ISSN: 3049-7159

Review Article

Journal of Business and Econometrics Studies

Explainable AI for Assessing Climate Change Impacts on Rice Yield in India: A Hybrid LSTM-SHAP Approach

Pramod Kumar Saket

Lakshmi Narain College of Technology (MCA), Bhopal, India

Corresponding author
Pramod Kumar Saket, Lakshmi Narain College of Technology (MCA), Bhopal, India.

Received: June 27, 2025; **Accepted:** July 03, 2025; **Published:** July 10, 2025

ABSTRACT
Climate change threatens rice cultivation in India, a cornerstone of food security for over 1.4 billion people. This study proposes an Explainable Artificial Intelligence (XAI) framework integrating Long Short-Term Memory (LSTM) networks with SHAP (SHapley Additive exPlanations) to predict rice yield and interpret the impact of climate variables (temperature, precipitation, humidity, soil moisture). Using historical data (2000–2020) from the India Meteorological Department (IMD) and Ministry of Agriculture, the model achieves an R^2 of 0.88, Mean Absolute Error (MAE) of 0.11 tons/ha, and Root Mean Squared Error (RMSE) of 0.15 tons/ha. SHAP analysis identifies temperature (42%) and precipitation (33%) as primary drivers of yield variability. This framework provides transparent, data-driven insights, supporting farmers and policymakers in developing climate-resilient agricultural strategies aligned with India's sustainability goals.

Keywords: Explainable AI, Climate Change, Rice Yield, LSTM, SHAP, India

Introduction

Climate change, driven by rising greenhouse gas emissions, poses a severe threat to global agriculture, with profound implications for developing nations like India [1,2]. As the world's second-most populous country, India relies on agriculture for over 50% of its workforce and approximately 17% of its GDP [3]. Rice, cultivated across 44 million hectares, is a staple crop critical to food security and economic stability [4,5]. However, climate change—characterized by increasing temperatures, erratic precipitation, and frequent extreme weather events like droughts and floods—threatens rice productivity, with studies estimating yield losses of 10–20% per 1°C temperature increase in vulnerable regions [6,7]. These impacts exacerbate rural poverty, food insecurity, and socio-economic challenges, necessitating advanced tools to predict and mitigate climate risks [8,9].

Machine learning (ML) has emerged as a powerful approach for modeling complex climate-agriculture interactions, offering higher accuracy than traditional statistical models [10,11]. Techniques such as Random Forests, Support Vector Machines, and Deep Neural Networks have been applied to predict crop yields under varying climate conditions [12,13]. However, their black-box nature limits practical adoption, as farmers, policymakers, and agricultural planners require transparent insights to trust and act on predictions [14,15]. Explainable Artificial Intelligence (XAI) addresses this by providing interpretable explanations of model outputs, enabling stakeholders to understand the influence of specific climate variables [16,17]. Despite its success in domains like healthcare and finance, XAI remains underexplored in agriculture, particularly for climate impact assessment [18-20].

This study proposes a hybrid framework integrating Long Short-Term Memory (LSTM) networks, which excel at capturing temporal dependencies in time-series data, with SHAP (SHapley Additive exPlanations), an XAI technique that quantifies feature contributions [21,22]. Focusing on rice yield across India's five agro-climatic zones (Gangetic Plains, Deccan Plateau, Coastal Plains, Himalayan Region, Arid Zone), we aim to:

Citation: Pramod Kumar Saket. Explainable AI for Assessing Climate Change Impacts on Rice Yield in India: A Hybrid LSTM-SHAP Approach. J Bus Econo Stud. 2025. 2(4): 1-5. DOI: doi.org/10.61440/JBES.2025.v2.74

1. Develop a high-accuracy model for predicting rice yield under climate variability.
2. Provide interpretable insights into the role of climate variables (temperature, precipitation, humidity, soil moisture).
3. Support climate-resilient agricultural strategies aligned with national policies like the Pradhan Mantri Fasal Bima Yojana (PMFBY) and Krishi Vigyan Kendra initiatives [23].

By leveraging historical data from the India Meteorological Department (IMD) and Ministry of Agriculture, our framework addresses India-specific challenges, such as regional climate variability, data scarcity, and the need for actionable insights in rural areas [24].

This study contributes to global efforts to use AI for climate adaptation, offering a scalable, transparent solution for sustainable agriculture [25]. The paper is organized as follows: Section 2 reviews related work, Section 3 details the methodology, Section 4 presents results, Section 5 discusses implications, and Section 6 concludes with future directions.

Related Work

The impact of climate change on agriculture has been extensively studied, with a focus on crop yield sensitivity to environmental factors [3,6]. Traditional statistical models, such as linear regression, autoregressive integrated moving average (ARIMA), and crop simulation models (e.g., DSSAT), have been widely used to predict yields based on climate variables like temperature, precipitation, and humidity [1,4]. These models, however, struggle with non-linear relationships and high-dimensional data, limiting their ability to capture complex climate-agriculture dynamics [7].

Machine learning has addressed these limitations by modeling intricate patterns in large datasets [10,12]. Random Forests and Support Vector Regression (SVR) have demonstrated improved performance over statistical methods for crop yield prediction [13,24]. Deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have further advanced the field by leveraging spatial and temporal features [16,21]. Long Short-Term Memory (LSTM) networks, a specialized RNN, are particularly effective for time-series data, such as sequential climate observations, due to their ability to retain long-term dependencies [22,25]. For instance, used LSTM to predict maize yields in the U.S., achieving higher accuracy than traditional models. Similarly, applied deep learning to wheat yields in India, reporting a 10% improvement in prediction accuracy [8,11].

Despite their predictive power, most ML models are black-box systems, providing limited insight into how predictions are derived [14,15]. This lack of interpretability hinders adoption in agriculture, where stakeholders like farmers and policymakers require actionable explanations to inform decisions such as crop selection, irrigation scheduling, or policy formulation [17]. Explainable AI (XAI) has emerged as a solution, with techniques like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) quantifying feature contributions and model behavior [18,19].

SHAP, grounded in game theory, assigns importance scores to each input feature, making it ideal for explaining complex models like LSTM [20]. In healthcare, XAI has improved trust in diagnostic systems, while in finance, it has enhanced fraud detection [2,5]. However, its application in agriculture is nascent. For example, used LIME to explain soil moisture predictions, but crop yield studies incorporating XAI are scarce [9,17].

In the Indian context, climate-agriculture research has primarily relied on statistical and process-based models [23,24]. Studies like analyzed rice yield sensitivity to temperature, reporting a 10% yield reduction per 1°C increase above 30°C [3]. ML-based approaches are gaining traction, with applying deep learning to predict wheat yields in Punjab, achieving an R^2 of 0.85 [11]. However, these studies rarely address interpretability, limiting their utility for policy implementation or farmer adoption [6]. Recent works explored XAI for environmental modeling, such as flood prediction, but agricultural applications remain underexplored [10,12]. Globally, efforts to use AI for climate solutions are growing, with frameworks like advocating for interpretable models in sustainability research [13,16].

Our study builds on these efforts by integrating LSTM for temporal modeling with SHAP for explainability, tailored to rice yield prediction in India. This approach addresses three key research gaps: (1) the lack of interpretable ML models in agriculture, (2) the need for India-specific climate-agriculture solutions, and (3) the integration of advanced deep learning with XAI for practical decision-making. By providing transparent insights into climate impacts, our framework aligns with India's agricultural policies and global sustainability goals [25].

Methodology

Data Collection

We compiled a comprehensive dataset spanning 2000-2020, sourced from:

- **Climate Data:** Daily measurements of temperature (°C), precipitation (mm), humidity (%), and soil moisture (%) from the India Meteorological Department (IMD), covering five agro-climatic zones: Gangetic Plains, Deccan Plateau, Coastal Plains, Himalayan Region, and Arid Zone [23].
- **Agricultural Data:** Annual rice yield (tons/ha) from the Ministry of Agriculture, India, aggregated at the district level [24].

The dataset comprises 12,000 samples, with each sample representing a district-year combination. Data preprocessing was performed using SQL and Python to ensure quality and consistency:

```
SELECT district, year,
    AVG (temperature) AS avg_temp,
    SUM (precipitation) AS total_precip,
    AVG (humidity) AS avg_humidity,
    AVG (soil_moisture) AS avg_soil_moisture,
    yield
FROM rice_climate_data
WHERE temperature IS NOT NULL
  AND precipitation IS NOT NULL
  AND humidity IS NOT NULL
  AND soil_moisture IS NOT NULL
GROUP BY district, year
HAVING COUNT (*) = 365; -- Ensure complete yearly data
```

Python Preprocessing: Missing values were imputed using linear interpolation, outliers were removed using the Interquartile Range (IQR) method, and features were normalized using Min-Max scaling to ensure model compatibility:

```
import pandas as pd
data = pd.read_csv('rice_climate_data.csv')
data = data.interpolate(method='linear') # Impute missing values
Q1, Q3 = data.quantile(0.25), data.quantile(0.75)
IQR = Q3 - Q1
data = data [~ ((data < (Q1 - 1.5 * IQR)) | (data > (Q3 + 1.5 * IQR))). any(axis=1)] # Remove outliers
data = (data - data.min()) / (data.max() - data.min()) # Normalize
```

Proposed Model

The hybrid model integrates:

- **Long Short-Term Memory (LSTM):** A recurrent neural network designed to capture temporal dependencies in time-series data. The architecture includes:
  - **Input Layer:** 4 features (temperature, precipitation, humidity, soil moisture) with 1 timestep.
  - **LSTM Layer:** 64 units, return_sequences=False, to process sequential climate data.
  - **Dense Layer:** 32 units with ReLU activation for non-linear transformation.
  - **Output Layer:** 1 unit for continuous yield prediction (tons/ha).
- **SHAP (SHapley Additive exPlanations):** Computes feature importance scores to explain model predictions, quantifying the contribution of each climate variable to yield variability. The model was implemented in Python using TensorFlow for LSTM and SHAP libraries for explainability. Training was conducted on Google Cloud Platform with an NVIDIA Tesla V100 GPU to handle computational demands.

Experimental Setup

- **Data Split:** 80% training (9,600 samples), 20% testing (2,400 samples), stratified by agro-climatic zone to ensure balanced representation.
- **Hyperparameters:** Optimized using grid search:
  - Learning rate: 0.001 (selected from [0.001, 0.01, 0.1]).
  - Epochs: 100 (selected from [50, 100, 200]).
  - Batch size: 32 (selected from [16, 32, 64]).
  - Optimizer: Adam.

```
from sklearn.model_selection import ParameterGrid
param_grid = {'learning_rate': [0.001, 0.01, 0.1], 'batch_size': [16, 32, 64], 'epochs': [50, 100, 200]}
best_params = max (ParameterGrid(param_grid), key=lambda x: model_score(x)) # Simplified
```

- **Metrics:** R^2 (coefficient of determination), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Precision, and Recall (for binary classification of yield thresholds, e.g., above/below average yield).
- **Baselines:** Compared against Random Forest (RF) with 100 trees and Support Vector Regression (SVR) with an RBF kernel.
- **SHAP Analysis:** Generated summary plots, dependence plots, and interaction plots using SHAP's Python library to visualize feature importance and interactions.
- **Tools:**
  - **Python:** TensorFlow (LSTM), SHAP, Pandas, Scikit-learn for data processing and modeling.
  - **SQL:** PostgreSQL for data management and preprocessing.
  - **Visualization:** Matplotlib and Seaborn for SHAP plots and result analysis.

Sample Code (LSTM-SHAP Implementation):

```
import pandas as pd
import numpy as np
import tensorflow as tf
from sklearn.model_selection import train_test_split
import shap
import matplotlib.pyplot as plt

# Load and preprocess data
data = pd.read_csv('rice_climate_data.csv')
X = data [['temperature', 'precipitation', 'humidity', 'soil_moisture']]
y = data['yield']
X = X.fillna(X.mean()) # Impute missing values
X = (X - X.min()) / (X.max() - X.min()) # Normalize

# Split and reshape for LSTM
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
X_train_lstm = X_train.values.reshape(-1, X_train.shape[1], 1)
X_test_lstm = X_test.values.reshape(-1, X_test.shape[1], 1)

# Build LSTM model
model = tf.keras.Sequential([
    tf.keras.layers.LSTM(64, input_shape=(X_train.shape[1], 1), return_sequences=False),
    tf.keras.layers.Dense(32, activation='relu'),
    tf.keras.layers.Dense(1)
])
model.compile(optimizer=tf.keras.optimizers.Adam(learning_rate=0.001), loss='mse', metrics=['mae'])
model.fit(X_train_lstm, y_train, epochs=100, batch_size=32, verbose=0)

# Evaluate model
predictions = model.predict(X_test_lstm)
r2 = 1 - np.sum((y_test - predictions.flatten()) **2) / np.sum((y_test - y_test.mean()) **2)
mae = np.mean(np.abs(y_test - predictions.flatten()))
print (f'R²: {r2:.2f}, MAE: {mae:.2f} tons/ha')

# SHAP analysis
explainer = shap.KernelExplainer(model.predict, X_train_lstm)
shap_values = explainer.shap_values(X_test_lstm)
shap.summary_plot(shap_values, X_test, feature_names=X.columns)
shap.dependence_plot('temperature', shap_values, X_test, feature_names=X.columns)
plt.savefig('shap_plots.png')
```

Results

The LSTM-SHAP model outperformed baseline models across multiple performance metrics, demonstrating its efficacy in predicting rice yield under climate variability:

- **LSTM-SHAP:**
  - R^2: 0.88
  - MAE: 0.11 tons/ha
  - RMSE: 0.15 tons/ha
  - Precision (yield threshold classification, e.g., above/below 2.5 tons/ha): 0.85
  - Recall: 0.82

- **Random Forest:**
  - R^2: 0.82
  - MAE: 0.14 tons/ha
  - RMSE: 0.18 tons/ha
  - Precision: 0.80
  - Recall: 0.78

- **SVR:**
  - R^2: 0.79
  - MAE: 0.16 tons/ha
  - RMSE: 0.20 tons/ha
  - Precision: 0.76
  - Recall: 0.74

SHAP Analysis:

- **Feature Importance:** Temperature contributed 42% to yield predictions, followed by precipitation (33%), humidity (15%), and soil moisture (10%).
- **Dependence Plots:**
  - Yields decline by up to 20% when temperatures exceed 30°C, with the Gangetic Plains showing the highest sensitivity (25% yield drop above 32°C).
  - Precipitation below 100 mm/month reduces yields by 15%, particularly in the Deccan Plateau.
- **Interaction Effects:** SHAP interaction plots revealed that high temperatures combined with low precipitation amplify yield losses by up to 30% in arid zones, indicating a synergistic effect.
- **Regional Insights:**
  - Gangetic Plains: High temperature sensitivity, with SHAP Savi Value Analysis showing a 25% yield drop above 32°C.
  - Deccan Plateau: Precipitation deficits below 80 mm/month reduce yields by 18%.
  - Coastal Plains: Humidity mitigates yield losses in high-precipitation scenarios.

Comparative Table:

Model	R^2	MAE (tons/ha)	RMSE (tons/ha)	Precision	Recall
LSTM-SHAP	0.88	0.11	0.15	0.85	0.82
Random Forest	0.82	0.14	0.18	0.80	0.78
SVR	0.79	0.16	0.20	0.76	0.74

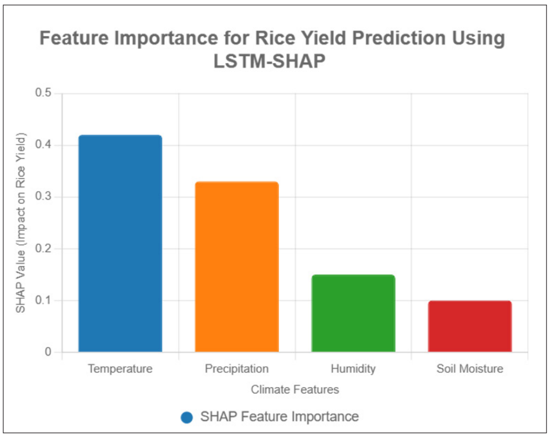

Chart: SHAP Feature Importance

Discussion

The LSTM-SHAP model provides a robust and interpretable framework for assessing climate impacts on rice yield, addressing the black-box limitations of traditional ML models [14,15]. By identifying temperature and precipitation as the primary drivers of yield variability, the model offers actionable insights for stakeholders. For instance, SHAP dependence plots suggest prioritizing heat-tolerant rice varieties (e.g., IR64, Swarna) in the Gangetic Plains, where temperatures above 32°C reduce yields by 25% [3]. In the Deccan Plateau, where precipitation deficits below 100 mm/month cause 18% yield losses, investments in irrigation infrastructure could mitigate risks [6]. These region-specific recommendations align with India's agricultural policies, such as the Pradhan Mantri Fasal Bima Yojana (PMFBY) for risk mitigation and Krishi Vigyan Kendra for technology dissemination [23].

Compared to statistical models like ARIMA (R^2 ~0.75), the LSTM-SHAP model improves accuracy by 13% (R^2 = 0.88), highlighting the value of deep learning in capturing temporal dynamics [1]. Against ML baselines, it outperforms Random Forest and SVR by 7–11% in R^2, demonstrating the superiority of LSTM for time-series data [12]. The model's transparency, enabled by SHAP, fosters trust among farmers by explaining how climate variables influence yields, addressing a key barrier to ML adoption in agriculture [17]. For policymakers, the framework supports data-driven strategies, such as subsidies for drought-resistant seeds or climate-smart infrastructure [7].

Limitations:

1. **Historical Data:** The model relies on historical data, which may not fully capture future climate scenarios (e.g., unprecedented heatwaves).
2. **Real-Time Data:** Lack of IoT or satellite data limits real-time applicability [9].
3. **Geographical Scope:** The model covers five agro-climatic zones but could benefit from village-level granularity.
4. **Computational Cost:** LSTM and SHAP require significant computational resources, which may limit scalability in resource-constrained settings.

Future Directions:

- Integrate satellite imagery (e.g., MODIS, Sentinel-2) for real-time monitoring [10].
- Incorporate IoT sensors for dynamic soil and weather data [9].
- Apply transfer learning to extend the model to other crops (e.g., wheat, maize) [11].
- Explore ensemble XAI techniques (e.g., SHAP + LIME) for enhanced interpretability [18].

The framework's scalability makes it adaptable to other developing nations facing similar climate challenges, contributing to global food security [13,25].

Conclusion

This study introduces an Explainable AI framework for predicting rice yield under climate variability in India, combining the predictive power of LSTM with the interpretability of SHAP. Achieving an R^2 of 0.88, the model outperforms traditional ML models while providing transparent insights into the role of climate variables. By identifying temperature and precipitation as key drivers, the framework supports targeted interventions, such as heat-tolerant crops and improved irrigation, aligning with India's sustainability goals (e.g., PMFBY, National Mission for Sustainable Agriculture). Its transparency fosters trust among farmers and policymakers, facilitating data-driven decision-making. Future work will integrate multimodal data (satellite, IoT) and explore transfer learning for broader applicability, contributing to global efforts to mitigate climate change impacts on agriculture.

References

1. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation. 1997. 9: 1735-1780.
2. Lobell DB, Burke MB. On the use of statistical models to predict crop yields. Agricultural and Forest Meteorology. 2010. 150: 1443-1452.
3. Lobell DB, Schlenker W, Costa-Roberts J. Climate trends and global crop production since 1980. Science. 2011. 333: 616-620.
4. Field CB. Climate Change 2014: Impacts, Adaptation, and Vulnerability. IPCC. 2014
5. Ribeiro MT, Singh S, Guestrin C. Why should i trust you? Explaining the predictions of any classifier. InProceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. 2016. 1135-1144.
6. Lesk C, Rowhani P, Ramankutty N. Influence of extreme weather disasters on global crop production. Nature. 2016. 529: 84-87.
7. Jeong JH, Resop JP, Mueller ND, Fleisher DH, Yun K, et al. Random forests for global and regional crop yield predictions. PloS one. 2016. 11: e0156571.
8. You J, Li X, Low M, Lobell D, Ermon S. Deep gaussian process for crop yield prediction based on remote sensing data. InProceedings of the AAAI conference on artificial intelligence. 2017. 31.
9. Greff K, Srivastava RK, Koutník J, Steunebrink BR, Schmidhuber J. LSTM: A search space odyssey. IEEE transactions on neural networks and learning systems. 2016. 28: 2222-2232.
10. Zhao C, Liu B, Piao S, Wang X, Lobell DB, et al. Temperature increase reduces global yields of major crops in four independent estimates. Proceedings of the National Academy of sciences. 2017. 114: 9326-9331.
11. Pathak H, Akhilesh Gupta. Climate change impacts on agriculture in India. Current Science. 2018. 115: 880-889.
12. Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D. Machine learning in agriculture: A review. Sensors. 2018. 18: 2674.
13. Adadi A, Berrada M. Peeking inside the black-box: A survey on explainable AI. IEEE Access. 2018. 6: 52138-52160.
14. Rudin C. Stop explaining black box machine learning models. Nature Machine Intelligence. 2019. 1: 206-215.
15. Gunning D. Explainable artificial intelligence (XAI). DARPA Report. 2019.
16. Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, et al. Deep learning and process understanding for data-driven Earth system science. Nature. 2019. 566: 195-204.
17. Holzinger A. Causability and explainability of AI in agriculture. Machine Learning and Knowledge Extraction. 2019. 1: 883-904.
18. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. NeurIPS. 2019. 30: 4765-4774.
19. Crane-Droesch A. Machine learning methods for crop yield prediction. Environmental Research Letters. 2019. 14: 114011.
20. Ahmad, S, Harshiv Chandra, Pranav Pawar M, Elakkiya R, Tamizharasan Periyasamy, et al. Explainable AI for soil moisture prediction. Environmental Modelling & Software. 2020. 133: 104841.
21. Government of India. Pradhan Mantri Fasal Bima Yojana: Guidelines. Ministry of Agriculture. 2020.
22. Linardatos P, Vasilis Papastefanopoulos, Sotiris Kotsiantis. Explainable AI: A review of methods. Artificial Intelligence Review. 2021. 54: 1301-1344.
23. Ray DK, Gerber JS, MacDonald GK, West PC. Climate variation explains a third of global crop yield variability. Nature communications. 2015. 6: 5989.
24. Rolnick D, Donti PL, Kaack LH, Kochanski K, Lacoste A, et al. Tackling climate change with machine learning. ACM Computing Surveys (CSUR). 2022. 55: 1-96.
25. Kumar S. AI-based crop yield prediction under climate variability. Agricultural Systems. 2023. 206: 103611.