The Scientific Temper (2025) Vol. 16 (10): 4916-4922
Doi: 10.58414/SCIENTIFICTEMPER.2025.16.10.10
E-ISSN: 2231-6396, ISSN: 0976-8653
https://scientifictemper.com/

RESEARCH ARTICLE

Location Specific Paddy Yield Prediction using Monte Carlo Simulation incorporated Long Short-Term Memory

V. Mahalakshmi[1*], M. Manimekalai[2]

Abstract

Accurately predicting paddy yield is vital for food security and efficient farm management. This work proposes LSPYP-ML, a framework that combines fuzzy logic, Monte Carlo simulation, and Long Short-Term Memory (LSTM) networks to improve prediction accuracy. The fuzzy module cleans and classifies uncertain data such as rainfall, temperature, and pesticide use. The Monte Carlo module simulates extreme weather scenarios to account for environmental variability. Finally, the LSTM module captures temporal patterns in climate and yield data for robust forecasting. Experiments show that the framework achieves higher accuracy, precision, sensitivity, specificity, and F-Score compared to existing methods. LSPYP-ML offers a reliable decision-support tool for farmers and policymakers to enhance productivity and manage climate risks.

Keywords: Paddy yield prediction, Fuzzy logic, Monte Carlo simulation, LSTM, Agricultural forecasting

Introduction

Predicting crop yields accurately is essential for ensuring food security and helping farmers and policymakers make better decisions. Paddy, being a staple crop, is particularly vulnerable to changes in rainfall, temperature, and pest activity, which makes its yield difficult to forecast [1], [2]. With climate change adding to this unpredictability, traditional models struggle to capture the true complexity of yield fluctuations [3].Older statistical and machine learning models tend to rely too much on fixed patterns in historical data. This makes them less effective when dealing with uncertainties like irregular rainfall, sudden temperature shifts, or noisy and incomplete records [4], [5]. Modern techniques such as fuzzy logic, probabilistic simulations, and deep learning are better suited because they can handle these uncertainties and reveal complex patterns in agricultural data [6].

Recent innovations in artificial intelligence have further improved yield prediction. Methods like LSTM [8] and CNNs [9] are now capable of capturing both temporal and spatial variations in crop growth. At the same time, IoT devices and satellite-based remote sensing provide continuous field data [10], while cloud and edge computing make analysis faster and more accessible. Even quantum computing is being explored for its ability to process vast datasets more efficiently [7].Since growing conditions vary widely across regions, location-specific prediction models are far more reliable than generalized ones. By tailoring forecasts to local soil, climate, and farming practices, farmers can optimize irrigation, fertilization, and pest control. This study presents LSPYP-ML, a framework that combines fuzzy logic, Monte Carlo simulation, and LSTM networks to better manage uncertainties, capture time-based trends, and deliver accurate location-specific forecasts for paddy yield.

2. Existing Methods

Several advanced approaches have been proposed for crop yield prediction, combining deep learning, hybrid modeling, and emerging computing technologies. A hybrid quantum deep learning model integrated Bi-LSTM, XGBoost, and

[1]Research Scholar (PT), Assistant Professor, PG & Research Department of Computer Science, Shrimati Indira Gandhi College, Affiliated to Bharathidasan University, Trichy–620002, Tamil Nadu, India.

[2]Research Head & Supervisor, PG & Research Department of Computer Science, Shrimati Indira Gandhi College, Affiliated to Bharathidasan University, Trichy–620002, Tamil Nadu, India.

***Corresponding Author:** V. Mahalakshmi, Research Scholar (PT), Assistant Professor, PG & Research Department of Computer Science, Shrimati Indira Gandhi College, Affiliated to Bharathidasan University, Trichy–620002, Tamil Nadu, India, E-Mail: mahalakshmiv51979@gmail.com

How to cite this article: Mahalakshmi, V., Manimekalai, M. (2025). Location Specific Paddy Yield Prediction using Monte Carlo Simulation incorporated Long Short-Term Memory. The Scientific Temper, **16**(10): 4916-4922.

Doi: 10.58414/SCIENTIFICTEMPER.2025.16.10.10

Source of support: Nil

Conflict of interest: None.

Received: 13/09/2025 **Accepted:** 05/10/2025 **Published:** 16/10/2025

quantum circuits to boost rice yield forecasting accuracy [11]. Deep learning models such as CNN-LSTM with attention have been applied to multi-source datasets in Northeast China for predicting maize, rice, and soybean yields [12]. Other works introduced deep neural network frameworks that considered climate change impacts on crop yields in Australia [13], while hybrid feature selection with optimized machine learning improved predictive efficiency by focusing on relevant variables [14]. These methods demonstrate the potential of AI-driven forecasting but still face challenges in handling environmental uncertainties, extreme weather events, and location-specific variations, highlighting the need for more robust solutions.

The key points about these methods are enumerated in Table 1.

Existing methods improve yield prediction but still struggle with environmental uncertainties, extreme weather, and location-specific variations, highlighting the need for more advanced models.

Background

Fuzzy logic and Long Short-Term Memory (LSTM) networks have emerged as key techniques for predictive modeling under uncertainty. Fuzzy logic is effective for handling vague or imprecise environmental factors such as rainfall, temperature, and pesticide use by classifying them into linguistic categories like “low,” “moderate,” or “high” [15]. This approach improves reliability when data is incomplete or noisy, making it more adaptable for real-world agricultural applications. It also supports expert-driven rule-based decision-making, which enhances interpretability and transparency.

LSTM networks, on the other hand, are powerful deep learning models designed to capture long-term dependencies in sequential data, overcoming the vanishing gradient issues of traditional RNNs [16], [17]. They are particularly suited for agricultural forecasting, where historical weather and soil patterns influence future yields. By integrating fuzzy logic preprocessing with LSTM forecasting, predictive frameworks can combine structured expert knowledge with advanced temporal modeling, resulting in more accurate and robust paddy yield predictions under dynamic environmental conditions.

Proposed Method

The LSPYP-ML framework consists of three modules—Fuzzy-based Domain-driven Data Preprocessor (FDDP), Monte Carlo Simulation for Environmental Variables (MCSEV), and LSTM-based Yield Prediction (LYP)—which work together to handle uncertainty, model environmental variability, and deliver accurate yield forecasts.

Fuzzy based Domain-driven Data Preprocessor

The Fuzzy-based Domain-driven Data Preprocessor (FDDP) uses fuzzy logic to manage uncertainty in agricultural data such as rainfall, temperature, pesticide use, and crop yield. It converts continuous values into categories for easier interpretation, handles missing or noisy records with membership functions, and applies imputation techniques to keep datasets accurate and complete. Key yield prediction factors and their units are listed in Table 2.

These input parameters are treated as a Set in FDDP module, represented as $P_{t_0} = \{\rho_{1_{t_0}}, \rho_{2_{t_0}} \dots \rho_{12_{t_0}}\}$ where $\rho_{1_{t_0}}, \rho_{2_{t_0}} \cdots$ are the input parameters at timestamp t_0, given in the same sequence as in Table 2.

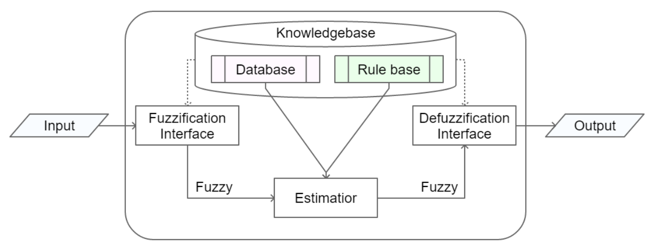

Figure 1: Fuzzy Logic

Table 1: Existing methods Key points

Author	*Work*	*Methodology*	*Advantages*	*Limitations*
De Rosal Ignatius Moses Setiad et.al.,	Hybrid Quantum Deep Learning Model (HQDLM)	Quantum computing, Bi-LSTM, XGBoost	High predictive accuracy for rice yield, Enhanced feature extraction	Relies on quantum simulators, Requires high computational resources, .
Jian Lu et al.	Deep Learning for Multi-Source Data-Driven Crop Yield Prediction (DLMSDD)	CNN, LSTM, Attention mechanism, kNDVI,	Superior accuracy in crop yield prediction, Enhances feature extraction	Heavily relies on high-quality multi-source data, Computational complexity
Haydar Demirhan	Deep Learning Framework for Crop Yield Prediction in Australia (DLFCC)	Deep neural network (DNN),	Outperforms benchmark models, High flexibility in handling low-frequency data,	Short observation period and limited availability of climate change proxies
Mahmoud Abdel-Salam et al.	Hybrid Feature Selection Approach and Optimized Machine Learning (HFSAOML)	Hybrid feature selection, K-means clustering,	Hybrid feature selection improves model efficiency,	Limited applicability to other agricultural contexts or crop types.

Table 2: Input parameters' details

S. No	*Parameter*	*Context*	*Unit*
1	Rainfall	Climatic	mm
2	Temperature	Climatic	°C
3	Humidity	Climatic	%
4	Solar Radiation	Climatic	W/m^2
5	Wind Speed	Climatic	km/h
6	Soil pH	Soil	H^+
7	Soil Moisture	Soil	%
8	Organic Matter	Soil	%
9	Pesticide Use	Agronomic	kg/ha
10	Fertilizer Use	Agronomic	kg/ha
11	Land Surface Temperature	Environmental	°C
12	Evapotranspiration	Environmental	mm/day

Interpolation based Missing Value Estimator (IMVE), Fuzzy Anomaly Detector (FAD) and Fuzzy Data Classifier (FDC) are the three fundamental components used to build the FDDP Module.

Interpolation based Missing Value Estimator

The Interpolation-based Missing Value Estimator (IMVE) fills data gaps by using interpolation to predict missing values from existing patterns. This keeps datasets consistent, reduces bias from random imputation, and preserves quality for reliable analysis, especially when gaps are small and trends are predictable.Since the input parameters are measured periodically in different timestamps, the parameters sets are portioned out for different timestamps $t_1, t_2, t_3 \ldots$ such as $P_{t_1}, P_{t_2}, P_{t_3} \ldots$. Then for any missing x^{th} parameter value $\rho_{x_{t_y}}$ at $t_y{}^{th}$ timestamp can be calculated using equation 1.

$$\rho_{x_{t_y}} = \frac{(t_y - t_{y-2})(t_y - t_{y-3})(t_y - t_{y-4})}{(t_{y-1} - t_{y-2})(t_{y-1} - t_{y-3})(t_{y-1} - t_{-4})} \times \rho_{x_{t_{y-1}}} + \frac{(t_y - t_{y-1})(t_y - t_{y-3})(t_y - t_{y-4})}{(t_{y-2} - t_{y-1})(t_{y-2} - t_{y-3})(t_{y-1} - t_{-4})} \times \rho_{x_{t_{y-2}}} + \frac{(t_y - t_{y-1})(t_y - t_{y-2})(t_y - t_{y-4})}{(t_{y-3} - t_{y-1})(t_{y-3} - t_{y-2})(t_{y-1} - t_{-4})} \times \rho_{x_{t_{y-3}}} + \frac{(t_y - t_{y-1})(t_y - t_{y-2})(t_y - t_{y-3})}{(t_{y-4} - t_{y-1})(t_{y-4} - t_{y-2})(t_{y-4} - t_{-3})} \times \rho_{x_{t_{y-4}}}$$

Equation (1)

IMVE works by assigning higher weights to nearby known data points and lower weights to distant ones, ensuring smooth transitions through missing values. This weighted interpolation preserves natural trends, maintains consistency, and provides reliable estimates for use in agriculture, environmental monitoring, and predictive modeling.

Fuzzy Anomaly Detector

The Fuzzy Anomaly Detector (FAD) uses fuzzy logic to spot outliers in uncertain or noisy agricultural data. Instead of strict thresholds, it applies graded membership functions, making detection more flexible and reducing false positives. This helps identify errors or abnormal events in large, variable datasets with greater reliability.

Algorithm 1: FAD

Input: $P_{t_1}, P_{t_2}, P_{t_3} \ldots P_{t_n}$

Output: Input parameter anomalies

Step 1: Load input data Sets $P_{t_1}, P_{t_2}, P_{t_3} \ldots P_{t_n}$

Step 2: $\forall i = 1 \rightarrow 12 :=$

Step 3: Let $\rho_{i_{t_y}}^{min}$ and $\rho_{i_{t_y}}^{max}$ be the minimum and maximum values of any parameter $\rho_{i_{t_y}}$

Step 4: $\forall j = 1 \rightarrow (n-1) :=$

Step 5: Compute $\Delta = \left| \rho_{i_{t_j}} - \rho_{i_{t_{j+1}}} \right|$

Step 6: $\rho_{i_{t_y}}^{min} = \begin{cases} \Delta \, if \, \Delta < \rho_{i_{t_y}}^{min} \\ \rho_{i_{t_y}}^{min} \, otherwise \end{cases}$

Step 7: $\rho_{i_{t_y}}^{max} = \begin{cases} \Delta \, if \, \Delta > \rho_{i_{t_y}}^{max} \\ \rho_{i_{t_y}}^{max} \, otherwise \end{cases}$

Step 8: end $\forall j$

Step 9: end $\forall \mathrm{i}$

Step 10: Let $P_{t_{n+1}}$ be the new input data with members $\left\{ \rho_{1_{t_{n+1}}}, \rho_{2_{t_{n+1}}} \ldots \rho_{12_{t_{n+1}}} \right\}$

Step 11: $\forall i = 1 \rightarrow 12 :=$

Step 12: $Anomaly = \begin{cases} FALSE \, if \, \rho_{i_{t_y}}^{min} \leq \rho_{i_{t_{n+1}}} \leq \rho_{i_{t_y}}^{max} \\ TRUE \, otherwise \end{cases}$

Step 13: end $\forall \mathrm{i}$

Step 14: return

The Fuzzy Anomaly Detector module identifies anomalies by applying above algorithm 1, which utilizes fuzzy logic for flexible and precise detection. By handling uncertainty and imprecise data, it effectively differentiates normal patterns from outliers. This ensures accurate anomaly detection in large, noisy datasets across various domains.

Fuzzy Data Classifier

The Fuzzy Data Classifier (FDC) assigns data to overlapping categories using fuzzy logic, allowing partial membership instead of rigid boundaries. This flexible approach is well-suited for ambiguous agricultural datasets, providing more nuanced classifications and supporting better decision-making where traditional methods fall short. Let $\rho_{x_{t_y}}^{norm}$ be the normalized value of an input parameter ρ_x at y^{th} timestamp which is calculated using equation 2.

$$\rho_{x_{t_y}}^{norm} = \frac{\left(\rho_{x_{t_y}} - \rho_{x_{t_y}}^{min} \right)}{\rho_{x_{t_y}}^{max} - \rho_{x_{t_y}}^{min}} \qquad \text{Equation (2)}$$

Then the input parameters values are classified by equation 3 as follows

$$Classification\,\Psi\left(\rho_{x_{t_y}}\right) = \begin{cases} Low\,if\,\rho_{x_{t_y}}^{norm} < \frac{1}{3} \\ Moderate\,if\,\frac{1}{3} \le \rho_{x_{t_y}}^{norm} < \frac{2}{3} \\ High\,otherwise \end{cases} \quad \text{Equation (3)}$$

In this module, fuzzy logic classifies environmental variables into categories such as "Low," "Moderate," or "High," making data more interpretable. This ensures the dataset is clean, consistent, and reliable for further analysis.

Monte Carlo Simulation for Environmental Variables

Monte Carlo Simulation for Environmental Variables (MCSEV) models uncertainty by simulating different scenarios for rainfall, temperature, humidity, and solar radiation, producing a range of outcomes that make yield forecasts more resilient to real-world variability.For every normal input parameter Set P_{t_y}, the Monte Carlo simulated variable set $P_{t_y}^{\Xi}$ is computed using the following algorithm.

Algorithm 2: Monte Carlo Simulated Variable Calculation

Input: $P_{t_1}, P_{t_2}, P_{t_3} \ldots P_{t_n}$

Output: $P_{t_1}^{\Xi}, P_{t_2}^{\Xi}, P_{t_3}^{\Xi}, \ldots P_{t_n}^{\Xi}$

Step 1: Load input data $P_{t_1} \rightarrow P_{t_n}$

Step 2: Let Γ be the set of mean and standard deviation 2-tuple set as $\left\{\left(\mu_{\rho_1}, \sigma_{\rho_1}\right), \ldots \left(\mu_{\rho_{12}}, \sigma_{\rho_{12}}\right)\right\}$

Step 3: $\forall i = 1 \rightarrow 12 :=$ Compute μ_{ρ_i} and $\sigma_{\rho_i} \in \Gamma$

Step 4. For $\mathrm{i} = 1 \rightarrow n$

Step 5: Set Distribution $N(\Gamma)$

Step 6: Compute $P_{t_i}^{\Xi} = \mathrm{e}^{P_{t_i}}$ as Log-normal

Step 7: End for i

Step 8: return

The normal input parameters sets $P_{t_1} \ldots P_{t_n}$, along with their individual members classifications, and the Monte Carlo simulated parameters sets $P_{t_1}^{\Xi} \ldots P_{t_n}^{\Xi}$ will serve as the foundational data for the LSTM-based Yield Prediction module.

LSTM based Yield Prediction (LYP)

The LYP module uses outputs from FDDP and MCSEV—combining deterministic and probabilistic data—to enhance yield prediction accuracy. It applies LSTM's core components, with equations for the Forget Gate, Input Gate, Candidate Cell State, Cell State Update, Output Gate, and Hidden State Update, ensuring robust temporal modeling.

LYP Forget Gate equation:

$$f_t = \sigma\left(w_f \cdot \left[h_{t-1}, \forall\lambda = 1 \rightarrow n := \left(P_{t_\lambda} || P_{t_\lambda}^{\Xi}\right) \otimes \Psi\left(P_{t_\lambda}\right)\right] + b_f\right) \quad \text{Equation (4)}$$

where f_t is the activation function of forget gate, w_f is the forget gate weight, b_f is the bias value, ⊗ refers a dedicated impact operator

LYP Input Gate equation:

$$i_t = \sigma\left(w_i \cdot \left[h_{t-1}, \forall\lambda = 1 \rightarrow n := \left(P_{t_\lambda} || P_{t_\lambda}^{\Xi}\right) \otimes \Psi\left(P_{t_\lambda}\right)\right] + b_t\right) \quad \text{Equation (5)}$$

Where i_t is the Input gate activation function, w_i and b_t are the weight and bias of input gate

LYP Candidate Cell State equation:

$$\tilde{C}_t = tanh\left(w_c \cdot \left[h_{t-1}, \forall\lambda = 1 \rightarrow n := \left(P_{t_\lambda} || P_{t_\lambda}^{\Xi}\right) \otimes \Psi\left(P_{t_\lambda}\right)\right] + b_C\right) \quad \text{Equation (6)}$$

where $\tilde{C}_t$ is the cell state activation function, w_c and b_c are the weight and bias of candidate cell state function

LYP Cell State Update equation:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad \text{Equation (7)}$$

where C_t refers the updated cell state

LYP Output Gate equation:

$$o_t = \sigma\left(w_o \cdot \left[h_{t-1}, \forall\lambda = 1 \rightarrow n := \left(P_{t_\lambda} || P_{t_\lambda}^{\Xi}\right) \otimes \Psi\left(P_{t_\lambda}\right)\right] + b_o\right) \quad \text{Equation (8)}$$

Where o_t is the Input gate activation function, w_o and b_o are the weight and bias of output gate

LYP Hidden State Update equation:

$$h_t = o_t \odot \tanh\left(C_t\right) \quad \text{Equation (9)}$$

where h_t is the final output of LYP-LSTM cell

The LYP module processes both normal input parameters and Monte Carlo simulated variables to enhance yield prediction accuracy. By integrating uncertainty-aware simulations, the LYP module dynamically updates its memory state, allowing it to learn complex patterns in agricultural data. The combination of deterministic and probabilistic inputs ensures that the model accounts for variations in environmental conditions, making predictions more robust. The LYP module leverages these diverse inputs to refine its internal representations and improve decision-making. An illustration of proposed LYP Architecture is provided in Figure 2.

This approach significantly enhances the reliability of yield forecasts by incorporating real-world uncertainties. Ultimately, the LYP module provides a more adaptive and data-driven framework for agricultural yield prediction.

Experimental Setup

The experimental setup for the LSPYP-ML framework was implemented on a system with an Intel Core i7 (3.8 GHz) processor, 16GB RAM, and a 1TB NVMe M.2 SSD. Development was conducted in Visual Studio IDE [19] using C++ 23.0 [20] with Microsoft Foundation Classes (MFC) [21] for the dedicated UI and Advanced C 24.0 Library for optimized computations, efficient memory management and parallel processing capabilities to handle large-scale agricultural datasets seamlessly.. The FDDP module utilized

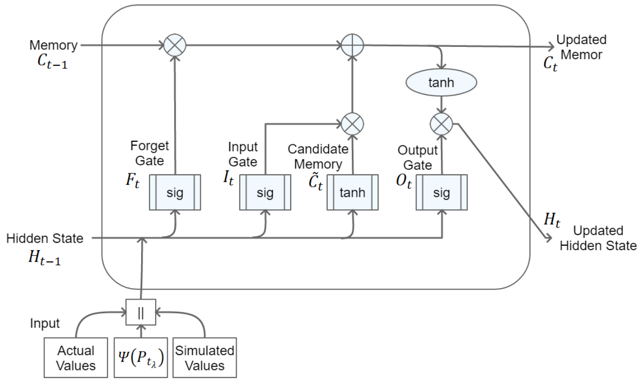

Figure 2: LYP LSTM Architecture

fuzzy logic to classify environmental variables (rainfall, temperature, pesticide usage) and handle missing or noisy data. MCSEV employed probabilistic modeling to generate synthetic datasets, simulating extreme weather scenarios. The LYP module leveraged LSTM networks to capture temporal dependencies in climate and yield data, enhancing location-specific forecasting. The UI is designed in a way to prepare working folders, load dataset, extract dataset, load different methodologies one-by-one in sequence, log the benchmark parameters at different timestamps during the execution, generate report and graphs.

Results and Analysis

The LSPYP-ML framework consistently outperformed existing methods (HQDLM, DLMSDD, DLFCC, HFSAOML) across all evaluation metrics—Accuracy, Precision, Sensitivity, Specificity, and F-Score. It achieved an average accuracy of about 97.07%, precision of 96.04%, sensitivity of 98.19%, specificity of 96.22%, and F-Score of 97.14%, ranking first in every category. Performance fluctuations across timestamps were minimal, demonstrating stability and robustness. These results confirm that integrating fuzzy preprocessing, Monte Carlo simulations, and LSTM

Table 3: Accuracy

Accuracy (%)					
Timestamp	*HQDLM*	*DLMSDD*	*DLFCC*	*HFSAOML*	*LSPYPML*
1	95.30437	80.31421	81.29726	77.521339	97.335625
2	95.15336	80.13043	80.85357	77.2724	97.410736
3	95.15365	79.83714	81.40273	77.92601	97.175606
4	95.6113	80.30492	81.34583	77.854393	96.872795
5	94.9491	80.23478	81.21953	77.626511	97.033646
6	95.13113	80.1753	81.17243	77.527046	97.09581
7	95.03204	80.41753	81.57651	77.401398	96.826363
8	94.93655	80.27249	81.70128	77.409554	97.253227
9	95.2278	80.0099	81.09645	77.571686	96.923164
10	95.53943	80.53426	81.02496	77.631439	96.752747

Table 4: Precision

Precision (%)					
Timestamp	*HQDLM*	*DLMSDD*	*DLFCC*	*HFSAOML*	*LSPYPML*
1	94.63306	82.44935	79.47427	77.901299	96.028694
2	94.47423	82.23538	79.02873	77.853188	96.423721
3	94.28311	81.78342	79.59053	78.144264	96.274254
4	94.62369	82.62614	79.77153	78.632759	95.695915
5	93.72188	82.22311	79.17042	77.886955	95.659805
6	94.56387	82.4773	78.72392	78.116646	95.890755
7	93.93199	82.37688	79.59149	77.586342	96.126762
8	93.8521	82.08076	79.57638	78.248611	96.198143
9	93.93322	82.2313	79.35983	77.903648	96.086571
10	94.75791	82.55996	78.51454	78.08007	95.758224

Table 5: Sensitivity

Sensitivity (%)					
Timestamp	*HQDLM*	*DLMSDD*	*DLFCC*	*HFSAOML*	*LSPYPML*
1	95.92091	79.07272	82.48154	77.313774	98.606125
2	95.7751	78.91322	82.02229	76.959244	98.365494
3	95.95374	78.71924	82.58369	77.804642	98.041656
4	96.53037	78.96044	82.36486	77.427429	98.002663
5	96.08012	79.07843	82.55366	77.483353	98.362495
6	95.64902	78.84718	82.77757	77.206238	98.258911
7	96.04507	79.27051	82.88194	77.300423	97.490852
8	95.93278	79.21589	83.10832	76.957176	98.271835
9	96.42995	78.73333	82.21538	77.389839	97.721634
10	96.26254	79.34537	82.66502	77.385712	97.701553

Table 6: Specificity

Specificity (%)					
Timestamp	*HQDLM*	*DLMSDD*	*DLFCC*	*HFSAOML*	*LSPYPML*
1	94.70417	81.66646	80.19631	77.732079	96.129852
2	94.54828	81.45464	79.76716	77.592918	96.49295
3	94.38094	81.04561	80.30438	78.048447	96.340225
4	94.72783	81.7803	80.389	78.294868	95.794891
5	93.87228	81.48691	79.99046	77.771172	95.775871
6	94.62486	81.63162	79.71717	77.855522	95.987465
7	94.06262	81.65811	80.37078	77.503136	96.180206
8	93.98261	81.40839	80.40896	77.877365	96.276711
9	94.08633	81.40517	80.05265	77.755966	96.15097
10	94.83859	81.82356	79.54172	77.881607	95.84095

Table 7: F-Score

F-Score (%)					
Timestamp	*HQDLM*	*DLMSDD*	*DLFCC*	*HFSAOML*	*LSPYPML*
1	95.27264	80.72575	80.94998	77.60643	97.300354
2	95.12022	80.54005	80.49769	77.403641	97.384926
3	95.11109	80.22208	81.05949	77.974091	97.149918
4	95.56753	80.75171	81.04746	78.025444	96.835556
5	94.88635	80.62012	80.82665	77.684624	96.992325
6	95.10335	80.6214	80.69988	77.658775	97.060394
7	94.97678	80.79384	81.20339	77.443123	96.804008
8	94.88104	80.62289	81.304	77.597519	97.22393
9	95.1652	80.44431	80.76237	77.645889	96.897209
10	95.5043	80.92075	80.53634	77.731339	96.720131

significantly improves yield prediction reliability under uncertain and variable. The measured results during the evaluation are provided in the following tables.

Conclusion

This study introduced the LSPYP-ML framework, which combines fuzzy preprocessing (FDDP), Monte Carlo simulations (MCSEV), and LSTM forecasting (LYP) to improve paddy yield prediction. By addressing uncertainty, simulating environmental variability, and capturing temporal patterns, the model achieved superior performance across all evaluation metrics, proving robust and reliable for location-specific forecasting. The system supports real-time decision-making, offering farmers and policymakers actionable insights to boost productivity and manage climate risks. Future work may extend the framework by incorporating additional environmental factors, optimizing deep learning architectures, and integrating satellite or IoT-based data for broader agricultural applications.

Code Availability

Code and datasets are provided online and the links will be provided through E-Mail requests to the authors

References

Lu, J., Fu, H., Tang, X. et al. GOA-optimized deep learning for soybean yield estimation using multi-source remote sensing data. Sci Rep 14, 7097 (2024). https://doi.org/10.1038/s41598-024-57278-6

Van Touch, Daniel K.Y. Tan, Brian R. Cook, De Li Liu, Rebecca Cross, Thong Anh Tran, Ariane Utomo, Sophea Yous, Clemens Grunbuhel, Annette Cowie, Smallholder farmers' challenges and opportunities: Implications for agricultural production, environment and food security, Journal of Environmental Management, Volume 370, 2024, 122536, ISSN 0301-4797, https://doi.org/10.1016/j.jenvman.2024.122536

Onyeaka, H., Nwauzoma, U. M., Akinsemolu, A. A., Tamasiga, P., Duan, K., Al-Sharify, Z. T., &Siyanbola, K. F. (2024). The ripple effects of climate change on agricultural sustainability and food security in Africa. Food and Energy Security, 13(5), e567. doi: https://doi.org/10.1002/fes3.567

Meghraoui K, Sebari I, Pilz J, Ait El Kadi K, Bensiali S. Applied Deep Learning-Based Crop Yield Prediction: A Systematic Analysis of Current Developments and Potential Challenges. Technologies. 2024; 12(4):43. https://doi.org/10.3390/technologies12040043

Iqbal N, Shahzad MU, Sherif E-SM, Tariq MU, Rashid J, Le T-V, Ghani A. Analysis of Wheat-Yield Prediction Using Machine Learning Models under Climate Change Scenarios. Sustainability. 2024; 16(16):6976. https://doi.org/10.3390/su16166976

Wang Y, Zhang Q, Yu F, Zhang N, Zhang X, Li Y, Wang M, Zhang J. Progress in Research on Deep Learning-Based Crop Yield Prediction. Agronomy. 2024; 14(10):2264. https://doi.org/10.3390/agronomy14102264

Shams, M.Y., Gamel, S.A. & Talaat, F.M. Enhancing crop recommendation systems with explainable artificial intelligence: a study on agricultural decision-making. Neural Comput&Applic 36, 5695–5714 (2024). https://doi.org/10.1007/s00521-023-09391-2

Deborah V. Gaso, Laura Elena Cue La Rosa, Laila A. Puntel, Juan I. Rattalino Edreira, Allard de Wit, Lammert Kooistra, Predicting spatial and temporal variability in soybean yield using deep learning and open source data, European Journal of Agronomy, Volume 164, 2025, 127498, ISSN 1161-0301, https://doi.org/10.1016/j.eja.2024.127498

Tomoaki Yamaguchi, Taiga Takamura, Takashi S.T. Tanaka, Taiichiro Ookawa, Keisuke Katsura, A study on optimal input images for rice yield prediction models using CNN with UAV imagery and its reasoning using explainable AI, European Journal of Agronomy, Volume 164, 2025, 127512, ISSN 1161-0301, https://doi.org/10.1016/j.eja.2025.127512

Praveen Sankarasubramanian, Enhancing precision in agriculture: A smart predictive model for optimal sensor selection through IoT integration, Smart Agricultural Technology, Volume 10,,2025, 100749, ISSN 2772-3755, https://doi.org/10.1016/j.atech.2024.100749

Setiadi DRIM, Susanto A, Nugroho K, Muslikh AR, Ojugo AA, Gan H-S. Rice Yield Forecasting Using Hybrid Quantum Deep Learning Model. Computers. 2024; 13(8):191. https://doi.org/10.3390/computers13080191

Lu J, Li J, Fu H, Tang X, Liu Z, Chen H, Sun Y, Ning X. Deep Learning for Multi-Source Data-Driven Crop Yield Prediction in Northeast China. Agriculture. 2024; 14(6):794. https://doi.org/10.3390/agriculture14060794

Haydar Demirhan, "A deep learning framework for prediction of crop yield in Australia under the impact of climate change," in Information Processing in Agriculture, 2024, ISSN 2214-3173, https://doi.org/10.1016/j.inpa.2024.04.004

Abdel-salam, M., Kumar, N. & Mahajan, S. A proposed framework for crop yield prediction using hybrid feature selection approach and optimized machine learning. Neural Comput&Applic 36, 20723–20750 (2024). https://doi.org/10.1007/s00521-024-10226-x

Bin L, Shahzad M, Khan H, Bashir MM, Ullah A, Siddique M. Sustainable Smart Agriculture Farming for Cotton Crop: A Fuzzy Logic Rule Based Methodology. Sustainability. 2023; 15(18):13874. https://doi.org/10.3390/su151813874

Yin L, Wang L, Li T, Lu S, Tian J, Yin Z, Li X, Zheng W. U-Net-LSTM: Time Series-Enhanced Lake Boundary Prediction Model. Land. 2023; 12(10):1859. https://doi.org/10.3390/land12101859

Park S-H, Lee B-Y, Kim M-J, Sang W, Seo MC, Baek J-K, Yang JE, Mo C. Development of a Soil Moisture Prediction Model Based on Recurrent Neural Network Long Short-Term Memory (RNN-LSTM) in Soybean Cultivation. Sensors. 2023; 23(4):1976. https://doi.org/10.3390/s23041976

Raphael Santana Almeida, Flávio Vasconcelos da Silva, Sávio S.V. Vianna, Combining the bow-tie method and fuzzy logic using Mamdani inference model, Process Safety and Environmental Protection, Volume 169, 2023, Pages 159-168, ISSN 0957-5820, https://doi.org/10.1016/j.psep.2022.11.005

https://visualstudio.microsoft.com/

https://en.cppreference.com/w/cpp/23

https://learn.microsoft.com/en-us/cpp/mfc/mfc-concepts?view=msvc-170