

ENHANCING AGRICULTURAL SUPPLY–DEMAND ANALYSIS THROUGH AGRICULTURAL PRICE INDEX FORECASTING USING HYBRID FEATURE SELECTION AND INTERPRETABLE MACHINE LEARNING

B. Rajmohan¹ and A. Latha²

¹School of Computational Engineering, Takshashila University, India

²School of Core Engineering, Takshashila University, India

Abstract

Agricultural Supply–Demand systems are shaped by dynamic and interdependent factors, including climate variability, market fluctuations, and policy interventions, which challenge traditional forecasting models in both accuracy and interpretability. This study proposes CLIF (Closed-Loop Adaptive Interpretable Forecasting), a framework in which SHAP (SHapley Additive exPlanations)-based explanations actively guide model learning. By incorporating explanation-driven feedback and a temporal forgetting mechanism, the framework enables adaptive feature relevance under evolving market conditions. LASSO is employed for sparse feature selection, while Random Forest captures nonlinear relationships, forming a feedback-driven predictive system responsive to structural shocks and policy changes. Evaluation on a decade dataset of Indian wheat and rice markets demonstrates the feasibility of improving forecasting performance compared with ARIMA, SVM, LASSO, and standalone Random Forest models, achieving the lowest RMSE (0.55) and highest R^2 (0.91). Feature attribution analysis identifies fertilizer prices, climate variability, and export tariffs as key drivers, providing actionable insights. The framework demonstrates potential as an interpretable and adaptive approach for Agricultural Price Index forecasting and Supply–Demand analysis in dynamic agricultural environments.

Keywords:

Supply Chain, Feature Selection, Interpretable Machine Learning, SHAP, LASSO, CLIF

1. INTRODUCTION

The global agricultural sector faces increasing pressure to meet the needs of a population projected to reach 9.7 billion by 2050. The global Agri-food supply chain supports the livelihoods of over 1.3 billion people worldwide. Agricultural Supply–Demand systems are governed by complex interactions among production costs, climatic conditions, trade policies, market dynamics, and consumer behavior.

Despite technological advancements, inefficiencies in agricultural supply chains continue to result in significant economic losses, food waste, and price volatility. According to the Food and Agriculture Organization (FAO), approximately one-third of all food produced globally is lost or wasted, resulting in substantial economic losses. In developing countries, more than 20% of agricultural produce is lost after harvest due to inadequate demand forecasting and inefficient logistics.

Since 2020, staple crop prices, particularly wheat and rice, have increased significantly due to the COVID-19 pandemic and the Russia–Ukraine conflict. These disruptions resulted in an estimated 30% decline in wheat exports and a 40% increase in wheat prices [1]. Furthermore, global food demand is projected to increase by 50% by 2050 owing to continued population growth.

However, annual agricultural productivity growth has slowed to 1.1%, which is considerably below the estimated 2.4% annual growth required to satisfy future food demand.

Climate change has further intensified these challenges through extreme weather events that reduce agricultural productivity and increase yield variability. According to Ortiz-Bobea et al. [2], global agricultural productivity has declined by approximately 21% since 1961 due to anthropogenic climate change. Extreme climatic events, including the floods in Pakistan and the droughts in East Africa during 2023, contributed to a 15% reduction in cereal production [3], thereby exacerbating food insecurity. In addition, geopolitical and economic disruptions have strained global agricultural supply chains. Export restrictions, such as India's 2022 wheat export ban, have further destabilized agricultural markets and exposed the fragility of global food systems [4].

In 2018, NITI Aayog established a Working Group led by Dr. Parmod Kumar to assess agricultural production trends, Agricultural Price Index forecasting for Supply–Demand analysis, and resource management in India. The committee developed projections for major agricultural commodities and recommended policy measures to improve resource allocation, supply chain efficiency, and national food security [5].

Accurate forecasting of agricultural market dynamics remains challenging due to data limitations, nonlinear relationships, and scalability constraints. Inconsistent data collection practices and missing values reduce dataset reliability, while more than 40% of agricultural datasets in some developing regions remain incomplete, making reliable forecasting difficult. Conventional statistical models, such as ARIMA, are unable to effectively capture complex nonlinear interactions among variables including fertilizer costs, climatic conditions, and crop yields, thereby limiting forecasting accuracy. Although deep learning models provide strong predictive capability, their black-box nature limits interpretability, reducing stakeholder confidence and restricting their adoption in policy-oriented decision making.

In this study, the Annual Agricultural Price Index (API) is adopted as the target variable and serves as a proxy indicator of agricultural Supply–Demand equilibrium. The API reflects the combined influence of agricultural production, consumption, trade policies, and climatic variability on market behavior, making it an effective indicator for forecasting Agricultural Price Index dynamics as a proxy for Supply–Demand equilibrium.

Scalability and real-time adaptability also remain significant challenges because many forecasting models are computationally intensive and struggle to respond to sudden policy interventions or extreme climatic events. Furthermore, many agricultural organizations continue to face difficulties integrating real-time

data into forecasting systems, highlighting the need for adaptive, interpretable, and computationally efficient forecasting frameworks.

To address these challenges, this study proposes a Closed-Loop Adaptive Interpretable Forecasting (CLIF) framework that integrates LASSO-based feature selection, Random Forest-based nonlinear learning, and SHAP-driven explanation-guided adaptive feature updating for forecasting the API. The proposed framework aims to improve predictive accuracy while maintaining model interpretability and adaptability under evolving agricultural market conditions.

The remainder of this paper is organized as follows. Section 2 reviews the related literature and identifies the research gap. Section 3 presents the proposed CLIF framework and experimental methodology. Section 4 discusses the experimental results and performance evaluation. Finally, Section 5 concludes the paper and outlines future research directions.

2. RELATED WORKS

Feature selection is crucial to optimize the Agri-supply chains due to increased forecasting accuracy and low computing complexity. The Least Absolute Shrinkage and Selection Operator (LASSO) is widely used because it can effectively manage high-dimensional datasets by selecting important features and applying regularization, which helps improve model accuracy and prevent overfitting. Assumption of linearity makes it unable to adequately represent the complex, nonlinear relationships found in agricultural systems. Other methods, such as deep learning and tree-based models, provide greater flexibility in modelling nonlinear connections, even if LASSO is widely employed in supply chain analytics.

Kamble et al. [6] highlighted that agricultural supply chains face significant challenges related to data integration, traceability, and real-time decision-making, emphasizing the importance of advanced data analytics.

Lobell et al. [7] demonstrated that climate variability significantly influences global crop productivity, highlighting the need for adaptive forecasting models to reduce climate risks in farming.

To improve the accuracy of agricultural yield forecasts, Chlingaryan et al. [8] showed that machine learning models effectively capture nonlinear relationships in agricultural systems, leading to improved crop prediction accuracy. Their approach captures both linear and nonlinear patterns in complex agricultural systems, showing the benefit of using multiple modeling strategies.

Mohanarajan et al. [9] developed a Random Forest-based crop prediction system that integrates soil nutrients, climatic variables, and a web-based decision support interface to assist farmers in selecting suitable crops. That demonstrated the effectiveness of Random Forest for agricultural prediction tasks.

LASSO employs L1 regularization to simultaneously perform feature selection and coefficient shrinkage, thereby reducing overfitting and improving model interpretability [10]. Prior studies have demonstrated the effectiveness of LASSO for sparse feature selection in agricultural forecasting [11], particularly for identifying dominant yield drivers while reducing model

complexity. An advanced predictive model was designed by Kalaiarasi et al. [12] that use hybrid machine learning algorithms within an intelligent decision support system. By combining multiple techniques, the system enhances planning and resource allocation in agriculture. This approach helps improve productivity and contributes to food security.

Regression analysis for agricultural yield prediction was compared by Priti Prakash Jorvekar et al. [13]. The author assessed various predictive modeling methodologies in order to identify the best regression strategies for crop yield forecasting. The study aimed to improve the accuracy and reliability of yield forecasts, supporting better decisions in farming and agribusiness.

IoT and AI applications in smart agriculture, encompassing pre, during, and post-harvest operations, were reviewed by Kasera et al. [14]. For increased agricultural efficiency, the study emphasizes developments in supply chain optimization, automation, and precision farming.

Kumar et al. [15] explored how IoT and machine learning can improve modern agriculture by boosting productivity, efficiency, and decision-making. Their study highlighted the role of these technologies in automation, yield prediction, and precision farming.

SHAP, a unified framework for interpreting complex machine learning model predictions, was introduced by Lundberg and Lee [16]. This work establishes the theoretical foundations for understanding how different features influence model outputs. SHAP has become a widely used interpretability tool, offering clear and quantitative insights into feature contributions across various fields, including agriculture. This helps build trust in model results and supports informed decision making.

FAOStat (Food and Agriculture Organization Statistical Database) (2010–2024) and Indian Agricultural Ministry Reports [17] [18] provide key datasets widely used in agricultural forecasting research. FAOStat offers broad, global agricultural data over more than a decade, while the Indian Agricultural Ministry reports deliver detailed, region-specific insights. Combining these sources helps researchers capture both global trends and local specifics, which is essential for building accurate forecasting models for agricultural supply networks.

Recent studies highlight the transformative role of AI, ML, and feature selection in agriculture and supply chains. Varma et al. [19] showed that algorithms like RF, LASSO, and forward/backward selection improve crop yield forecasting by reducing irrelevant features and computational costs. Elufioye et al. [20] confirmed that AI enhances demand forecasting and operational efficiency in Agri-supply chains, though challenges like data quality and infrastructure persist.

In broader SCM contexts, Younis et al. [21] and Singh [22] emphasized the growing but early-stage adoption of AI/ML for logistics and predictive decision-making. IoT's role is also notable, with Onike [23] reporting significant gains in yield, water savings, and traceability. Ault and Dief [24] further stressed that AI's impact is strongest when integrated across supply chain locations collaboratively. These insights collectively reflect how digital technologies are reshaping agriculture and supply systems toward greater efficiency and sustainability.

2.1 PROBLEM STATEMENT

The agricultural Supply–Demand system is a complex, dynamic ecosystem influenced by interconnected factors such as climate variability, geopolitical policies, market trends, and consumer behavior. These factors collectively influence the Agricultural Price Index (API), which reflects the interaction between supply and demand conditions. In this study, the API is considered the target variable and is used as a proxy indicator of agricultural Supply–Demand equilibrium. Accurate forecasting of the API is essential for supporting resource allocation, reducing food waste, and maintaining market price stability.

Traditional forecasting models face significant challenges, including poor data quality (e.g., missing values, inaccuracies, and inconsistencies), limited ability to capture nonlinear relationships among agricultural variables (e.g., fertilizer costs, crop yields, and climate variability), and lack of interpretability in advanced machine learning models. Mishra et al. [25] emphasized that building resilient agri-food supply chains requires a combination of dynamic capabilities rather than relying on a single strategy. Therefore, there is a need to develop a well-contextualized and interpretable forecasting framework for supporting decision-making in agri-food supply chains.

Accurate forecasting of the API, which serves as a proxy for Supply–Demand equilibrium, is inherently challenging due to non-stationary market dynamics arising from seasonal variability, climate fluctuations, and policy-induced shocks. Traditional statistical and machine learning models often assume static feature importance and operate in sequential pipelines where feature selection, prediction, and interpretability are treated as independent stages. Such approaches limit adaptability under structural changes and restrict interpretability to post hoc explanation rather than using it to improve the learning process.

There is therefore a need for a forecasting framework that can (i) adapt to evolving agricultural market conditions, (ii) maintain interpretability for policymakers and decision-makers, and (iii) integrate explanation into the learning process rather than treating it as a separate analytical layer.

This study proposes a Closed-Loop Adaptive Interpretable Forecasting (CLIF) framework for forecasting the Annual API while providing interpretable insights into agricultural Supply–Demand equilibrium. The framework is characterized by the following contributions:

1. **Constrained Hierarchical Feature Filtering:** LASSO-based sparsity and Random Forest nonlinear ranking are combined in a structured dual-stage feature selection process to stabilize learning and reduce multicollinearity.
2. **Explanation-Driven Learning:** SHAP-based explanations are incorporated into the model update process, enabling interpretability to actively guide subsequent learning rather than serving only descriptive purposes.
3. **Time-Adaptive Feature Relevance:** A temporal weighting mechanism with a forgetting factor enables feature importance to evolve across forecasting windows in response to climate variability and policy changes.
4. **Shock-Aware System-Level Modeling:** The framework jointly models production, consumption, trade, climate,

logistics, and policy-related variables to capture structural changes affecting agricultural market prices.

3. PROPOSED METHODOLOGY

The proposed methodology introduces a Closed-Loop Adaptive Interpretable Forecasting Framework (CLIF) designed to address the challenges of Agricultural Price Index forecasting for Supply–Demand analysis by combining the strengths of statistical methods and machine learning with interpretability. The framework consists of three phases as shown in Figure 1

- Stage 1: Data Pre-processing
- Stage 2: Hybrid Feature Selection.
- Stage 3: Interpretability Enhanced Prediction

3.1 DATA PRE-PROCESSING

In the data pre-processing stage, raw agricultural datasets are cleaned to handle missing values by using KNN (K-Nearest Neighbors) algorithm, Z-score normalization technique is adopted for uniformity, and features are enriched with derived features such as seasonal indices and climate volatility metrics to capture temporal and spatial patterns. The output of this stage is cleaned and normalized dataset ready for the next stage of processing.

3.1.1 K-Nearest Neighbors Imputation:

KNN imputation was employed to handle missing values in the agricultural dataset before feature selection and model training. The KNN imputer estimates missing values by identifying the five nearest neighboring observations ($k = 5$) using the Euclidean distance metric and imputes missing entries based on the uniform average of their corresponding values. Since all variables considered in this study are numerical, no additional handling of categorical attributes was required, and the imputation was performed feature-wise. Compared with simple mean or median imputation, KNN preserves the local relationships among agricultural variables such as soil conditions, weather, fertilizer costs, and logistics, thereby maintaining the underlying data distribution. However, because KNN considers all available features during distance computation, it is sensitive to irrelevant or redundant variables that may introduce noise. Therefore, KNN imputation is followed by the proposed hybrid feature selection framework, where LASSO removes redundant features and Random Forest identifies the most informative predictors before forecasting, as formulated in Eq.(1).

$$d(X, Y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (1)$$

where,

$X_i = (x_1, x_2, \dots, x_n)$ represents the feature values of the first data point.

$Y_i = (y_1, y_2, \dots, y_n)$ represents the feature values of the second data point.

n is the total features.

The KNN imputation was performed before feature selection to preserve the maximum amount of information available for subsequent analysis. Since the percentage of missing values was

relatively low, KNN imputation was selected because it better preserves local data structures than simple statistical imputation techniques.

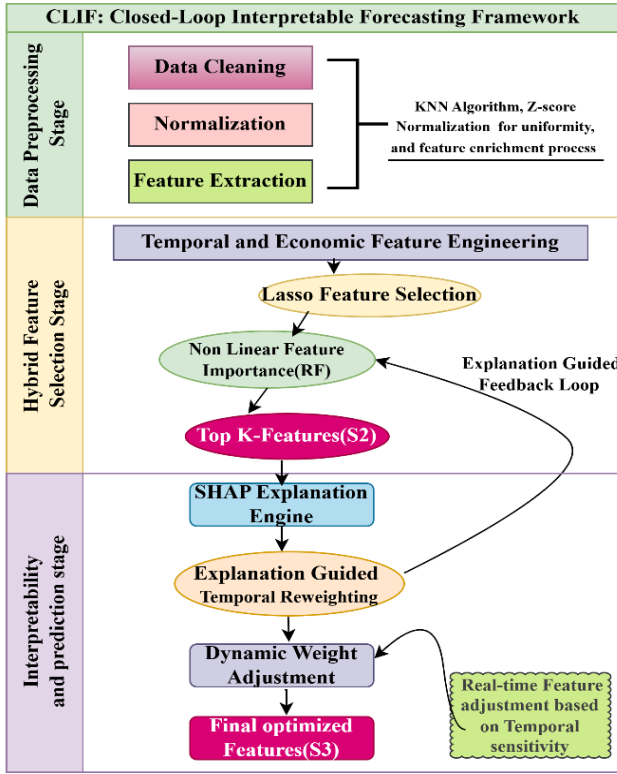


Fig.1. Closed-Loop Adaptive Interpretable Forecasting Framework (CLIF)

3.1.2 Z Score Normalization:

Following missing value imputation, all numerical features were standardized using the Z-score normalization technique to ensure a uniform scale and prevent features with larger numerical ranges from dominating the learning process. Agricultural variables such as temperature (°C), soil pH, humidity (%), fertilizer cost, and transportation time (hours) are measured in different units and scales. Standardizing these variables improves numerical stability and enables fair comparison among features during model training, particularly for distance-based preprocessing and regularization methods such as KNN imputation and LASSO feature selection.

For a given feature X , each value x_i is transformed using:

$$z_i = \frac{x_i - \mu}{\sigma} \quad (2)$$

where,

z_i = Standardized value of x_i

x_i = Original value.

μ = Mean of feature X .

σ = Standard deviation of feature X .

3.2 HYBRID FEATURE SELECTION STAGE

3.2.1 LASSO Feature Selection:

The first stage of the proposed hybrid feature selection framework employs the Least Absolute Shrinkage and Selection

Operator (LASSO) to eliminate irrelevant and redundant features from the preprocessed agricultural dataset. LASSO performs feature selection by incorporating L1 regularization, which shrinks less informative feature coefficients toward zero while retaining only the most significant predictors. This reduces model complexity, mitigates multicollinearity, and improves generalization, particularly for datasets containing correlated agricultural variables.

In the proposed CLIF framework, LASSO serves as the initial filtering mechanism to generate an optimized feature subset (S_1), which is subsequently refined using Random Forest-based nonlinear feature ranking. This hierarchical feature selection strategy combines the sparsity-inducing capability of LASSO [10] with the nonlinear learning capability of Random Forest to improve forecasting performance.

$$\beta = \operatorname{argmin} \left(\|Y - X\beta\|^2 + \lambda \|\beta\| \right) \quad (3)$$

where,

Y : Target variable (Annual Agricultural Price Index).

X : Feature matrix.

β : Coefficient vector.

λ : Regularization parameter.

3.2.2 Random Forest Feature Ranking:

Random Forest (RF) is employed to capture nonlinear interactions among agricultural variables and rank feature importance within the candidate feature set S_1 . The RF model was implemented using the scikit-learn library with decision trees ($n_estimators = 100$), a maximum tree depth of ($max_depth = 5$), minimum samples required to split an internal node of ($min_samples_split = 2$), minimum samples required at each leaf node of ($min_samples_leaf = 2$), and bootstrap sampling enabled. A fixed $random_state = 42$ was used to ensure experimental reproducibility.

The Random Forest model estimates feature importance by measuring the cumulative reduction in Gini impurity contributed by each feature across all decision trees. Agricultural variables such as climate variability, fertilizer costs, production, trade policies, and logistics exhibit complex nonlinear interactions that are difficult to model using linear approaches. Random Forest effectively captures these nonlinear relationships without requiring explicit feature engineering. Based on the computed importance scores, the top 10 ranked features were selected to form the refined feature subset S_2 .

$$I(X_i) = \sum_{t=1}^T \frac{\Delta G_t(X_i)}{N_t} \quad (4)$$

$I(X_i)$ = Importance score of feature X_i .

T = Total number of trees in the forest.

$\Delta G_t(X_i)$ = Gini impurity reduction by feature X_i in tree t

N_t = Total number of Nodes in tree t .

To preserve temporal dependence, model evaluation was performed using Rolling-Origin Cross-Validation (RO-CV), where each training fold contained only historical observations and the subsequent annual observation was reserved for testing. Overfitting was mitigated through the combined use of LASSO-based feature selection, Random Forest ensemble learning, and

temporal validation, thereby reducing model variance while maintaining robustness for the limited annual dataset.

Algorithm 1 Closed-Loop Feature Selection

Require: Dataset $D(n \times m)$, Target Y , forgetting factor $\alpha=0.6$

Ensure: Optimized adaptive feature set S^*

- 1: Preprocess D (impute missing values, normalize)
- 2: Initialize feature matrix X
- 3: for each Rolling-Origin Cross-Validation (RO-CV) training fold do
- 4: Apply LASSO feature selection:

$$\beta = \operatorname{argmin} \left(\|Y - X\beta\|^2 + \lambda \|\beta\| \right)$$

- 5: Retain non-zero coefficient features $\rightarrow S_1$
- 6: Train Random Forest on S_1 , compute nonlinear feature importance
- 7: Retain top- k features $\Rightarrow S_2$
- 8: Train forecasting model using S_2
- 9: Compute SHAP values for all selected features
- 10: Calculate temporal feature importance drift

$$\Delta I = I_t - I_{t-1}$$
- 11: Update feature weights

$$W_t = \alpha W_{t-1} + (1 - \alpha) \Delta I$$
- 12: Re-rank features using updated weights
- 13: Obtain optimized adaptive feature subset S^*
- 14: End for
- 15: Return S^*

The closed-loop update is executed after each Rolling-Origin Cross-Validation (RO-CV) training fold. SHAP values are computed using the trained Random Forest model for the current fold, and the resulting feature importance drift is used to update feature weights for the subsequent fold through the temporal weighting mechanism.

3.2.3 Closed-Loop Update Mechanism:

The closed-loop update mechanism enables the proposed CLIF framework to adapt feature importance after each Rolling-Origin Cross-Validation (RO-CV) fold. Rather than using SHAP solely for post-hoc interpretation, the computed SHAP values are incorporated into the feature-weight update process, allowing the model to adapt to evolving agricultural market conditions. SHAP does not modify the Random Forest model parameters directly instead, SHAP-derived importance scores are used to update the adaptive feature-weighting mechanism for feature ranking in subsequent RO-CV folds.

Algorithm 2 Closed-Loop Feature Weight Update Mechanism

Input: Training dataset D , selected feature set S_2 , forgetting factor α

Output: Updated feature weights and Annual API predictions

- 1: For each RO-CV training fold do
- 2: Predict the next-year Agricultural Price Index
- 3: Compute SHAP values for all selected features
- 4: Estimate feature importance drift:

$$\Delta I = I_t - I_{t-1}$$

- 5: Update feature weights

$$W_t = \alpha W_{t-1} + (1 - \alpha) \Delta I$$

- 6: End for

- 7: Return final API predictions and updated feature weights

The closed-loop update is executed after each RO-CV fold. SHAP values are computed using the trained Random Forest model for the current fold, and the resulting feature importance drift is incorporated into the temporal weighting mechanism. The updated feature weights are then used in the subsequent fold, enabling the CLIF framework to adapt to changes in agricultural market dynamics while preserving model interpretability.

3.3 INTERPRETABILITY AND PREDICTION STAGE

3.3.1 SHAP based Interpretability:

The interpretability and prediction stage employs SHAP to quantify the contribution of each feature toward predicting the Annual Agricultural Price Index. The Random Forest model is trained using the optimized feature subset (S_2) obtained from the hybrid feature selection stage. SHAP is then applied to explain individual predictions and derive global feature importance patterns. Unlike conventional forecasting frameworks, the proposed CLIF framework incorporates SHAP-derived feature importance into the adaptive learning process, enabling feature relevance to be continuously updated across successive forecasting windows.

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f(S \cup \{i\}) - f(S)] \quad (5)$$

where,

ϕ_i = SHAP value of feature i .

F = Set of all features.

S = Any subset of features excluding i .

$|S|$ = Size of subset S .

$f(S)$ = Model output when only subset S of features is used.

SHAP values are computed after training the Random Forest model within each Rolling-Origin Cross-Validation (RO-CV) fold. These values quantify the contribution of each selected feature to the current prediction of the API. The computed SHAP importance scores are subsequently incorporated into the hybrid feature-weight update mechanism through the temporal weighting strategy. The updated feature weights are then utilized during the next RO-CV training fold, thereby transforming model interpretability into an active learning signal rather than a post-hoc explanation.

3.3.2 Dynamic Feature Weighting:

Agricultural markets exhibit non-stationary behavior, where the influence of individual features changes over time due to climate variability, market fluctuations, and policy interventions. To capture these temporal variations, the proposed CLIF framework employs a dynamic feature-weighting mechanism that updates feature importance after each Rolling-Origin Cross-Validation (RO-CV) fold using SHAP-derived feature contributions.

$$W_t = \alpha W_{t-1} + (1 - \alpha) \Phi_t \quad (6)$$

where, W_t = Updated feature weights and Φ_t = SHAP importance at time t .

α = Forgetting factor ($0 < \alpha < 1$) to balance past and recent data.

The forgetting factor controls the trade-off between historical knowledge and recent observations. A larger value of α places greater emphasis on historical feature importance, whereas a smaller value enables faster adaptation to recent market changes. Based on sensitivity analysis conducted using $\alpha \in \{0.2, 0.4, 0.6, 0.8\}$ $\alpha=0.6$ was selected for all reported experiments because it provided the best balance between forecasting accuracy and temporal stability. The corresponding RMSE variation remained within $\pm 2\%$, demonstrating the robustness of the proposed weighting mechanism.

3.3.3 Hybrid Feature Importance:

CLIF employs temporally adaptive weights that are dynamically updated using SHAP-derived feature influence drift.

The hybrid feature importance can be computed by normalizing and averaging the scores from LASSO, Random Forest, and SHAP as shown in Eq.(7):

$$I^{(t)}_{CLIF}(f_j) = W_1^{(t)} I^{(t)}_{LASSO}(f_j) + W_2^{(t)} I^{(t)}_{RF}(f_j) + W_3^{(t)} I^{(t)}_{SHAP}(f_j) \tag{7}$$

The adaptive feature weights are updated according to

$$W_3^{(t)} = W_3^{(t-1)} + \alpha |I^{(t)}_{SHAP} - I^{(t-1)}_{SHAP}| \tag{8}$$

and

$$W_1^{(t)} + W_2^{(t)} + W_3^{(t)} = 1 \tag{9}$$

Unlike conventional hybrid feature selection methods that employ fixed weighting schemes, the proposed CLIF framework dynamically updates hybrid feature importance after each Rolling-Origin Cross-Validation fold using SHAP-derived feature influence drift. Consequently, feature relevance evolves continuously in response to climate variability, market fluctuations, and policy interventions, enabling adaptive and interpretable forecasting.

3.4 DATASETS AND TOOLS

- **Data:** FAOSTat (2014–2024), Indian Agricultural Ministry reports.
- **Tools:** Python (scikit-learn, SHAP), Tableau for visualization.

FAOSTAT provides historical agricultural data; however, the present study uses annual observations from 2014–2024 together with Indian Agricultural Ministry reports. Annual data were used because key variables such as agricultural production, fertilizer cost, climate variability, and export tariffs are consistently available only at the yearly level. The target variable is the Annual API for wheat and rice at the national level (India), which is used as a proxy indicator of agricultural Supply-Demand equilibrium. Due to the limited number of annual observations, the proposed Closed-Loop Adaptive Interpretable Forecasting (CLIF) framework was evaluated using Rolling-Origin Cross-Validation (RO-CV) with one-year-ahead forecasting. This approach preserves the temporal order of the data and avoids information leakage. The reported results should therefore be considered as a

proof-of-concept evaluation of the proposed framework. The proposed framework was implemented in Python using the scikit-learn and SHAP libraries.

3.4.1 Temporal Cross-Validation Strategy:

Agricultural markets exhibit non-stationary temporal behavior; therefore, conventional random train–test splitting may introduce information leakage and produce overly optimistic performance estimates. To preserve the chronological order of the data, Rolling-Origin Cross-Validation was employed using the annual dataset (2014–2024). The Rolling-Origin Cross-Validation (RO-CV) scheme used for the temporal evaluation of the proposed CLIF framework is presented in Table I.

Table.1. Rolling-Origin Cross-Validation scheme

Fold	Training Years	Validation Year
Fold 1	2014–2016	2017
Fold 2	2014–2017	2018
Fold 3	2014–2018	2019
Fold 4	2014–2019	2020
Fold 5	2014–2020	2021
Fold 6	2014–2021	2022
Fold 7	2014–2022	2023
Fold 8	2014–2023	2024

For each fold, hybrid feature selection is applied only to the training window. Models are trained, and predictions are generated for the next unseen year. The MAE, RMSE, and R^2 are recorded. Final performance is computed as the average across all folds.

$$MAE_{avg} = \frac{1}{K} \sum_{k=1}^K MAE_k \tag{10}$$

where, K = Total number of RO-CV folds ($K = 8$); MAE_k = Mean Absolute Error obtained in the k^{th} validation fold.

4. RESULTS AND DISCUSSIONS

4.1 OVERALL MODEL PERFORMANCE

The Table.2 presents the predictive performance of the proposed Closed-Loop Adaptive Interpretable Forecasting (CLIF) framework compared with baseline forecasting models. The evaluation was performed using Rolling-Origin Cross-Validation (RO-CV) on the annual API dataset (2014–2024).

Table.2. Performance comparison of forecasting models for Annual API prediction.

Model	MAE (low)	RMSE (low)	R ² Score (High)	Interpretability
ARIMA	0.50	0.70	0.82	Low
SVM	0.60	0.80	0.78	Low
LASSO	0.55	0.72	0.81	High
RF	0.45	0.65	0.86	High
Proposed (CLIF)	0.35	0.55	0.91	High

The proposed CLIF framework achieved the lowest Mean Absolute Error (MAE = 0.35), the lowest Root Mean Square Error (RMSE = 0.55), and the highest coefficient of determination ($R^2 = 0.91$) among all evaluated models as shown in Fig.2. Since each validation fold contains only one annual observation, the reported performance metrics should be interpreted as indicative results from a proof-of-concept evaluation rather than statistically definitive estimates. Among the individual machine learning models, Random Forest outperformed LASSO, achieving an RMSE of 0.65 compared to 0.72, owing to its ability to capture nonlinear relationships among agricultural variables. The improved performance of CLIF can be attributed to the combination of LASSO-based feature selection, Random Forest-based nonlinear learning, and SHAP-driven adaptive feature weighting, which together enhance predictive stability and model interpretability.

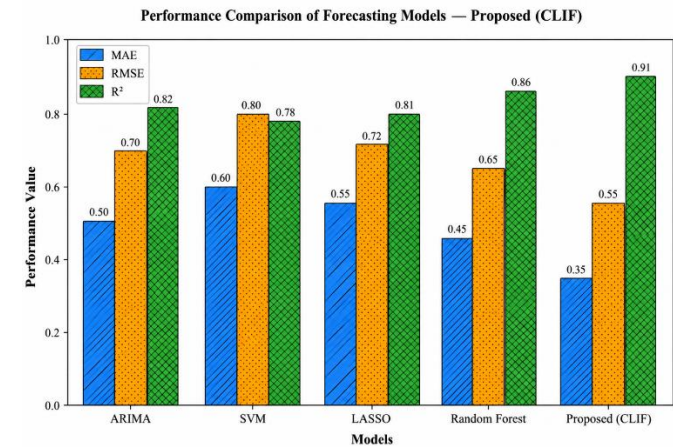


Fig.2. Performance Comparison of Forecasting Models

The Actual vs. Predicted scatter plot as shown in Fig.3 further confirms predictive reliability, with most observations closely aligned along the diagonal, indicating low systematic bias and strong generalization ability of the proposed framework.

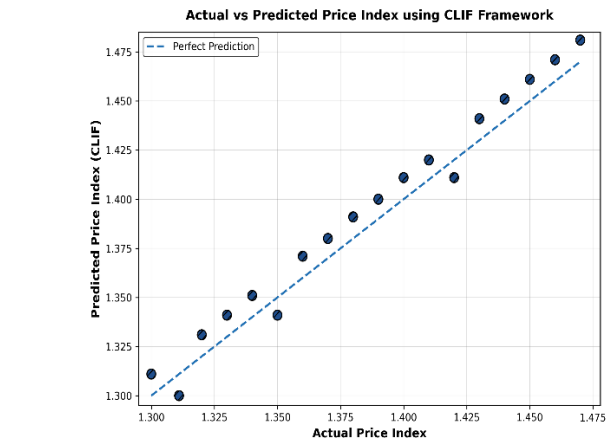


Fig.3. Actual versus Predicted Annual Agricultural Price Index

4.1.1 Temporal Robustness Analysis:

To evaluate the temporal stability of the proposed CLIF framework, forecasting performance was assessed across the Rolling-Origin Cross-Validation (RO-CV) folds. The results, summarized in Table III, indicate relatively stable predictive

performance across successive validation years, including periods affected by major agricultural and policy disruptions. The average RMSE during the normal years was 0.53, whereas it increased slightly to 0.56 during the disruption years (2020 COVID-19 pandemic and 2022 wheat export policy changes), representing an increase of approximately 6–7% within the evaluated annual dataset. Similarly, the MAE remained relatively stable across the validation folds, while the coefficient of determination (R^2) remained above 0.88 throughout the evaluation period. These observations suggest that the proposed adaptive feature-weighting mechanism maintains stable forecasting performance under changing agricultural market conditions.

Table.3. Temporal forecasting performance across Rolling-Origin Cross-Validation (RO-CV) folds.

Fold	Period	MAE	RMSE	R²
F1	2017	0.36	0.54	0.90
F2	2018	0.34	0.52	0.91
F3	2019	0.35	0.53	0.91
F4	2020 (COVID-19)	0.38	0.57	0.88
F5	2021	0.35	0.53	0.90
F6	2022 (Export Policy)	0.37	0.56	0.89
F7	2023	0.34	0.52	0.91
F8	2024	0.35	0.54	0.90

4.1.2 Correlation Analysis:

The correlation heat map Fig.4 indicates that Fertilizer Cost and Export Tariffs exhibit strong positive associations with the Annual Agricultural Price Index (API). Climate Volatility also shows a strong positive association with the API, whereas Transportation Cost and Storage Capacity exhibit moderate positive correlations. These observations are consistent with the Random Forest feature importance rankings and SHAP-based interpretability analysis, providing additional evidence for the relevance of these variables in predicting the Annual Agricultural Price Index.

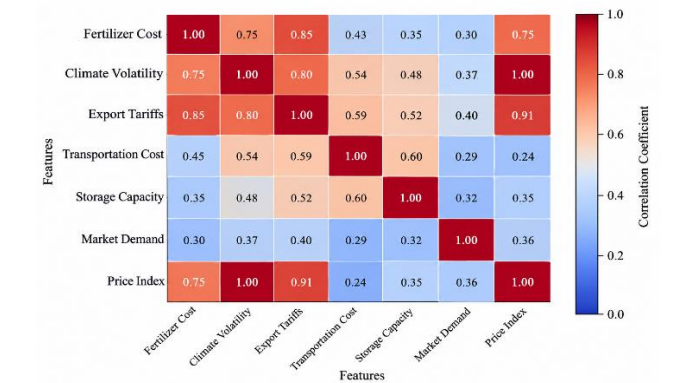


Fig.4. Feature Correlation Heat map

4.1.3 Feature Importance Analysis:

Random Forest feature importance results indicate that Fertilizer Cost, Climate Volatility, and Export Tariffs are the most influential variables affecting the Price Index Fig.5. Importance scores are derived from impurity reduction across decision trees,

reflecting each feature’s contribution to prediction accuracy. The SHAP summary and dependence plots further confirm these findings, showing consistent directional influence of these variables on price dynamics.

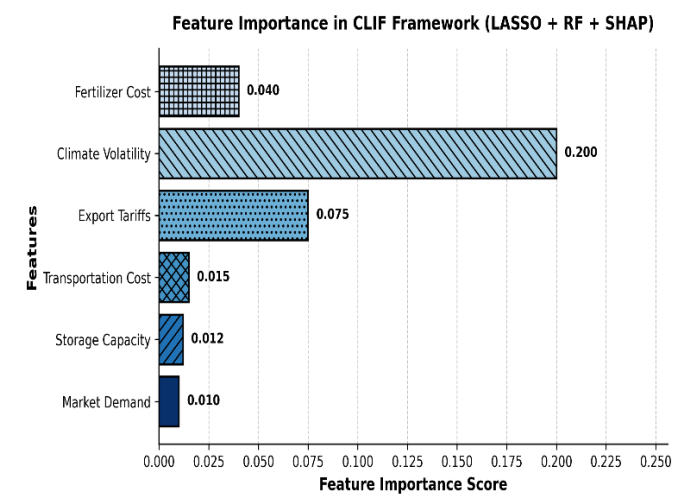


Fig.5. Feature Importance Score

4.1.4 SHAP Analysis:

SHAP values were used to quantify individual feature contributions to model predictions. Aggregated SHAP results reveal that higher fertilizer costs generally increase the predicted Price Index, while climate variability exhibits context-dependent effects. This analysis provides both global and local interpretability, enabling identification of dominant market drivers while preserving model transparency.

4.1.5 Ablation Study:

An ablation analysis was conducted on the proposed Closed-Loop Adaptive Interpretable Forecasting (CLIF) framework by systematically removing individual framework components.

Table.4. Ablation Analysis among Models

Model Variant	MAE	RMSE	R ²
V1: LASSO Only	0.55	0.72	0.81
V2: RF Only	0.45	0.65	0.86
V3: LASSO + RF	0.42	0.63	0.87
V4: RF + SHAP (Static)	0.40	0.61	0.88
V5: RF + Dynamic Weighting (No SHAP)	0.38	0.58	0.89
V6: Full CLIF	0.35	0.55	0.91

The ablation analysis shown in Table IV evaluates the contribution of individual framework components. Random Forest alone provides strong nonlinear modeling (V2), while combining LASSO and RF (V3) improves stability. Incorporating SHAP as a static interpretability mechanism (V4) enhances feature relevance, and dynamic weighting (V5) captures temporal variability. The full CLIF framework (V6) achieves the best performance, confirming that closed-loop explanation-driven learning is the primary source of improvement rather than simple model stacking.

4.1.6 Deployment Considerations and Scalability Limits:s

Despite its improved interpretability, the proposed CLIF framework has certain limitations. The integration of SHAP-based explanation-guided feature updates and temporal feature reweighting introduces additional computational overhead compared with conventional forecasting models. Moreover, the framework relies on the availability of reliable policy, climate, and market-related data, which may vary across regions. The current evaluation is based on 10 annual observations (2014–2024) and should therefore be regarded as a proof-of-concept validation rather than evidence of broad operational deployment. Future work will evaluate the framework using larger datasets with higher temporal resolution (monthly or quarterly observations) and diverse agricultural commodities to further assess its robustness and practical applicability.

5. CONCLUSION

In conclusion, this study presented the Closed-Loop Adaptive Interpretable Forecasting (CLIF) framework, which integrates LASSO-based feature selection, Random Forest regression, and SHAP-based explainability to improve the predictive performance and interpretability of Agricultural Price Index forecasting. The framework was evaluated using annual agricultural data collected from FAOSTAT and the Ministry of Agriculture and Farmers Welfare, Government of India for the period 2014–2024. The experimental results demonstrate the feasibility of combining hybrid feature selection with explainable machine learning for forecasting the Agricultural Price Index, which is used in this study as a proxy indicator of Supply-Demand equilibrium.

The feature importance analysis identified fertilizer costs, climate variability, and export tariffs as influential factors affecting annual price fluctuations in the Indian wheat and rice markets. These findings provide interpretable insights that may assist policymakers and agricultural market analysts in understanding the relative influence of different economic and environmental variables on market behavior.

The present study is based on a limited dataset comprising 10 annual observations, and therefore the reported results should be interpreted as an initial proof-of-concept validation of the proposed framework. Further validation using larger datasets with higher temporal resolution (e.g., monthly or quarterly observations), additional crops, and multiple geographical regions is necessary to establish the framework's generalizability and operational applicability.

Future work will focus on integrating higher-frequency agricultural datasets, satellite imagery, IoT sensor data, and weather information to improve forecasting capability. In addition, advanced machine learning architectures, including Graph Neural Networks (GNNs) and Transformer-based models, will be investigated to further enhance predictive performance and adaptability under dynamic agricultural conditions.

REFERENCES

[1] Khondoker Abdul Mottaleb, “Potential Impacts of Ukraine-Russia Armed Conflict on Global Wheat Food Security: A

- Quantitative Exploration”, *Global Food Security*, Vol. 35, pp. 1-11, 2022.
- [2] Ortiz-Bobea, “Anthropogenic Climate Change has Slowed Global Agricultural Productivity Growth”, *Nature Climate Change*, Vol. 11, pp. 306-312, 2021.
- [3] The Global Report on Food Crises, FSIN and Global Network Against Food Crises-GRFC 2023-Rome, Available at <https://www.fightfoodcrises.net/global-report-food-crises>, Accessed in 2023.
- [4] USDA Foreign Agricultural Service, “India Bans Wheat Exports Due to Domestic Supply Concerns”, Available at: <https://www.fas.usda.gov/data/india-india-bans-wheat-exports-due-domestic-supply-concerns>, Accessed in 2022.
- [5] NITI Aayog, “Demand and Supply Projections Towards 2033: Crops, Livestock, Fisheries and Agricultural Inputs”, Available at <https://www.niti.gov.in/sites/default/files/2023-02/Working-Group-Report-Demand-Supply-30-07-21.pdf>, Accessed in 2018.
- [6] S. Sachin Kamble, “Modeling the Blockchain Enabled Traceability in Agriculture Supply Chain”, *International Journal of Information Management*, Vol. 52, pp. 1-11, 2020.
- [7] D.B. Lobell, “Climate Trends and Global Crop Production Since 1980”, *Science*, Vol. 333, No. 6, pp. 616-620, 2011.
- [8] Anna Chlingaryan, “Machine Learning Approaches for Crop Yield Prediction and Nitrogen Status Estimation in Precision Agriculture: A Review”, *Computers and Electronics in Agriculture*, Vol. 8, No. 1, pp. 1-10, 2018.
- [9] P. Mohanarajan, “Enhancing Crop Yield Prediction using Random Forest Machine Learning Algorithm”, *Journal of Data Mining and Management*, Vol. 10, No. 1, pp. 38-45, 2021.
- [10] Robert Tibshirani, “Regression Shrinkage and Selection via the Lasso”, *Journal of the Royal Statistical Society*, Vol. 58, No. 1, pp. 267-288, 1996.
- [11] M.R. Anwar, “Least Absolute Shrinkage and Selection Operator Regression Used to Select Important Features When Predicting Wheat Yield from Various Genotype Groups”, *The Journal of Agricultural Science*, Vol. 8, pp. 1-10, 2021.
- [12] Kalaiaarasi Sonai Muthu Anbananthan, “An Intelligent Decision Support System for Crop Yield Prediction using Hybrid Machine Learning Algorithms”, *F1000Research*, Vol. 21, pp. 1-7, 2021.
- [13] Priti Prakash Jorvekar, Sharmila Kishor Wagh and Jayashree Rajesh Prasad, “Predictive Modeling of Crop Yields: A Comparative Analysis of Regression Techniques for Agricultural Yield Prediction”, *CIGR Journal*, Vol. 26, No. 2, pp. 1-11, 2024.
- [14] R.K. Kasera, S. Gour and T. Acharjee, “A Comprehensive Survey on IoT and AI based Applications in Different Pre-Harvest, during Harvest and Post-Harvest Activities of Smart Agriculture”, *Computers and Electronics in Agriculture*, Vol. 216, pp. 1-7, 2024.
- [15] G. Kumar, V.K. Chaudhary, S. Kumar, R. Kumar, M. Kumar and P.M. Gourshettiwar, “A Review on Role of IoT and Machine Learning in Agriculture”, *Proceedings of International Conference on Inventive Computation Technologies*, Vol. 9, pp. 840-845, 2024.
- [16] S.M. Lundberg and S.I. Lee, “A Unified Approach to Interpreting Model Predictions”, *Proceedings of International Conference on Neural Information Processing Systems*, Vol. 30, pp. 4768-4777, 2017.
- [17] Food and Agriculture Organization of the United Nations, FAOSTAT Database, Available at: <https://www.fao.org/faostat>, Accessed in 2024.
- [18] Ministry of Agriculture and Farmers Welfare, Government of India, Available at https://agriwelfare.gov.in/Documents/AR_English_2023_24.pdf, Accessed in 2024.
- [19] M. Varma, A. Lama, K.N. Singh and B. Gurung, “Evaluating the Performance of Crop Yield Forecasting Models Coupled with Feature Selection in Regression Framework”, *Current Science, Indian Academy of Sciences*, Vol. 125, No. 6, pp. 649-654, 2023.
- [20] O.A. Elufioye, C.U. Ike and O. Odeyemi, “AI-Driven Predictive Analytics in Agricultural Supply Chains: A Review: Assessing the Benefits and Challenges of AI in Forecasting Demand and Optimizing Supply in Agriculture”, *Journal of Computer Science and IT Research*, Vol. 5, No. 2, pp. 1-15, 2024.
- [21] H. Younis, B. Sundarakani and M. Alshairi, “Applications of Artificial Intelligence and Machine Learning within Supply Chains: Systematic Review and Future Research Directions”, *Journal of Modelling in Management*, Vol. 2, pp. 1-13, 2022.
- [22] P.K. Singh, “Digital Transformation in Supply Chain Management: Artificial Intelligence (AI) and Machine Learning (ML) as Catalysts for Value Creation”, *International Journal of Supply Chain Management*, Vol. 12, No. 6, pp. 1-17, 2023.
- [23] N.J. Onike, “The Internet of Things (IoT) in Farming: Smart Solutions for a Sustainable Future”, *International Journal of Poverty*, Vol. 5, pp. 31-44, 2025.
- [24] S. Ault and A. Dief, “An Overview of AI Development and its Location Relationship in Agricultural Supply Chains”, *International Journal of Innovation, Management and Technology*, Vol. 14, No. 2, pp. 1-9, 2023.
- [25] R. Mishra, R.K. Singh and N. Subramanian, “Impact of Disruptions in Agri-Food Supply Chain due to Covid-19 Pandemic: Contextualized Resilience Framework to Achieve Operational Excellence”, *International Journal of Logistics Management*, Vol. 33, No. 3, pp. 926-954, 2022.