTRADHAR

2431963

MSc Data And Artificial Intelligence

Abstract

Agriculture in Bangladesh is highly vulnerable to climate variability, including irregular rainfall, flooding, and rising temperatures. Traditional crop yield prediction methods are largely experience-driven or based on simplified statistical assumptions, limiting their ability to adapt to complex environmental interactions. This research proposes an artificial intelligence-based framework for multi-crop yield prediction in Bangladesh, leveraging publicly available agricultural and climatic datasets. By comparing traditional predictive models with machine learning approaches, the study aims to improve yield forecasting accuracy and support data-driven agricultural decision-making in climate-sensitive farming systems.

1 Introduction

1.1 Context and Background

Agriculture plays a central role in Bangladesh's food security and rural livelihoods, yet it remains highly exposed to climate variability. Despite the country's dependence on agriculture, most farming decisions continue to rely on experience rather than predictive analytics. Conventional yield prediction techniques struggle to capture non-linear relationships among climate variables, soil conditions, and crop responses.

Recent advances in artificial intelligence provide an opportunity to move toward predictive and adaptive agricultural systems. By learning complex patterns from large-scale datasets, AI-based models can improve yield stability and reduce climate-related risks.

This research focuses specifically on Bangladesh as the target country. A multi-crop approach is adopted to reflect real agricultural diversity rather than restricting the analysis to a single crop. The study is based on a publicly available dataset hosted on Mendeley, which contains yield and associated environmental variables for multiple

crops cultivated in Bangladesh. The dataset is documented in a peer-reviewed descriptor published in the *Data in Brief* journal, ensuring data reliability and academic rigor.

1.2 Research Problem

Although agricultural and climatic datasets are increasingly available, Bangladesh lacks integrated analytical frameworks that translate these data into actionable yield predictions. Existing approaches rely primarily on linear or simplified statistical models that perform poorly under climate variability and fail to generalize across crops.

This research addresses the problem of unreliable yield forecasting by investigating whether AI-based models can improve prediction accuracy and robustness. Improved yield prediction directly supports climate preparedness, resource planning, and risk reduction in a highly climate-sensitive agricultural system.

1.3 Research Aim and Objectives

The primary aim of this study is to design and evaluate an AI-driven framework for predicting crop yield in Bangladesh under variable climatic conditions.

1. To develop a data-driven predictive framework for multi-crop yield estimation in Bangladesh using publicly available climatic and environmental datasets.
2. To compare the performance of traditional statistical models with machine learning-based approaches for crop yield prediction.
3. To evaluate model robustness under varying climatic conditions using standard regression and error-based performance metrics.
4. To identify and analyze the most influential climatic and environmental factors affecting crop yield through feature importance assessment.
5. To assess the practical applicability of the proposed predictive model for supporting precision agriculture and informed decision-making.

1.4 Significance of the Research

The novelty of this work lies in its geographic and contextual focus. Despite its dependence on agriculture, Bangladesh remains underrepresented in AI-driven yield prediction research.

Improved predictive capability can support farmers and agricultural planners by providing data-driven insights into the relationship between climate variability and crop

productivity. The research may contribute to improved agricultural planning, climate preparedness, and more efficient resource management.

While the present work focuses on numerical environmental variables, the framework can be extended in the future through integration with remote sensing and computer vision-based agricultural monitoring systems.

2 Literature Review

Previous research demonstrates that machine learning models, including neural networks and ensemble methods, outperform conventional statistical models in capturing non-linear agricultural processes. However, much of the existing literature focuses on controlled datasets or single-factor analysis, limiting real-world applicability.

There is also a notable gap in studies that balance predictive accuracy with interpretability, which is essential for farmer trust and adoption. This research positions itself within this gap by emphasizing both performance and usability, synthesizing insights from AI, agronomy, and climate science.

Table 1: Comparison of Selected Studies (Part I)

Paper	Summary	Methodology	Limitations
Comparative Analysis of Machine Learning Models for Crop Yield Prediction Across Multiple Crop Types (2025) Yashraj Patil et al.	This study compares several machine learning regression models—Gaussian Process, Random Forest, KNN, XGBoost, and AdaBoost—for predicting yields of six crops. Results show that no single model performs best across all crops; performance is crop-specific.	Multiple ML models were trained and evaluated on six crop datasets with systematic feature testing.	Dataset-specific findings limit generalization across regions.
Enhanced precision Multi-Crop farming using Machine Learning Recommendations (2024) Prem Kumar et al.	A real-time crop recommendation system using soil nutrient sensor data to determine optimal crop selection.	Classification algorithms trained on soil nutrient datasets.	Limited environmental and regional validation.
Simultaneous multi-crop land suitability prediction from remote sensing data using semi-supervised learning (2023) Amanjot Bhullar et al.	Semi-supervised deep learning model for multi-crop land suitability prediction.	MLP architecture using environmental datasets.	Limited dataset size and missing spatial dependencies.

Table 2: Comparison of Selected Studies (Part II)

Paper	Summary	Methodology	Limitations
Hybrid deep learning with a fusion model of weather and soil-based multi-crop recommendations (2025) Mohan M. et al	Integrated deep learning framework combining CNN and graph-based LSTM for multi-crop recommendation.	Spatial-temporal modelling with optimization techniques and explainable AI.	Computational complexity and deployment limitations.
UniCrop: Universal Data Pipeline for Crop Yield Prediction (2026) Emiliya Khidirova et al.	Automated data engineering pipeline integrating satellite and environmental data sources.	Feature engineering and ML models including LightGBM.	Validated using limited datasets.
Multi-Crop Classification Using Feature Selection-Coupled ML Classifiers (2022) Shan He et al.	Evaluation of feature selection techniques combined with ML classifiers for crop classification.	Feature extraction from remote sensing data with ML classifiers.	Limited geographic and temporal coverage.

3 Research Design and Methodology

3.1 Research Design

A quantitative research design is adopted. Historical crop yield data are combined with weather records and environmental variables obtained from publicly available agricultural datasets. Predictive models will be developed and evaluated to determine whether artificial intelligence techniques can improve crop yield prediction accuracy.

3.2 Research Methods

Baseline statistical models such as linear regression will be compared with machine learning models including K-Nearest Neighbors, Random Forest, and ensemble techniques.

Model performance will be evaluated using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and coefficient of determination (R^2).

3.3 Materials

Research Data

Primary data are sourced from the publicly available Mendeley dataset:

<https://data.mendeley.com/datasets/cphdw4z5kw/2>

Software

Python, Pandas, NumPy, Scikit-learn, Matplotlib.

Hardware

Standard computing hardware capable of running machine learning experiments.

Table 3: Summary of Recent Crop Modeling and Classification Studies

Study	Focus	Data Source	Methods / Models	Key Contribution	Limitations
Multi-crop Regression	ML	Crop datasets	Gaussian Process, RF, KNN, XGBoost	Demonstrated crop-specific model performance	Limited generalization
Real-time based Recommendation	Soil-Crop	Soil nutrient sensors	Decision Tree	Real-time recommendation	Limited environmental variables
Semi-supervised Land Suitability		Remote sensing datasets	MLP	Multi-crop suitability modelling	Small dataset
Hybrid Learning Framework	Deep	Soil and weather data	CNN + LSTM	Spatial-temporal modelling	High computational cost
UniCrop Pipeline	Data	Satellite and environmental datasets	LightGBM, RF, SVR	Automated feature engineering	Limited validation
FS-ML coupling	Cou-	Remote sensing data	RF, SVM, KNN	Improved classification accuracy	Limited scope

4 Expected Outcomes

At the proposal stage, empirical results are not yet available. However, it is anticipated that machine learning models will demonstrate improved predictive performance compared with traditional statistical approaches. The research may also identify environmental variables that significantly influence crop production patterns in Bangladesh.

5 Timeline

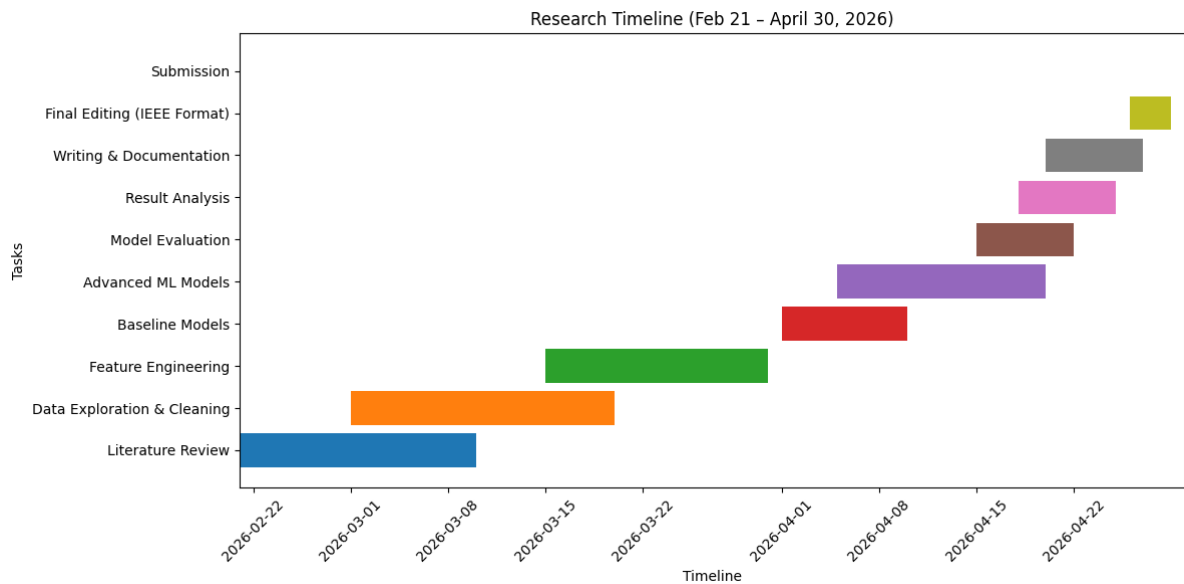


Figure 1: Research Timeline

References

- [1] M. A. Rahman, M. R. Islam, and S. Ahmed, “Multi-crop yield and climate data for Bangladesh,” Mendeley Data, 2025.
- [2] M. A. Rahman, M. R. Islam, and S. Ahmed, “Dataset on crop yield and associated climatic variables in Bangladesh,” *Data in Brief*, 2025.
- [3] Y. Patil et al., “Comparative Analysis of Machine Learning Models for Crop Yield Prediction Across Multiple Crop Types,” 2025.
- [4] P. Kumar et al., “Enhanced Precision Multi-Crop Farming Using Machine Learning Recommendations,” 2024.
- [5] A. Bhullar et al., “Simultaneous Multi-Crop Land Suitability Prediction from Remote Sensing Data Using Semi-Supervised Learning,” 2023.
- [6] M. Mohan et al., “Hybrid Deep Learning with a Fusion Model of Weather and Soil-Based Multi-Crop Recommendations,” 2025.
- [7] E. Khidirova et al., “UniCrop: A Universal Data Pipeline for Crop Yield Prediction,” 2026.
- [8] S. He et al., “Multi-Crop Classification Using Feature Selection-Coupled Machine Learning Classifiers,” 2022.