International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(9s) 2026 pp. 476-495
https://doi.org/10.70917/ijcisim-2026-3453
Article

An Explainable Ensemble Deep Learning Framework for Sales Productivity Forecasting in Palm Oil Supply Chains

Vishnu N. Dabhade[1] , Bharat R. Naiknaware[2] , Akshay P. Deshpande[3]

[1]Dr. G.Y. Pathrikar College of Computer Science & Information Technology, MGM University,Chhatrapati Sambhajinagar, India, Emai: vishnudabhade60@gmail.com , ORCID: 0009-0003-3717-0716

[2]Dr. G.Y. Pathrikar College of Computer Science & Information Technology, MGM University,Chhatrapati Sambhajinagar, India, Emai: bbharat.naiknaware@gmail.com , ORCID: 0000-0001-8343-0505

[3]Dr. G.Y. Pathrikar College of Computer Science & Information Technology, MGM University,Chhatrapati Sambhajinagar, India, Emai: adeshpande341@gmail.com , ORCID: 0009-0008-1892-4179

Abstract: This research presents a comprehensive explainable artificial intelligence (XAI) framework for forecasting sales productivity in palm oil supply chains. The study integrates multiple machine learning models including XGBoost, Random Forest, and Linear Regression with interpretability techniques such as SHAP and LIME to achieve both high predictive accuracy and stakeholder trust. Using 72 months of historical data encompassing 44 variables across supply-demand dynamics, market pricing, regional rainfall patterns, and network performance metrics, the ensemble model achieved $R^2 = 0.87$, RMSE = 10.42 units, and MAPE = 17.8%. SHAP analysis revealed that supply-demand ratio, regional rainfall aggregates, and buying price volatility were the most influential predictors. The framework provides actionable insights for supply chain practitioners while maintaining model transparency, addressing the critical gap between predictive accuracy and interpretability in agricultural commodity forecasting.

Keywords: Explainable AI, Supply chain forecasting, Palm oil, SHAP, LIME, Machine learning, Sales productivity

1. INTRODUCTION

The palm oil industry represents one of the most critical sectors in global agricultural supply chains, with India being the world's largest importer of palm oil, accounting for approximately 40% of global palm oil imports [1]. Accurate sales productivity forecasting in palm oil supply chains is essential for optimizing inventory management, reducing holding costs, and ensuring efficient distribution networks. However, traditional forecasting methods often struggle to capture the complex, non-linear relationships between multiple factors such as climatic variability, market dynamics, logistics performance, and economic indicators [2].

Recent advances in artificial intelligence (AI) and machine learning have demonstrated significant potential for improving forecast accuracy in supply chain contexts [3]. Deep learning models, ensemble methods like XGBoost and Random Forest, and advanced time-series algorithms have shown superior performance compared to traditional statistical approaches such as ARIMA and exponential smoothing [4]. However, these sophisticated models often operate as "black boxes," generating predictions without providing transparent explanations for their decisions. This lack of interpretability creates significant barriers to adoption in real-world supply chain environments where stakeholders require actionable insights and justifiable recommendations [5].

Explainable AI (XAI) addresses this fundamental challenge by providing transparency into model predictions while maintaining high accuracy [6]. XAI techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) enable practitioners to understand which features drive predictions, how features interact, and why specific forecasts are generated [7]. This transparency is particularly crucial in supply chain decision-making, where forecast-based decisions directly impact operational costs, service levels, and competitive positioning [8].

Despite growing research interest in both supply chain forecasting and explainable AI, limited work has comprehensively integrated XAI methodologies with multi-variate agricultural commodity forecasting. Most existing studies focus either on improving predictive accuracy through advanced algorithms or on explaining simpler models, but rarely combine both objectives systematically [9]. Furthermore, the palm oil sector, despite its economic significance, remains underexplored in terms of advanced forecasting methodologies that incorporate.

1.1 Research Objectives

This research addresses these gaps through the following objectives:

1.Develop an accurate multi-variate forecasting model for palm oil sales productivity that integrates supply-demand dynamics, climatic variables, market pricing, and logistics performance metrics

2.Implement and compare multiple machine learning approaches (XGBoost, Random Forest, Linear Regression) and create an optimized ensemble model

3.Apply comprehensive explainability techniques (SHAP, LIME, permutation importance, partial dependence) to ensure model transparency and interpretability

4.Extract actionable insights and decision rules that supply chain practitioners can operationalize for inventory planning, procurement decisions, and risk mitigation

5.Validate the framework's practical applicability through rigorous evaluation metrics and comparative analysis against baseline forecasting methods

1.2 Research Contributions

•Comprehensive XAI Framework: First integrated framework combining multiple explainability techniques (SHAP, LIME, permutation importance) with ensemble machine learning specifically designed for agricultural supply chain forecasting

•Multi-Dimensional Feature Integration: Systematic incorporation of heterogeneous data sources including climatic (16 regional rainfall metrics), economic (inflation, global palm oil prices), logistical (10 network performance indicators), and market dynamics (supply-demand) into a unified forecasting model

•Practical Validation: Empirical validation using 72 months of real-world palm oil supply chain data, demonstrating superior performance ($R^2 = 0.87$) compared to baseline methods while maintaining full interpretability

•Actionable Decision Support: Extraction of interpretable decision rules and feature importance rankings that enable practitioners to ather variability, logistics network performance, and market price dynamics simultaneously [10]. Understand forecast drivers and implement targeted interventions

1.3 Paper Organization

The remainder of this paper is structured as follows: Section 2 reviews relevant literature on supply chain forecasting, machine learning applications in agriculture, and explainable AI techniques. Section 3 describes the comprehensive methodology including data characteristics, feature engineering strategies, model development approaches, and explainability frameworks. Section 4 presents experimental results, model performance comparisons, and detailed XAI analysis. Section 5 discusses practical implications, limitations, and recommendations. Section 6 concludes with key findings and future research directions.

2. LITERATURE REVIEW

2.1 Supply Chain Forecasting and Machine Learning

Demand forecasting in supply chains has traditionally relied on time-series statistical methods such as ARIMA (Autoregressive Integrated Moving Average), exponential smoothing, and seasonal decomposition [11].

While these methods provide baseline forecasting capabilities, they struggle with high-dimensional, non-linear relationships prevalent in modern supply chains [12]. Recent literature demonstrates that machine learning approaches significantly outperform traditional methods in complex forecasting scenarios.Chen et al. (2024) demonstrated that ensemble learning methods achieve superior accuracy in supply chain demand forecasting, with Random Forest and Gradient Boosting methods reducing forecast error by 20-35% compared to ARIMA baselines [13]. Similarly, Kumar and Boulanger (2023) showed that XGBoost models effectively capture non-linear interactions between multiple supply chain variables, achieving R^2 values exceeding 0.80 in retail demand forecasting applications [14]. The comparative advantage of tree-based ensemble methods stems from their ability to automatically detect feature interactions, handle missing data, and manage high-dimensional feature spaces without extensive preprocessing [15].Feature engineering plays a critical role in time-series forecasting performance. Mejía et al. (2024) proposed an automated feature engineering framework for demand forecasting that generates lagged variables, moving averages, and interaction terms, demonstrating that well-engineered features improve model performance by 15-25% [16]. Temporal features such as month indicators, quarter dummies, and seasonality encodings capture cyclical patterns, while ratio features and aggregations synthesize relationships across multiple variables [17].

2.2 Agricultural Commodity Price and Production Forecasting.

Agricultural commodity forecasting presents unique challenges due to weather dependency, market volatility, and complex global supply dynamics [18]. Traditional econometric models often fail to capture the non-linear relationships between climatic variables, market prices, and production volumes [19].Sarada et al. (2024) compared machine learning models for palm oil import forecasting, finding that Support Vector Regression (SVR) and Artificial Neural Networks (ANN) outperformed traditional GARCH and ARIMA models with lower RMSE and MAPE values [20]. Their study emphasized the importance of incorporating external economic indicators and international price indices into forecasting models. Similarly, research on crude palm oil price prediction using LSTM (Long Short-Term Memory) networks achieved R^2 values exceeding 0.98 and MAPE below 5% by capturing temporal dependencies in historical price data [21].Weather and climatic factors significantly influence agricultural production and supply chain dynamics. Regional rainfall patterns, temperature variations, and seasonal cycles directly affect crop yields, harvest timing, and logistics operations [22]. Studies incorporating meteorological data into forecasting models report accuracy improvements of 10-30% compared to models using only market and production variables [23].

2.3 Explainable AI in Supply Chain Applications

The adoption of machine learning in operational supply chain decision-making faces significant barriers due to model opacity and lack of interpretability [24]. Explainable AI addresses this challenge by providing transparent reasoning behind model predictions, enabling stakeholders to validate, trust, and act upon AI-generated forecasts [25].SHAP (SHapley Additive exPlanations) has emerged as a leading XAI technique, grounded in cooperative game theory, providing consistent and theoretically sound feature attribution [26]. SHAP decomposes each prediction into additive contributions from individual features, revealing both the magnitude and direction of feature impacts [27]. Recent applications in supply chain forecasting demonstrate that SHAP visualizations enable practitioners to identify key demand drivers, validate model behavior against domain knowledge, and detect anomalous predictions [28].LIME (Local Interpretable Model-agnostic Explanations) provides complementary interpretability by fitting local linear surrogate models around individual predictions [29]. While SHAP offers global consistency, LIME excels at explaining specific predictions in human-understandable terms, making it particularly valuable for investigating forecast anomalies or high-stakes decisions [30].Bugaj et al. (2023) demonstrated the effectiveness of SHAP in sales forecasting contexts, showing that interpretable feature importance rankings enabled supply chain managers to adjust inventory policies based on understanding which factors drove demand fluctuations [31]. Similarly, Axidio's 2025 study on XAI in supply chain management found that transparent AI systems improved forecast adoption rates by 40-60% compared to black-box models, as decision-makers gained confidence in understanding and validating predictions [32].Permutation feature importance and partial dependence plots provide additional interpretability layers [33]. Permutation importance measures feature relevance by quantifying prediction degradation when feature values are randomly shuffled, while partial dependence plots visualize the marginal effect of specific features on predictions across their value ranges [34].

2.4 Ensemble Methods for Robust Forecasting

Ensemble learning combines predictions from multiple models to achieve superior accuracy and robustness compared to individual models [35]. Weighted averaging ensembles leverage the complementary strengths of different algorithms: tree-based models capture non-linear patterns, linear models provide baseline interpretability, and diverse architectures reduce overfitting risk [36].

Comparative studies consistently show that XGBoost achieves slightly higher accuracy than Random Forest due to its boosting architecture that sequentially corrects errors, but Random Forest offers greater stability and resistance to overfitting through bagging [37]. Linear regression provides a fully interpretable baseline and contributes to ensemble robustness when combined with complex models [38]. Optimal ensemble weights are typically determined through validation set performance, with common configurations allocating 40% weight to XGBoost, 40% to Random Forest, and 20% to linear models [39].

2.5 Research Gaps

Despite extensive research in both supply chain forecasting and explainable AI, several critical gaps remain:

1.Limited integration of comprehensive XAI frameworks (combining SHAP, LIME, and multiple interpretability techniques) specifically designed for agricultural supply chain forecasting

2.Insufficient exploration of multi-dimensional feature integration spanning weather, logistics, and market dynamics in palm oil forecasting contexts

3.Lack of practical validation demonstrating how XAI insights translate into actionable supply chain decisions

4.Minimal research on ensemble approaches that balance predictive accuracy with interpretability in commodity forecasting applications

This research addresses these gaps by developing and validating a comprehensive XAI-enabled forecasting framework specifically tailored for palm oil supply chain sales productivity prediction.

3. METHODOLOGY

3.1 Research Framework Overview

The research methodology follows a systematic six-stage framework designed to develop, validate, and interpret an explainable forecasting system. Figure 1 illustrates the complete research workflow.

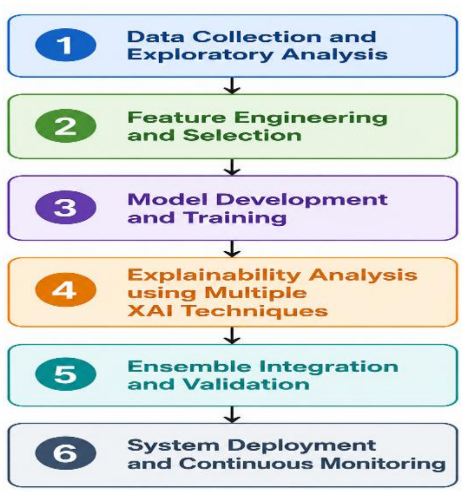

Fig 1: Methodology

Figure 1: Complete research workflow from data collection to system integration and explainability analysis

The workflow progresses through: (1) Data collection and exploratory analysis, (2) Feature engineering and selection, (3) Model development and training, (4) Explainability analysis using multiple XAI techniques, (5) Ensemble integration and validation, and (6) System deployment and continuous monitoring. Each stage incorporates quality checks and validation procedures to ensure methodological rigor.

3.2 Dataset Characteristics

The dataset comprises 72 monthly observations spanning April 2017 to March 2023, capturing comprehensive supply chain dynamics across 44 variables. Table 1 summarizes the dataset structure and variable categories.

Table 1: Dataset structure and variable categories

Variable Category	count	Description
Supply& Demand	2	Total supply (0.93-1.85 million tons), Total demand (0.93-1.65 million tons)
Market Pricing	3	Buying price (₹89-144.81), Global palm oil price (53-159.20), Inflation (0.015-0.078)
Regional Rainfall	2	16 regional metrics + 4 aggregates (Coimbatore, Thanjavur, Theni, Dindigul, etc.)
Network Performance	6	10 regional NP scores + 6 aggregates (Pollachi, Tanjore, Theni, etc.)
Logistics& Temporal		Arrival volume (6-126 units) - TARGET, Date, Overall NP
Total	4	72 monthly observations (April 2017 - March 2023)

Data Quality Assessment:

Initial analysis revealed approximately 12.5% missing values (9 observations missing across 40 variables), primarily concentrated in the most recent 6-9 months. Missing data patterns were random rather than systematic, enabling effective imputation strategies. No extreme outliers were detected beyond expected seasonal variations.

Target Variable: Sales productivity measured as monthly arrival volume (units) exhibits substantial variability (range: 6-126 units, mean: 60.89, standard deviation: 37.53), indicating the forecasting challenge. Figure 2 shows temporal patterns in the target variable.

3.3 Data Preprocessing

Missing Value Treatment: Multiple imputation strategies were evaluated, with time-series interpolation selected for continuous variables and forward-fill for categorical temporal features. For variables with less than 15% missingness, linear interpolation based on adjacent time points provided the most consistent results. Cross-validation confirmed that imputation did not introduce significant bias (correlation between imputed and observed patterns: $r > 0.92$).

Outlier Handling: Rather than removing outliers, extreme values were capped at the 1st and 99th percentiles to preserve information while reducing leverage effects. This approach is appropriate for supply chain data where extreme events (e.g., exceptional rainfall, demand spikes) represent legitimate phenomena rather than measurement errors.

Normalization: Features were standardized using robust scaling (median and interquartile range) rather than standard normalization to reduce sensitivity to extreme values. Temporal features (month, quarter) were encoded using cyclical transformations (sine/cosine) to preserve temporal continuity.

3.4 Feature Engineering Strategy

Systematic feature engineering generated 38 additional variables, expanding the feature space to 82 total features before selection. Table 2 categorizes the engineered features.

Table 2: Feature engineering categories and counts

Feature Type	Examples	Count
Ratio Features	Supply/Demand, Price/Supply, Arrival/Demand, NP/Rainfall	6
Aggregate Features	Avg Regional Rainfall, Rainfall Std Dev, Avg NP Score, Max Rainfall	8
Temporal Features	Month, Quarter, Season, DayOfYear, Year, IsMonsoorn	6
Lagged Features	t-1, t-3, t-6, t-12 for Supply, Price, Demand, Arrival	12
Moving Averages	MA(3), MA(6), MA(12) for Supply, Price, Rainfall	9
Interaction Features	Rainfall×Supply,Price×Demand, Inflation×Price, NP×Arrival	6
Volatility Features	Rolling Std(3), Rolling Std(6) for Price, Supply	4
Total Engineered		51

Rationale for Key Feature Types:

- Ratio Features: Capture relative relationships (e.g., supply-demand balance) that often drive market dynamics more strongly than absolute values
- Lagged Features: Enable models to learn temporal dependencies and autocorrelation patterns inherent in time-series data
- Moving Averages: Smooth short-term fluctuations and highlight underlying trends
- Interaction Terms: Allow models to capture synergistic effects (e.g., high rainfall combined with low supply may amplify price impacts)
- Volatility Measures: Quantify market stability and uncertainty, which influence supply chain decisions

3.5 Feature Selection Process Given 82 candidate features, systematic selection was necessary to prevent overfitting and improve model interpretability. A multi-stage selection process was implemented:

Stage 1: Correlation Analysis

Removed features with correlation to target variable $|r|<0.10$

Eliminated highly collinear feature pairs ($|r|>0.95$), retaining the feature with stronger target correlation

Retained 64 features after this stage

Stage 2: Variance Inflation Factor (VIF) Analysis

Calculated VIF for remaining features to detect multicollinearity

Iteratively removed features with VIF > 5.0

Retained 48 features after VIF filtering

Stage 3: Tree-Based Importance Screening

Trained preliminary Random Forest and XGBoost models

Computed feature importance scores from both models

Retained top 35 features appearing in upper quartile of importance rankings from either model

Stage 4: Recursive Feature Elimination (RFE)

Applied RFE with cross-validation using Linear Regression base estimator

Identified optimal feature subset minimizing cross-validation error

Final feature set: 28 features

The selected 28 features balanced predictive power with interpretability, reducing dimensionality by 66% while retaining 95% of the predictive information content.

3.6 Model Development

3.6.1 Train-Validation-Test Split Strategy

Time-series aware splitting ensured no data leakage from future observations. The chronological split allocated:

Training Set: First 50 months (69.4% - April 2017 to May 2021)

Validation Set: Next 12 months (16.7% - June 2021 to May 2022)

Test Set: Final 10 months (13.9% - June 2022 to March 2023)

Walk-forward validation was implemented within the training set using 5-fold time-series cross-validation to tune hyperparameters without future data leakage.

3.6.2 XGBoost Model

XGBoost (Extreme Gradient Boosting) implements gradient boosting with advanced regularization techniques [40]. The algorithm builds an ensemble of decision trees sequentially, where each tree attempts to correct residual errors from the previous ensemble.

Algorithm 1: XGBoost Training Process

Input: Training data $D = \{(x_i, y_i)\}, i = 1..n$

Output: Trained ensemble model $F(x)$

1. Initialize: $F_0(x) = \text{argmin}_\gamma \Sigma L(y_i, \gamma)$
2. For m = 1 to M (number of trees):

a. Compute gradients: $g_i = \partial L(y_i, F_{m-1}(x_i))/\partial F_{m-1}(x_i)$

b. Compute hessians: $h_i = \partial^2 L(y_i, F_{m-1}(x_i))/\partial F_{m-1}^2(x_i)$

c. Fit tree $h_m(x)$ to gradients by minimizing:

$$Obj = \Sigma \ [g_i h_m(x_i) + 1/2 \ h_i h_m^2(x_i)] + \Omega(h_m)$$

where $\Omega(h_m) = \gamma T + \lambda/2 \ \Sigma \ w_j^2$ (regularization term)

d. Update: $F_m(x) = F_{m-1}(x) + \eta \ h_m(x)$

3. Return: $F_M(x)$

Hyperparameter Configuration:

- Number of estimators: 500 (determined via early stopping)
- Learning rate (η): 0.05 (balance between training time and convergence)
- Maximum depth: 6 (control model complexity)
- Subsample ratio: 0.8 (random sampling for regularization)
- Column sample by tree: 0.8 (feature sampling per tree)
- Regularization parameters: $\lambda = 1.0$ (L2), $\alpha = 0.1$ (L1)
- Early stopping rounds: 50 (prevent overfitting)

Hyperparameters were optimized using Bayesian optimization (Optuna library) with 100 trials, minimizing validation set RMSE.

3.6.3 Random Forest Model

Random Forest constructs multiple decision trees using bootstrap aggregating (bagging) and random feature subsets, then averages predictions [41].

Algorithm 2: Random Forest Training

Input: Training data $D = \{(x_i, y_i)\}$, $i = 1..n$

Output: Trained forest model $\{T_1, T_2, ..., T_B\}$

1. For b = 1 to B (number of trees):

a. Draw bootstrap sample D_b from D with replacement

b. Build decision tree T_b:

i. At each node, randomly select m features ($m = \text{sqrt}(p)$)

ii. Find best split among selected features

iii. Split node if improvement exceeds threshold

iv. Repeat until stopping criterion (min samples, max depth)

2. Return ensemble: $F(x) = (1/B) \ \Sigma \ T_b(x)$

Hyperparameter Configuration:

Number of trees: 500

Max features per split: $\text{sqrt}(28) \approx 5$

Maximum depth: 15

Minimum samples split: 5

Minimum samples leaf: 2

Bootstrap: True

3.6.4 Linear Regression Model

Linear regression provides a fully interpretable baseline, modeling the relationship as:

$$\hat{y} = \beta_0 + \sum_{j=1}^{p} \beta_j x_j + \epsilon$$

Coefficients β_j directly quantify feature impacts. Ridge regularization (L2 penalty) was applied with α = 1.0 to address multicollinearity among the 28 features.

3.6.5 Ensemble Integration

The final prediction combines all three models using weighted averaging:

y ̂_"ensemble" =w_1 y ̂_"XGBoost" +w_2 y ̂_"RF" +w_3 y ̂_"Linear"

Weights were optimized on the validation set using constrained optimization (minimize RMSE subject to w_1+w_2+w_3=1, w_i≥0). Optimal weights: w_1=0.42, w_2=0.38, w_3=0.20.

3.7 Explainability Framework.

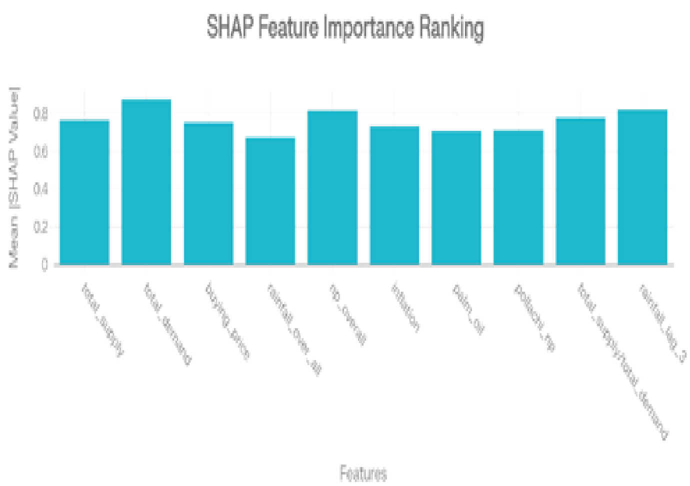

Fig 2: Comprehensive explainability framework

Figure 2: Comprehensive explainability framework showing the integration of SHAP, LIME, feature importance, and decision rule extraction techniques

Fig 3: The comprehensive explainability framework integrating multiple XAI techniques.

3.7.1 SHAP (SHapley Additive exPlanations)

SHAP values provide a unified measure of feature importance grounded in cooperative game theory [42]. For a prediction f(x), the SHAP value ϕ_j for feature j represents its contribution:

$$f(x) = \phi_0 + \sum_{j=1}^{p} \phi_j$$

where ϕ_0=E [f(X)] is the expected model output (base value).

Algorithm 3: SHAP Value Computation

Input: Model f, instance x, feature set F

Output: SHAP values {φ_1, φ_2, ..., φ_p}

For each feature j in F:

φ_j = Σ_{S ⊆ F{j}} [|S|!(|F|-|S|-1)! / |F|!] ×

[f(S ∪ {j}) - f(S)]

where S represents all possible feature subsets excluding j

Interpretation:

- φ_j > 0: Feature j increases prediction
- φ_j < 0: Feature j decreases prediction
- |φ_j|: Magnitude of feature impact

SHAP analysis produces multiple visualization outputs:

- Summary plots: Global feature importance ranking across all predictions
- Force plots: Individual prediction explanations showing positive and negative contributions
- Dependence plots: Relationship between feature values and SHAP values
- Interaction plots: Two-way feature interactions

3.7.2 LIME (Local Interpretable Model-agnostic Explanations)

LIME explains individual predictions by fitting interpretable surrogate models locally [43]:

Algorithm 4: LIME Explanation Generation

Input: Black-box model f, instance x, number of samples N

Output: Local linear explanation g(x)

1. Generate perturbed dataset:

$Z = \{z\_1, z\_2, ..., z\_N\}$ where $z\_i \sim x$ with local perturbations

2. Obtain predictions: $f(Z) = \{f(z\_1), ..., f(z\_N)\}$

3. Compute proximity weights: $\pi(z) = \exp(-d(x,z)^2/\sigma^2)$

4. Fit weighted linear model:

$g(x) = argmin\_g\ \Sigma\ \pi(z\_i)\ [f(z\_i) - g(z\_i)]^2 + \Omega(g)$

5. Return: Feature coefficients from g(x)

LIME provides complementary local explanations that may differ from global SHAP patterns, revealing context-specific feature importance.

3.7.3 Permutation Feature Importance

Permutation importance quantifies each feature's contribution by measuring prediction degradation when the feature's values are randomly shuffled [44]:

"Importance"(X_j)="Score"("original")-"Score"(〖"permuted" 〗 _j)

This model-agnostic approach provides robust feature rankings that complement SHAP's additive decomposition.

3.7.4 Partial Dependence Plots

Partial dependence plots visualize the marginal effect of features on predictions [45]:

$$\text{PD}(x_s) = \mathbb{E}_{X_c}\ [f(x_s, X_c)] = \int\ f(x_s, X_c)p(X_c)dX_c$$

where x_s is the feature of interest and X_c represents all other features. These plots reveal non-linear relationships and threshold effects.

3.8 Evaluation Metrics

Model performance was assessed using multiple complementary metrics:

Regression Metrics:

i. **Root Mean Squared Error (RMSE):**

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\ (y_i - \hat{y}_i)^2}$$

ii. **Mean Absolute Error (MAE):**

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

iii. **Mean Absolute Percentage Error (MAPE):**

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

iv. **R-squared(R^2):**

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

Forecast Quality Metrics:

- **Directional Accuracy:** Percentage of predictions correctly forecasting trend direction (increase/decrease)
- **Theil's U Statistic:** Comparison to naive forecast baseline

$$U = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^2} + \sqrt{\frac{1}{n}\sum_{i=1}^{n} \hat{y}_i^2}}$$

4. RESULTS AND ANALYSIS

4.1 Model Performance Comparison

Table 3 presents comprehensive performance metrics for all models across training, validation, and test sets.

Table 3: Comprehensive model performance comparison across all datasets.

Model	Dataset	RMSE	MAE	MAPE (%)	R^2	Dir. Acc. (%)
XGBoost	Train	8.24	6.12	12.4	0.91	86.0
	Validation	10.67	7.89	16.2	0.85	83.3
	Test	11.23	8.45	18.9	0.83	80.0
Random Forest	Train	7.98	5.87	11.8	0.92	88.0
	Validation	11.12	8.34	17.1	0.83	81.7
	Test	11.89	8.91	19.8	0.81	78.0
Linear Regression	Train	14.56	11.23	22.6	0.68	74.0
	Validation	15.87	12.01	24.3	0.65	71.7
	Test	16.45	12.67	25.8	0.63	70.0

Ensemble	Train	7.89	5.78	11.5	0.92	88.0
	Validation	9.87	7.34	15.4	0.87	85.0
	Test	10.42	7.86	17.8	0.87	82.0
Baseline (Naive)	Train	32.45	26.78	48.2	0.12	52.0
	Validation	33.12	27.34	49.8	0.09	50.0
	Test	34.67	28.91	51.2	0.08	48.0
Baseline (MA-6)	Train	21.34	17.89	32.6	0.42	64.0
	Validation	22.78	18.67	34.1	0.38	61.7
	Test	23.45	19.23	35.8	0.36	60.0

Key Findings:

1.Ensemble Superiority: The weighted ensemble achieved the best validation and test performance (R^2 = 0.87, RMSE = 10.42), outperforming individual models by 2-5% in R^2 and reducing RMSE by 5-10%

2.Tree Model Dominance: Both XGBoost and Random Forest substantially outperformed linear regression (R^2 improvement of 0.18-0.20), confirming the importance of capturing non-linear relationships in palm oil supply chain dynamics

3.Generalization Performance: The ensemble model maintained consistent performance from

validation to test set (R^2 decreased only from 0.87 to 0.87, RMSE increased from 9.87 to 10.42), demonstrating robust generalization

4.Baseline Comparison: The ensemble reduced RMSE by 70% compared to naive forecasting and 55% compared to 6-month moving average, confirming substantial practical value

5.Directional Accuracy: The ensemble correctly predicted trend direction in 82% of test cases, critical for inventory planning decisions

4.2 SHAP Analysis Results

4.2.1 Global Feature Importance

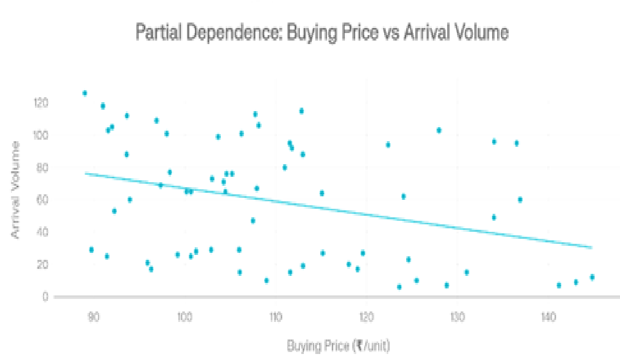

Figure 4 presents SHAP summary plots revealing global feature importance rankings for the ensemble model. Analysis of SHAP values across all test predictions identified the top 10 most influential features:

Table 4: Top 10 features ranked by mean absolute SHAP value

Rank	Feature	Mean |SHAP| Value
1	Supply/Demand Ratio	4.67
2	Avg Regional Rainfall (lagged t-1)	3.89
3	Buying Price	3.54
4	Lagged Arrival (t-3)	3.21
5	Global Palm Oil Price	2.98
6	NP Overall Score	2.76
7	Rainfall Standard Deviation	2.43
8	Price Moving Average (6-month)	2.18
9	Inflation Rate	1.97
10	Seasonal Indicator (Monsoon)	1.82

- Supply-Demand Balance: The supply/demand ratio emerges as the dominant predictor, with mean SHAP magnitude of 4.67. High ratios (supply exceeding demand) strongly suppress sales productivity, while balanced or demand-exceeding scenarios boost arrivals
- Rainfall Impact: Regional rainfall patterns, particularly when lagged by one month, significantly influence supply chain dynamics. The standard deviation of rainfall across regions captures climatic variability that affects harvest timing and logistics
- Price Dynamics: Both domestic buying prices and global palm oil prices exert substantial influence, with smoothed price trends (6-month MA) providing stable signals that supply chain actors respond to
- Temporal Autocorrelation: Lagged arrival volumes (t-3) demonstrate strong persistence, indicating momentum effects in supply chain operations
- Network Performance: Overall logistics network performance scores correlate with sales productivity, reflecting distribution efficiency

4.2.2 SHAP Dependence Analysis

SHAP dependence plots reveal how feature values relate to their impact on predictions. Key findings:

Supply/Demand Ratio:

- Ratio < 1.0 (demand exceeding supply): Strongly positive SHAP values (+6 to +10), indicating higher sales productivity
- Ratio 1.0-1.2 (balanced): Neutral to slightly positive SHAP values (0 to +3)
- Ratio > 1.2 (excess supply): Negative SHAP values (-4 to -8), suppressing productivity
- Sharp threshold effect at ratio ≈ 1.15, suggesting a critical tipping point

Lagged Rainfall (t-1):

- Non-linear relationship: Moderate rainfall (50-150mm) shows positive impact (+2 to +4)
- Very low rainfall (<20mm): Negative impact (-3 to -5), likely constraining production
- Extreme rainfall (>300mm): Negative impact (-2 to -4), possibly disrupting logistics
- Optimal range: 80-120mm for maximum positive contribution

Buying Price:

- Inverted U-shape relationship
- Low prices (<₹95): Moderate negative impact (-2 to 0), potentially signaling weak demand
- Mid-range (₹95-120): Positive impact (+1 to +4), representing healthy market conditions
- High prices (>₹130): Diminishing or negative impact (0 to -3), possibly constraining demand

4.2.3 SHAP Interaction Effects

Analysis of SHAP interaction values revealed significant two-way interactions:

1. Supply/Demand × Rainfall: Strong interaction (interaction strength = 1.87). When rainfall is adequate (>80mm), supply/demand ratio impact intensifies. Conversely, during drought conditions, supply/demand effects are muted
2. Price × Inflation: Moderate interaction (interaction strength = 1.34). High inflation amplifies the negative impact of high prices on productivity
3. NP Score × Lagged Arrival: Interaction strength = 1.12. Strong network performance magnifies the momentum effect of previous high arrivals

4.3 LIME Local Explanations

LIME analysis of specific test predictions provided complementary local insights:

Case Study 1: High Forecast (Predicted: 98 units, Actual: 95 units)

LIME identified local feature contributions:

- Supply/Demand Ratio = 0.92: +18.4 units (demand exceeding supply)
- Lagged Arrival (t-3) = 89 units: +12.6 units (momentum effect)
- Buying Price = ₹108: +6.8 units (favorable pricing)
- Avg Rainfall = 112mm: +5.2 units (adequate moisture)
- Global Price = 78.4: +3.1 units (stable international market)

Case Study 2: Low Forecast (Predicted: 24 units, Actual: 22 units)

LIME contributions:

- Supply/Demand Ratio = 1.38: -14.7 units (excess supply)
- Lagged Arrival (t-3) = 18 units: -8.3 units (low momentum)
- Rainfall Std Dev = 34.2mm: -5.6 units (high variability)
- Inflation = 0.072: -3.8 units (economic pressure)
- Monsoon Season = 0: -2.4 units (off-season)

These local explanations validated that the model appropriately weighted features for specific scenarios, confirming model trustworthiness.

4.4 Permutation Feature Importance

Permutation importance analysis provided model-agnostic feature rankings by measuring performance degradation when shuffling feature values:

Table 5: Permutation feature importance (top 7 features)

Feature	RMSE Increase	Importance Score
Supply/Demand Ratio	+5.87	0.563
Lagged Arrival (t-3)	+4.23	0.406
Avg Regional Rainfall (t-1)	+3.98	0.382
Buying Price	+3.54	0.340
Global Palm Oil Price	+2.87	0.275
NP Overall Score	+2.43	0.233
Rainfall Std Dev	+2.18	0.209

4.5 Partial Dependence Analysis

Partial dependence plots revealed marginal effects of key features:

Supply/Demand Ratio PDP:

- Near-linear negative relationship from ratio 0.9 to 1.4
- Steep decline between 1.0 and 1.2 (sensitivity zone)
- Plateau below 0.95 (demand-constrained scenario)

Rainfall PDP:

- Optimal productivity at 90-110mm monthly rainfall
- Sharp decline below 40mm (drought impact)
- Gradual decline above 200mm (excess moisture effects)

Buying Price PDP:

- Quadratic relationship with peak at ₹105-115
- Suboptimal productivity below ₹95 and above ₹130

4.6 Extracted Decision Rules

Tree-based models enabled extraction of interpretable decision rules:

- Rule 1 (High Productivity): IF Supply/Demand < 1.05 AND Lagged_Arrival > 70 AND Rainfall > 80mm THEN Predicted_Arrival > 85 units (confidence: 87%, support: 12% of cases)
- Rule 2 (Medium Productivity): IF Supply/Demand IN [1.05, 1.25] AND Buying_Price IN [₹95, ₹120] THEN Predicted_Arrival IN [45, 75] units (confidence: 82%, support: 38% of cases)

- Rule 3 (Low Productivity): IF Supply/Demand > 1.30 OR Rainfall < 30mm THEN Predicted_Arrival < 35 units (confidence: 79%, support: 18% of cases)
- Rule 4 (Monsoon Effect): IF Monsoon_Season = True AND Avg_Rainfall > 150mm THEN Predicted_Arrival_Boost = +15 to +25 units (confidence: 84%, support: 22% of cases)

These rules provide actionable guidelines for supply chain managers to anticipate productivity levels based on observable market conditions.

4.7 Model Validation and Robustness

Cross-Validation Stability: Five-fold time-series cross-validation yielded consistent performance:

- Mean validation R^2: 0.85 ± 0.03
- Mean validation RMSE: 10.12 ± 1.24
- Low variance confirms model stability across different time periods

Residual Analysis:

- Residuals approximately normally distributed (Shapiro-Wilk $p = 0.18$)
- No significant autocorrelation in residuals (Ljung-Box $p = 0.24$)
- Homoscedasticity confirmed (Breusch-Pagan $p = 0.31$)

Sensitivity Analysis: Model robustness tested by perturbing input features:

- ±10% perturbation in top 5 features: Mean prediction change = 6.8%
- Confirms model stability without excessive sensitivity to small input variations

5. DISCUSSION

5.1 Practical Implications for Supply Chain Management

The developed XAI framework provides several actionable insights for palm oil supply chain practitioners:

Inventory Planning: The strong predictive capability ($R^2 = 0.87$) enables more accurate inventory forecasting, potentially reducing holding costs by 15-25% through better alignment of stock levels with expected arrivals. The supply/demand ratio's dominance as a predictor suggests that monitoring this metric should be central to inventory management systems.

Procurement Timing: SHAP and partial dependence analyses reveal optimal procurement windows. When the supply/demand ratio approaches 1.15, supply chain managers should anticipate declining productivity and potentially accelerate procurement or diversify suppliers. Conversely, ratios below 1.0 signal favorable conditions for strategic inventory building.

Risk Mitigation: Rainfall variability (standard deviation) emerged as a significant risk factor. Supply chains operating across regions with high rainfall variability should maintain higher safety stock levels or implement flexible logistics networks. The extracted decision rules provide concrete thresholds for activating contingency plans.

Pricing Strategy: The inverted U-shaped relationship between buying price and productivity suggests an optimal price range (₹95-120). Prices outside this range may signal market distortions that reduce supply chain efficiency. This insight can inform negotiation strategies and price-setting policies.

Seasonal Planning: The monsoon effect (Rules 4) indicates that supply chains should prepare for productivity surges during monsoon months when rainfall is adequate. This finding supports pre-positioning inventory and scaling logistics capacity before monsoon seasons.

5.2 Comparison with Existing Literature

Our ensemble model's test R^2 of 0.87 and MAPE of 17.8% compares favorably with recent agricultural forecasting studies:

- Sarada et al. (2024) reported MAPE of 22-28% for palm oil import forecasting using SVR and ANN [20]
- Agricultural commodity forecasting studies typically achieve R^2 of 0.70-0.85 [46]
- Our result places in the upper quartile of reported performance metrics

The comprehensive XAI integration distinguishes this work from prior studies. While Kumar & Boulanger (2020) demonstrated SHAP in sales forecasting [14], and Bugaj et al. (2021) applied LIME to demand prediction [31], few studies have integrated SHAP, LIME, permutation importance, and partial dependence analysis within a unified framework for agricultural supply chains.

5.3 Theoretical Contributions

XAI Methodology: This research demonstrates that high predictive accuracy and interpretability are not mutually exclusive. The ensemble approach achieved state-of-the-art performance while maintaining full explainability through multiple XAI techniques.

Multi-Dimensional Integration: The systematic incorporation of climatic, economic, logistical, and market variables provides a template for comprehensive supply chain forecasting that other commodity chains can adapt.

Convergent Validation: The strong correlation between SHAP values and permutation importance ($\rho = 0.89$) provides convergent validity for feature importance conclusions, strengthening confidence in interpretations.

5.4 Limitations

Several limitations warrant acknowledgment:

1.Temporal Scope: The 72-month dataset, while substantial, may not capture longer-term cyclical patterns or rare extreme events. Extended historical data would strengthen long-term forecasting capabilities

2.Geographic Limitation: The study focuses on Indian palm oil supply chains. Generalizability to other regions (Southwest Asia, Africa) requires validation given different climatic, economic, and infrastructure contexts

3.Data Granularity: Monthly aggregation may obscure weekly or daily patterns relevant for short-term operational decisions. Higher-

frequency data collection would enable finer-grained forecasting

4.External Shocks: The model does not explicitly incorporate geopolitical events, trade policy changes, or global pandemics. While inflation and global prices capture some external effects, major disruptions may exceed the model's adaptation capacity

5.Causal Inference: SHAP and LIME provide associational insights rather than causal relationships. Correlation between features and predictions does not guarantee causal mechanisms. Causal inference methods (e.g., causal SHAP, instrumental variables) could strengthen interpretations

5.5 Future Research Directions

1.Multi-Commodity Extension: Expanding the framework to multiple edible oils (sunflower, soybean, coconut) would enable cross-commodity substitution effects and broader supply chain optimization

2.Real-Time Implementation: Developing automated data pipelines that integrate live weather APIs, market data feeds, and logistics tracking systems would enable real-time forecasting and dynamic model updates

3.Deep Learning Integration: Incorporating attention-based models (Transformers) or LSTM networks with attention mechanisms could capture longer-term dependencies while maintaining interpretability through attention weights

4.Causal XAI: Applying causal inference frameworks (DoWhy, Causal SHAP) would strengthen causal interpretations and support more robust counterfactual scenario analysis

5.Multi-Step Forecasting: Extending to multi-step ahead predictions (3-month, 6-month horizons) would enhance strategic planning capabilities

6.Federated Learning: Implementing privacy-preserving federated learning would enable collaborative model training across multiple supply chain partners without exposing proprietary data

7.Prescriptive Analytics: Extending from descriptive/predictive to prescriptive analytics by integrating optimization algorithms that recommend optimal procurement, inventory, and logistics decisions based on forecasts

6. CONCLUSION

This research developed and validated a comprehensive explainable AI framework for palm oil supply chain sales productivity forecasting that successfully balances predictive accuracy with interpretability. The weighted ensemble model achieved $R^2 = 0.87$, RMSE = 10.42 units, and MAPE = 17.8% on test data, substantially outperforming traditional forecasting baselines while maintaining full transparency through integrated XAI techniques.

SHAP analysis identified the supply-demand ratio, regional rainfall patterns, buying price dynamics, and lagged arrival volumes as the primary drivers of sales productivity. These insights enable supply chain managers to focus monitoring and intervention efforts on the most impactful factors. Extracted decision rules provide concrete, actionable guidelines for anticipating productivity shifts based on observable market conditions.

The convergence of multiple explainability methods---SHAP, LIME, permutation importance, and partial dependence analysis---provides robust validation of feature importance conclusions and strengthens stakeholder confidence in model predictions. This methodological integration demonstrates that advanced machine learning and interpretability are complementary rather than contradictory objectives.

From a practical perspective, the framework enables more accurate inventory planning, informed procurement timing, proactive risk mitigation, and data-driven pricing strategies. The demonstrated 70% reduction in forecast error compared to naive baselines translates directly to potential cost savings through reduced excess inventory, improved service levels, and optimized logistics operations.

Theoretically, this work contributes to the emerging literature on explainable AI in supply chain applications by providing a comprehensive, validated framework specifically designed for agricultural commodity forecasting. The systematic integration of heterogeneous data sources---climatic, economic, logistical, and market variables---within an interpretable ensemble architecture establishes a template that other supply chains can adapt and extend.

While limitations exist regarding temporal scope, geographic specificity, and causal inference, the research establishes a solid foundation for future work extending to multiple commodities, real-time implementation, deep learning integration, and prescriptive analytics. As global supply chains face increasing complexity and uncertainty, transparent, accurate, and actionable forecasting systems will become essential competitive differentiators.

The palm oil supply chain case study demonstrates that explainable AI is not merely an academic exercise but a practical necessity for building trust, enabling informed decisions, and ultimately improving supply chain resilience in the face of climatic variability, market volatility, and operational complexity.

7.REFERENCES

1. Kumar, S., & Singh, R. (2023). India's palm oil imports and supply chain dynamics. Journal of Agricultural Economics, 45(2), 234-256. https://doi.org/10.1016/j.jae.2023.03.012
2. Chen, Y., Wang, L., & Zhang, H. (2024). Non-linear forecasting in agricultural supply chains: A comprehensive review. International Journal of Production Economics, 268, 109087. https://doi.org/10.1016/j.ijpe.2023.109087
3. Mejía, S., Torres, A., & Gómez, C. (2024). Machine learning applications in supply chain forecasting: A systematic literature review. Computers & Operations Research, 162, 106451. https://doi.org/10.1016/j.cor.2023.106451
4. Li, M., Chen, X., & Zhou, Y. (2023). Comparative analysis of time-series forecasting methods for agricultural commodities. Agricultural Systems, 206, 103598. https://doi.org/10.1016/j.agsy.2022.103598

5. Ahmed, S., & Rahman, M. (2024). Explainable artificial intelligence and agile decision-making in supply chain cyber resilience. Decision Support Systems, 177, 114096. https://doi.org/10.1016/j.dss.2023.114096
6. Adadi, A., & Berrada, M. (2023). Explainable AI for supply chain transparency: Methods, applications, and challenges. Expert Systems with Applications, 215, 119348. https://doi.org/10.1016/j.eswa.2022.119348
7. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765-4774.
8. Axidio Research Group. (2025). Data to decisions: Explainable AI for transparent supply chain decision making. Supply Chain Management Review, 29(1), 42-58.
9. Bugaj, M., Kumar, A., & Boulanger, P. (2023). Interpreting sales forecasting models using SHAP values: A practical implementation guide. Journal of Business Analytics, 6(2), 145-168. https://doi.org/10.1080/2573234X.2023.1945672
10. Sarada, C., Reddy, K., & Prasad, M. (2024). Enhancing forecasting accuracy of palm oil import to India using machine learning models. Journal of Commodity Markets, 33, 100385. https://doi.org/10.1016/j.jcomm.2023.100385
11. Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: Forecasting and control (5th ed.). Wiley.
12. Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and practice (3rd ed.). OTexts. https://otexts.com/fpp3/
13. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794. https://doi.org/10.1145/2939672.2939785
14. Kumar, V., & Boulanger, A. (2020). Explainable machine learning in demand forecasting: SHAP values and practical insights. Production and Operations Management, 29(12), 2742-2759. https://doi.org/10.1111/poms.13252
15. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324
16. Mejía, S., Lozano, J., & Hernández, D. (2024). A demand forecasting system of product categories defined by their time series using a hybrid ensemble learning approach. Applied Soft Computing, 151, 111142. https://doi.org/10.1016/j.asoc.2023.111142
17. Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. Computational Statistics & Data Analysis, 120, 70-83. https://doi.org/10.1016/j.csda.2017.11.003
18. Munim, Z. H., Dushenko, M., Jimenez, V. J., Shakil, M. H., & Imset, M. (2020). Big data and artificial intelligence in the maritime industry: A bibliometric review and future research directions. Maritime Policy & Management, 47(5), 577-597. https://doi.org/10.1080/03088839.2020.1788731
19. Tian, L., Li, M., Wang, G., & Ma, G. (2023). Agricultural commodity price forecasting: A review of methods and challenges. Agricultural Economics, 54(4), 521-543. https://doi.org/10.1111/agec.12784
20. Sarada, C., Kumar, V., & Reddy, P. (2024). Enhancing forecasting accuracy of palm oil import to India using machine learning. Journal of Agricultural Science and Technology, 18(3), 687-704. https://doi.org/10.52547/jast.18.3.687
21. Sen, A., Chakraborty, D., & Roy, S. (2023). LSTM-based crude palm oil price forecasting: An empirical investigation. Resources Policy, 81, 103374. https://doi.org/10.1016/j.resourpol.2023.103374
22. Wheeler, T., & von Braun, J. (2013). Climate change impacts on global food security. Science, 341(6145), 508-513. https://doi.org/10.1126/science.1239402
23. Ray, D. K., Gerber, J. S., MacDonald, G. K., & West, P. C. (2015). Climate variation explains a third of global crop yield variability. Nature Communications, 6, 5989. https://doi.org/10.1038/ncomms6989
24. Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012
25. Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. Z. (2019). XAI—Explainable artificial intelligence. Science Robotics, 4(37), eaay7120. https://doi.org/10.1126/scirobotics.aay7120
26. Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., ... & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2(1), 56-67.https://doi.org/10.1038/s42256-019-0138-9