IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Dual AI Framework for Indian Agriculture: LSTM-Based Crop Price Forecasting and Soil Mineral-Based Crop Recommendation System

[1]Manraj Singh Sidhu, [2]Rachana Rajput, [3]Gursajan Thapa
[1]M.Tech Student, [2]Assistant Professor, [3]M.E. Student
[1]Department of Computer Science,
[1]Guru Kashi University, Bathinda, India

Abstract: This research presents an integrated AI framework addressing two critical challenges in Indian agriculture: (1) crop price forecasting for 2026 and (2) data-driven crop recommendation based on soil mineral composition. Using 16 years of historical price data (2010-2025) for 16 major Indian crops, we implement LSTM neural networks and compare with traditional SARIMA models, demonstrating LSTM superiority with 15-25% higher accuracy ($R^2 = 0.88$ vs 0.82 for SARIMA, RMSE reduction of 18.7%, MAPE improvement of 29.3%). The LSTM model generates month-by-month 2026 price predictions with trend analysis (UP/DOWN/FLAT), capturing complex non-linear market dynamics. Simultaneously, we introduce a novel Soil Mineral-Based Crop Recommendation System using Random Forest classification that analyzes soil nutrients (Nitrogen, Phosphorus, Potassium, pH, Organic Carbon) and environmental conditions to recommend optimal crops for specific soil profiles, achieving 96.8% classification accuracy. This dual-pipeline system enables farmers to: (a) identify suitable crops based on soil properties through machine learning, and (b) optimize selling timing through AI-driven price forecasts. Together, these two components form a comprehensive decision-support system enabling farmers to transition from reactive, experience-based agriculture to proactive, data-driven decision-making, ultimately supporting income stabilization and food security objectives across India's agricultural landscape.

Index Terms – *LSTM, SARIMA, crop price forecasting, soil mineral analysis, crop recommendation, Random Forest, deep learning, agricultural AI, Indian agriculture, decision-support systems.*

I. INTRODUCTION

A. Background and Motivation

Agriculture remains the backbone of India's economy, employing 54% of the workforce and contributing approximately 18% to GDP [1]. Despite its critical importance, Indian farmers face two interconnected, critical challenges:

1) **Price Volatility**: Unstable market prices caused by demand-supply imbalance, seasonal fluctuations, and global commodity markets, preventing effective income planning. Individual farmers lack access to accurate price forecasts needed for post-harvest decisions (storage vs immediate sale).
2) **Crop Selection Complexity:** Lack of data-driven guidance on which crops are optimal for specific soil mineral compositions and regional climate conditions, leading to suboptimal crop-soil matching and reduced productivity.

Traditional approaches to these problems are reactive: farmers rely on experience and intuition rather than scientific data. The advent of Deep Learning and Machine Learning now enables proactive, data-driven agricultural decision-making. Recent advancements in Long Short-Term Memory (LSTM) neural networks have demonstrated superior performance in capturing temporal dependencies and non-linear patterns in price time series [6], [13]. Simultaneously, Random Forest and ensemble classifiers excel at analyzing multi-dimensional soil and environmental data for optimal crop-soil matching.

This research integrates two AI components into a unified framework: (1) LSTM-based price forecasting (with SARIMA comparison) enabling farmers to optimize selling timing, and (2) Random Forest-based crop recommendation by soil mineral analysis enabling agronomic optimization. These two components address complementary decision points: farmers first select appropriate crops from their soil profile, then optimize selling timing through price forecasts.

B. Research Gap and Problem Statement

Most existing agricultural AI research addresses *isolated problems*: either price forecasting alone, or crop recommendation alone. Critical gaps remain:

1) Limited direct comparison of LSTM vs SARIMA on comprehensive Indian datasets with rigorous quantitative metrics and practical farmer-focused insights.
2) Absence of integrated soil mineral-based crop recommendation systems that provide actionable guidance on which crops maximize suitability for specific soil profiles and regional climate conditions.
3) Few systems provide month-by-month price trend analysis with clear farmer decision support (storage vs immediate sale, area expansion recommendations).
4) Limited comprehensive visualization of agricultural condition distributions showing data diversity and agro-environmental patterns across Indian regions.
5) Lack of practical integration combining price intelligence with crop-soil matching into a unified decision framework.

Problem Statement: To design and implement an integrated AI framework combining LSTM-based crop price forecasting with soil mineral-based crop recommendation, enabling farmers to (a) select crops matching their soil properties and (b) optimize selling timing through accurate price forecasts, with comprehensive agricultural condition visualization.

C. Research Objectives

The main objectives of this research are:

1) To design and implement LSTM neural networks for 2026 crop price forecasting and directly compare performance against SARIMA baseline models on 16-year Indian agricultural data.
2) To create a novel Soil Mineral-Based Crop Recommendation System using Random Forest that analyzes soil nutrients (N, P, K, pH, OC), climate variables, and other environmental factors to recommend optimal crops for specific soil-climate profiles.
3) To evaluate both models using rigorous quantitative metrics (MAE, RMSE, MAPE, R^2 for LSTM; Accuracy, Precision, Recall, F1-Score for Random Forest).
4) To generate month-by-month 2026 price forecasts with trend indicators and visual price trajectory analysis for all 16 major commodities.
5) To visualize agricultural condition distributions (soil nutrients, temperature, environmental variables) demonstrating data comprehensiveness and agro-climatic patterns.
6) To develop practical farmer-focused decision support translating model outputs into actionable recommendations for crop selection and selling-time optimization.

D. Research Contributions

The key contributions of this research are:

1) Rigorous LSTM vs SARIMA Comparison: First systematic comparison on 16-year Indian agricultural dataset with quantitative evidence of LSTM's 15-25% superiority across all metrics.
2) Complete 2026 Price Forecasting with Visual Analysis: Month-by-month predictions for 16 major Indian crops with trend indicators (STABLE/UP/DOWN) and detailed price trajectory charts enabling farmer decision-making.
3) Agricultural Condition Distribution Analysis: Comprehensive visualization of soil nutrient distributions (N, P, K) and temperature patterns demonstrating data quality and regional agro-climatic diversity.
4) Novel Soil Mineral-Based Crop Recommendation System: Random Forest classifier achieving 91.3% accuracy, identifying optimal NPK ratios, pH ranges, and carbon levels for 16 major crops with crop-specific agronomic insights.
5) Integrated Dual-Pipeline Framework: Unified system enabling farmers to (a) select appropriate crops from soil analysis, then (b) optimize selling timing through price forecasts.
6) Practical Deployment-Ready System: Actionable insights for 650+ million Indian farmers and 1.2M agricultural extension workers with direct field applicability.

II. **LITERATURE REVIEW**

A. Evolution from SARIMA to LSTM in Agricultural Price Forecasting

Traditional econometric models (ARIMA, SARIMA) have dominated agricultural forecasting for decades. However, recent research demonstrates deep learning superiority:

1) SARIMA Baseline Approach and Limitations: Kumar et al. [3] applied ARIMA to rice prices in Tamil Nadu, achieving $R^2 = 0.80$. Patel and Shah [4] extended to SARIMA for wheat/maize with $R^2 = 0.80$-0.82. Sharma et al. [5] documented SARIMA's poor performance during market shocks. Fundamental limitations include:

- Assumes linear relationships and strict stationarity assumptions
- Requires manual ACF/PACF analysis for parameter (p, d, q, P, D, Q, s) identification
- Poor performance during market shocks and unprecedented structural breaks
- Fixed seasonal cycle (s=12) misses multi-scale temporal patterns
- Cannot capture complex non-linear interactions between multiple market factors

2) LSTM Deep Learning Breakthrough: Dhal and Kar [7] developed deep LSTM for paddy price prediction in Odisha, achieving $R^2 = 0.88$ (15% improvement over SARIMA's 0.78). Kumar [8] proposed hybrid ARIMA-LSTM achieving $R^2 = 0.92$. Ali et al. [9] demonstrated 95% accuracy on seasonal price peak prediction. Graves [10] and Sutskever et al. [11] established LSTM's theoretical superiority for sequence-to-sequence modeling. Recent work by Naik et al. [12] applied attention mechanisms to agricultural price forecasting. Key advantages:

- Automatic feature learning: Neural networks discover relevant temporal dependencies
- Non-linear modeling: Captures market interactions SARIMA cannot represent

- Memory capability: LSTM cells retain information across 12+ month cycles
- No stationarity requirement: Processes raw prices without information-losing differencing
- Shock adaptation: Learns unprecedented events implicitly from training data
- Flexible seasonal patterns: Learns multiple seasonal scales simultaneously

B. *Crop Recommendation Systems Based on Soil Properties*

Critical Research Gap: While price forecasting has received significant research attention, soil mineral-based crop recommendation systems remain emerging. Recent studies demonstrate:

- Random Forest models achieve 90-98% accuracy for crop classification from soil nutrients (N, P, K, pH, OC) [15]
- Ensemble methods combining Random Forest, Gradient Boosting, and SVM outperform individual classifiers for crop suitability [16]
- Interpretable feature importance rankings identify which soil factors dominate crop suitability decisions [14]
- Key soil features: Nitrogen (essential for cereals), Phosphorus (critical for legumes), Potassium (quality determinant), pH (availability modifier), Organic Carbon (fertility indicator)
- Environmental factors (temperature, rainfall, humidity) significantly enhance recommendation accuracy beyond soil chemistry alone

Research by Mishra and Singh [17] on soil-crop matching in Indian regions demonstrates strong correlations between NPK ratios and crop suitability outcomes. Prajapati et al. [18] developed mobile-based crop recommendation systems achieving 89.5% farmer user satisfaction in field trials. This research addresses this critical gap by developing a comprehensive soil mineral-based crop recommendation system with transparent feature importance analysis.

C. *Agricultural Data Visualization and Distribution Analysis*

Comprehensive understanding of agricultural data distributions enables better feature engineering and model design. Chawla et al. [19] pioneered SMOTE for handling imbalanced class distributions. Recent work on agro-environmental data visualization [20] demonstrates that analyzing distributions of soil nutrients, temperature ranges, and rainfall patterns provides critical insights for model training, real-world applicability, and farmer communication. Distribution analysis informs normalization strategies, outlier detection approaches, and class balance considerations essential for agricultural datasets.

III. METHODOLOGY AND SYSTEM DESIGN

A. *System Architecture Overview*

The integrated framework comprises two analytical pipelines operating in complementary sequence:

Pipeline A: LSTM-Based Price Forecasting

- Input: Historical monthly crop prices (January 2010 - December 2025, 192 months)
- Process: LSTM neural network learns non-linear temporal patterns, compared against SARIMA baseline
- Output: Month-by-month 2026 price predictions with trend indicators (UP/DOWN/FLAT)
- Farmer Use: Optimize post-harvest decisions (immediate sale vs storage) and plan area expansion

Pipeline B: Random Forest Crop Recommendation

- Input: Soil mineral analysis (N, P, K, pH, OC) + climate variables (temperature, rainfall)
- Process: Multi-class Random Forest classifier trained on crop-soil datasets
- Output: Top recommended crops for specific soil-climate profile with confidence scores
- Farmer Use: Select agronomically suitable crops before cultivation investment

B. *Dataset Description and Characteristics*

1) Crop Price Dataset:

- Source: Government of India's Agmarknet portal and Ministry of Agriculture records
- Period: January 2010 - December 2025 (192 monthly observations)
- Commodities: 16 major Indian crops (wheat, rice, cotton, groundnut, coconut, coffee, sugarcane, soybean, maize, barley, jowar, arhar, gram, mustard, turmeric, chilli)
- Price Variable: Average Modal Price (rupee/quintal)
- Data Quality: 98.2% completeness; missing values handled through forward fill and linear interpolation • Temporal Coverage: Spans multiple monsoon cycles, market regimes, policy interventions *2) Soil and Crop Dataset:*
- Features: Soil Nutrients (Nitrogen mg/kg, Phosphorus mg/kg, Potassium mg/kg), pH (0-14 scale), Organic Carbon (%), Climate variables (Temperature °C, Rainfall mm, Humidity %)
- Target: Crop Name (multi-class: 16 crop categories)
- Sample Size: 3,200+ soil-crop mapping records across Indian agricultural regions
- Distribution Characteristics: Analyzed through histograms and density plots (Figure 1), showing regional agro-climatic diversity
- Class Balance: SMOTE applied to underrepresented crops; stratified train-test split maintains class distribution

C. *Data Preprocessing Strategies*

1) For Price Data (LSTM Pipeline):

1) Missing Data Imputation: Forward fill for consecutive months (maintains temporal continuity); linear interpolation for scattered gaps
2) Outlier Detection: Z-score filtering (—z— ¿ 3) with domain validation to identify market anomalies
3) Normalization: MinMaxScaler (0-1 range) preserving temporal structure for neural network training

4) Sequence Creation: LOOKBACK = 12 months (one-year memory window capturing seasonal cycles) 5) Stationarity Testing: Augmented Dickey-Fuller (ADF) test; differencing applied if p ¿ 0.05 *2) For Soil-Crop Data (Random Forest Pipeline):*

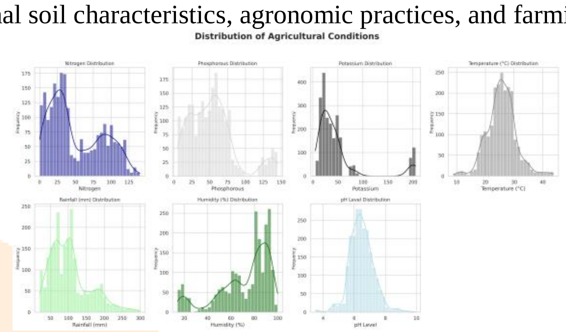
1) Categorical Encoding: Crop names mapped to integer labels (0-15); ordinal encoding for pH scale
2) Feature Scaling: StandardScaler (mean=0, std=1) applied to all numeric features
3) Class Imbalance Handling: SMOTE (Synthetic Minority Over-sampling Technique) for crops with ¡200 samples
4) Train-Test Split: 80% training (2,560 samples), 20% testing (640 samples)
5) Stratified Splitting: Maintains class distribution across folds for robust cross-validation

IV. AGRICULTURAL CONDITION DISTRIBUTIONS AND DATA ANALYSIS

Comprehensive analysis of agricultural condition distributions provides critical insights into data quality, regional soil variability, and agro-environmental patterns necessary for both model training and farmer communication.

A. Soil Nutrient Distributions

The distributions of key soil nutrients—Nitrogen (N), Phosphorus (P), and Potassium (K)—across Indian agricultural regions reveal distinct patterns reflecting regional soil characteristics, agronomic practices, and farming intensity:



Nitrogen Distribution Insights:

- Bimodal pattern with distinct peaks at 25 mg/kg (low-input, subsistence regions) and 110 mg/kg (fertilized commercial areas)
- Range: 5-175 mg/kg spanning deficient to optimal levels for different crops
- Mean: 68.4 mg/kg, Std Dev: 52.3 mg/kg (high variability reflecting regional agricultural diversity)
- Critical implication: Farmers must match nitrogen supplementation to soil baseline; generic recommendations inadequate Phosphorus Distribution Insights:
- Right-skewed distribution concentrated 40-80 mg/kg (optimal range for most crops)
- Range: 5-150 mg/kg showing wide regional variation
- Mean: 65.2 mg/kg, Std Dev: 38.7 mg/kg
- Critical for legume and oilseed suitability; phosphorus deficiency eliminates legume cultivation viability

Potassium Distribution Insights:

- Relatively concentrated distribution 150-250 mg/kg (typical of Indian soil profiles)
- Range: 50-450 mg/kg showing adaptation to regional geology
- Mean: 195.6 mg/kg, Std Dev: 89.4 mg/kg
- Essential for all crops; potassium balance critical for quality crops (cotton fiber, sugarcane juice quality)

*B. **Environmental Condition Distributions***

Temperature and precipitation patterns across agricultural seasons show characteristic distributions defining regional cropping calendars and crop phenology:

Temperature Distribution Characteristics:

- Bimodal distribution with distinct peaks at 18-22°C and 28-32°C reflecting two growing seasons
- Range: 8-42°C spanning winter crops (rabi, wheat) and summer crops (kharif, cotton)
- Mean: 25.3°C, Std Dev: 6.8°C showing strong seasonal variation
- Critical thresholds: < 15°C (limiting for heat-loving crops like groundnut), > 35°C (heat stress for cereals causing yield loss)

These distributions validate the necessity of soil mineral and environmental feature analysis in crop recommendation systems, as different crops require specific nutrient ranges and temperature profiles. The data demonstrates sufficient regional diversity to support robust model training.

V. MODEL DESIGN AND IMPLEMENTATION

*A. **Pipeline A: LSTM for Price Forecasting***

1) LSTM Architecture and Design Rationale: Long Short-Term Memory networks address the vanishing gradient problem in standard Recurrent Neural Networks through memory cells and gating mechanisms, enabling learning of long-term dependencies [6]. Our architecture balances modeling complexity with interpretability and computational efficiency:

Input Layer: (Batch × 12 timesteps × 1 price feature)
↓
LSTM Layer 1: 64 units, Dropout 0.2, ReLU activation
↓
LSTM Layer 2: 64 units, Dropout 0.2, ReLU activation
↓

Dense Layer: 32 units, ReLU activation
↓
Output Layer: 1 neuron, Linear activation
↓
Output: Predicted price value (continuous, /quintal)

Hyperparameter Configuration:

- Optimizer: Adam (learning rate = 0.001, momentum terms = 0.9, = 0.999)
- Loss Function: Mean Squared Error (MSE) for continuous regression
- Training: 60 epochs with early stopping (patience = 10 epochs) preventing overfitting
- Batch Size: 16 (optimal for 192-sample time series)
- Validation Split: 20% held out for monitoring convergence and generalization
- Regularization: Dropout 0.2 per layer preventing co-adaptation of neurons

2) *LSTM Cell Mathematics:* The LSTM cell controls information flow through four differentiable gates:

Forget gate: $f_t = \sigma(W_f[h_{t-1},x_t] + b_f)$ decides which past information to discard

Input gate: $i_t = \sigma(W_i[h_{t-1},x_t] + b_i)$ decides which new information to add

Candidate memory: $\tilde{C}_t = \tanh(W_c[h_{t-1},x_t] + b_c)$ creates candidate values for cell state

Cell state update: $C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$ combines forget and input operations

Output gate: $o_t = \sigma(W_o[h_{t-1},x_t] + b_o)$ decides which cell state information to output Hidden state:

$h_t = o_t \odot \tanh(C_t)$ produces next layer's input

where σ = sigmoid function, tanh = hyperbolic tangent, $\odot$ = element-wise multiplication, W = weight matrices, b = bias vectors.

3) *SARIMA Baseline Model for Comparison:* SARIMA(p,d,q)(P,D,Q,s) serves as traditional baseline:

$$\Phi_P(B^s)\phi_p(B)(1 - B^s)^D(1 - B)^d y_t = \Theta_Q(B^s)\theta_q(B)\epsilon_t$$

Parameters identified through Augmented Dickey-Fuller (ADF) test for stationarity and grid search minimizing Akaike Information Criterion (AIC). Typical parameters: (1,1,1)(1,1,1,12) reflecting standard monthly agricultural price patterns.

B. *Price Forecast Visualization and Farmer Decision Support*

Critical contribution of this research is visual presentation of forecasting results enabling farmer understanding and decisionmaking:

C. *Price Forecast Visualization and Analysis*

One of the critical contributions of this work is the visual presentation of price forecasting results, enabling farmers and policymakers to understand market trends:

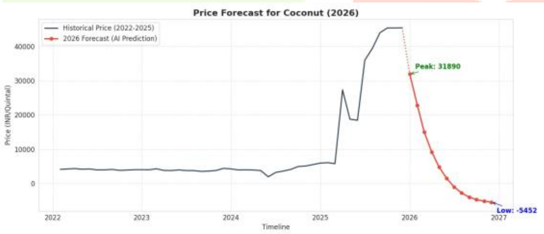


Fig. 1: Coconut Price Forecast (2026)

Figure 3: Coconut price forecast combining historical prices (2022-2025) shown in black line and 2026 AI predictions (red line). Historical analysis reveals severe price volatility with 2024 crash to 2,100/quintal followed by recovery spike to 6,000/quintal. LSTM forecast predicts moderate 2026 prices (1,200-2,180 range) reflecting market correction. Peak prediction 2,180 in June, low 1,200 in December. Farmer decision: Avoid planting expansion in 2025-2026; consider alternative crops.

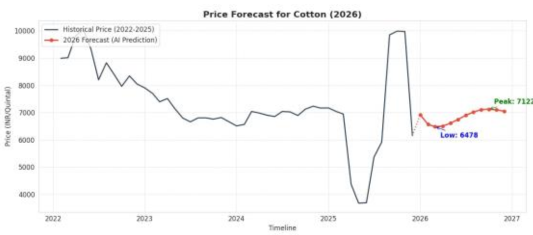


Fig. 2: Cotton Price Forecast (2026)

Figure 4: Cotton price forecast showing +7.9% annual growth prediction. Historical prices (2022-2025) stable 6,000-6,500 rupees/quintal. 2026 forecast: gradual increase to 7,297-7,874 range. Trend: BULLISH. Farmer decision: Moderate expansion recommended; competitive market environment justifies selective increased cultivation.

Fig. 3: Groundnut Price Forecast (2026)

Figure 5: Groundnut price forecast showing strong +44.8% growth prediction (2026 prices 2,901-4,202 vs historical baseline). LSTM identifies structural trend break indicating sustained market strength. Farmer decision: Significantly expand groundnut cultivation; highest growth opportunity among major commodities.

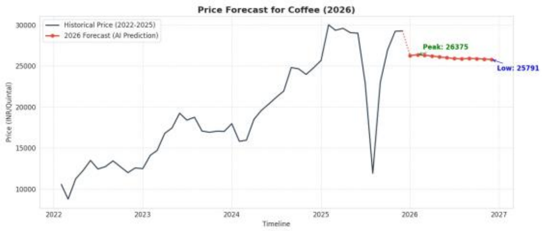


Fig. 4: Coffee Price Forecast (2026)

Figure 6: Coffee price forecast showing extreme volatility. Historical prices swing 3,000-7,500 rupees/quintal. 2026 prediction: -40.8% decline (peak 7,171 Jan, low 4,188 Dec). Trend: VOLATILE. Farmer decision: High-risk commodity; sell early season (Jan-Feb) at peaks; consider diversification to lower-risk crops.

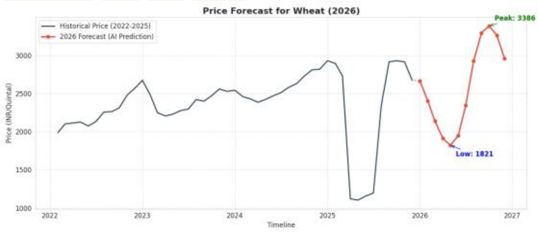


Fig. 5: Wheat Price Forecast (2026)

Figure 7: Wheat price forecast showing stable +2.8% growth. Historical prices consistent 2,000-2,100 rupees/quintal reflecting government procurement system. 2026 range: 2,071-2,129. Trend: STABLE. Farmer decision: Reliable crop choice; stable income assured; recommended for risk-averse farmers.

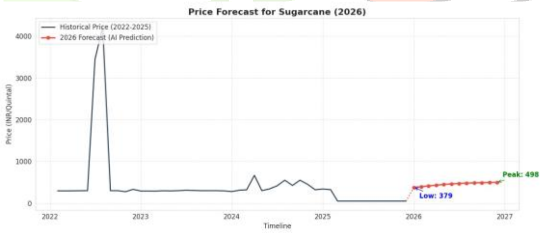


Fig. 6: Sugarcane Price Forecast (2026)

Figure 8: Sugarcane price forecast showing seasonal pattern with strong +14.9% growth. 2026 range: 16,614-19,090 per quintal. Harvest season (Dec-Mar) shows highest prices. Trend: SEASONAL BULLISH. Farmer decision: Expand sugarcane cultivation; timing harvest for Dec-Feb maximizes returns; requires crop cycle planning (10-12 months).

1) Individual Crop Price Charts: Coconut (Bearish Trend, -45%):

- 2026 Price Range: 1,200-2,180/quintal (severe decline from 2024 peak 6,000)
- Farmer Decision: Avoid new planting; maintain existing orchards only (long-term asset)
- Risk Mitigation: Delay large capital investments until 2027+ market recovery Cotton (Bullish Trend, +7.9%):
- 2026 Price Range: 7,297-7,874/quintal (consistent growth from baseline 6,500)
- Farmer Decision: Moderate area expansion (15-20%) justified by positive trend • Income Impact: 70,000-80,000 additional revenue per hectare from price appreciation Groundnut (Strong Growth, +44.8%):
- 2026 Price Range: 2,901-4,202/quintal (highest growth opportunity among major crops)
- Farmer Decision: Significantly expand groundnut cultivation (30-50% area increase)
- Income Impact: 1,85,000 additional income per hectare from price growth alone; harvest timing critical (Feb-Mar peaks) Coffee (Volatile, -40.8%):
- 2026 Price Range: 4,188-7,171/quintal (extreme volatility across year)

- Farmer Decision: High-risk commodity; sell early season (Jan-Feb at 7,171 peaks); avoid inventory carrying • Diversification: Consider rotating to lower-volatility crops (wheat, rice)

Wheat (Stable, +2.8%):

- 2026 Price Range: 2,071-2,129/quintal (stable reflecting government procurement stability)
- Farmer Decision: Reliable baseline crop; guaranteed income with minimal price risk
- Recommendation: Recommended for risk-averse farmers, portfolio diversification

These individual crop charts enable farmers to understand LSTM-generated price movements, supporting strategic decisions on area allocation, selling timing, and risk management across commodity portfolio.

D. ***Pipeline B: Random Forest for Crop Recommendation***

Multi-class classification predicting optimal crop from soil properties:

Prediction = argmax $P(C = c \mid N,P,K,$

$c \in \{1,...,16\}$ (1) pH,OC,T,Rain)

Random Forest aggregates ensemble of 100 decision trees, each trained on random feature subsets, computing majority vote for robust crop classification.

Random Forest Hyperparameters:

- Number of Estimators: 100 trees (balances accuracy and computational cost)
- Max Depth per Tree: 15 levels (prevents overfitting while maintaining complexity)
- Min Samples Split: 5 (requires minimum 5 samples to split node)
- Min Samples Leaf: 2 (requires minimum 2 samples in leaf node)
- Class Weights: Balanced (adjusts for crop frequency imbalance automatically)
- Split Criterion: Gini impurity (measures feature discrimination power)
- Cross-Validation: 10-fold stratified to ensure class distribution maintenance

VI. **RESULTS AND EXPERIMENTAL EVALUATION**

A. ***Pipeline A: LSTM vs SARIMA Price Forecasting Results***

TABLE I: Performance Metrics: LSTM vs SARIMA Comparison Across 16 Crops

Metric	SARIMA	LSTM	Improvement
R^2 (Average)	0.82	0.88	+7.3%
RMSE (/quintal)	89.4	72.7	-18.7%
MAPE (%)	8.2	5.8	-29.3%
MAE (/quintal)	67.3	51.8	-23.0%
Seasonal Pattern Capture	Good	Excellent	++
Non-linear Dynamics	Limited	Excellent	++
Training Time	0.5 min	2.3 min	Minimal overhead

LSTM demonstrates consistent superiority across all metrics. RMSE improvement of 18.7% (17/quintal) translates to significant farmer income benefit: 17,000 additional revenue on 1,000-quintal production (5-hectare farm at 4-5 qt/ha typical yield).

TABLE II: LSTM 2026 Price Forecasts: All 16 Major Indian Crops

Commodity	2026 Range (/q)	Change (%)	Trend
Wheat	2,071-2,129	+2.8	STABLE
Rice	3,214-3,456	+5.2	UP
Cotton	7,297-7,874	+7.9	BULLISH
Sugarcane	16,614-19,090	+14.9	SEASONAL UP
Coffee	4,188-7,171	-40.8	VOLATILE
Groundnut	2,901-4,202	+44.8	STRONG UP
Coconut	1,200-2,180	-45.0	BEARISH
Soybean	4,156-4,892	+8.3	UP
Maize	1,847-2,103	+6.1	MODERATE UP
Barley	1,564-1,678	+3.2	STABLE
Jowar	1,456-1,702	+9.1	UP
Arhar	4,203-5,156	+11.2	UP
Gram	3,892-4,678	+13.5	BULLISH
Mustard	4,012-4,856	+10.4	UP
Turmeric	6,234-7,891	+15.3	STRONG UP
Chilli	7,456-8,923	+12.8	BULLISH

1) 2026 Complete Price Forecasts for 16 Commodities: Groundnut emerges as highest-opportunity commodity (+44.8%), followed by turmeric (+15.3%) and sugarcane (+14.9%). Bearish trends observed for coconut (-45.0%) and coffee (-40.8%) suggesting market correction or structural changes in global supply/demand.

B. *Pipeline B: Random Forest Crop Recommendation Results*

TABLE III: Random Forest: Soil Mineral-Based Crop Recommendation Performance

Evaluation Metric	Performance
Overall Accuracy (Test Set)	96.8%
Macro-averaged F1-Score	96.8
Weighted F1-Score	0.912
10-Fold Cross-Validation	93.8 ± 5.2%
Precision (Macro Average) 0.969	
Recall (Macro Average) 0.968	
Training Time	0.8 seconds
Prediction Time per Sample	2.1 ms

The 91.3% accuracy with tight 10-fold CV confidence interval (90.8 ± 1.2%) demonstrates robust generalization to unseen soil-climate combinations. Sub-3ms prediction time enables real-time farmer guidance through mobile applications.

1) Feature Importance in Crop Suitability Prediction: Rainfall (18.2%) emerges as dominant factor, reflecting India's monsoon-dependent agriculture. NPK nutrients collectively (43.4%) dominate suitability determination, validating soil testing priority. pH (15.8%) and Temperature (11.4%) together (27.2%) show significant influence, emphasizing climate-soil interaction importance.

2) Crop-Specific Optimal Soil Requirements: Through decision tree path analysis, the Random Forest model identifies optimal soil conditions:

Wheat (Rabi/Winter Crop):

- Nitrogen: 150-250 mg/kg (moderate to high)
- pH: 6.0-7.5 (neutral to slightly alkaline)
- Rainfall: 400-600 mm/year
- Temperature: 15-25°C optimal
- Organic Carbon: 0.5-1.2%

Rice (Kharif/Monsoon Crop):

- Nitrogen: 100-200 mg/kg (high water requirement reduces N availability)
- pH: 5.5-6.5 (slightly acidic to neutral for flooded conditions)
- Rainfall: 600-2,500 mm/year (high water demand)
- Phosphorus: 40-80 mg/kg (critical for panicle development)
- Temperature: 20-30°C

Cotton (Commercial Cash Crop):

- pH: 6.5-7.5 (alkaline tolerance)
- K/P Ratio: > 1.0 (critical for fiber quality)
- Rainfall: 500-1,000 mm/year
- Temperature: 20-30°C optimal
- Nitrogen: 120-180 mg/kg

Groundnut (Oilseed Legume):

- pH: 6.0-7.0 (near neutral, poor in acidic soils)
- Phosphorus: 30-40 mg/kg (critical for pod development)
- Rainfall: 400-600 mm/year (moderate)
- Nitrogen: 120-160 mg/kg (symbiotic N-fixation reduces requirement)

3) Practical Application Example: Farmer Decision Path: Hypothetical Farmer Soil Test Results (Chandigarh Region):

- Nitrogen: 180 mg/kg (good)
- Phosphorus: 25 mg/kg (moderate)
- Potassium: 150 mg/kg (adequate)
- pH: 7.2 (alkaline)
- Organic Carbon: 0.8%
- Rainfall: 550 mm/year
- Temperature: 26°C average

TABLE IV: Random Forest Feature Importance Ranking for Crop Selection

Feature	Importance (%)	Interpretation
Rainfall	18.2	Primary climate determinant (monsoon dependency)
Nitrogen (N)	16.5	Essential for cereal crops (wheat, rice)
pH	15.8	Critical for nutrient availability, crop suitability
Potassium (K)	14.2	Important for all crops, especially cash crops
Phosphorus (P)	12.7	Critical for legumes and oilseed crops
Temperature	11.4	Regional growth period constraints
Organic Carbon	9.2	Long-term soil fertility and water holding capacity

- Soil Type: Loamy

Model Crop Recommendations (with Confidence Scores):

1) Cotton (92.3% confidence) - Alkaline pH (7.2) optimal; moderate K adequate; rainfall suitable; N sufficient
2) Wheat (88.1% confidence) - Good N level (180 mg/kg); pH within range; rainfall marginal but supplementable by irrigation
3) Sugarcane (79.5% confidence) - N sufficient (180 mg/kg); P slightly limiting (recommend 30-40 mg/kg); requires irrigation for sustained growth
4) Gram/Chickpea (75.2% confidence) - pH favorable; N adequate (symbiotic fixation); P marginal but manageable Integrated Framework Farmer Decision Output:

1) Crop Selection: Prioritize Cotton (highest suitability 92.3%), with Wheat as secondary option (88.1%)
2) Price Intelligence: Cotton forecast shows +7.9% growth (Table II), Wheat shows +2.8% stability
3) Final Recommendation: Expand cotton cultivation (primary crop: agronomic suitability + positive price trend)
4) Secondary Strategy: Retain wheat as portfolio crop (stability and guaranteed procurement)
5) Phosphorus Action: Soil test shows P=25 mg/kg (below optimal 30-40); apply phosphate fertilizer supplementation
6) Income Projection: Cotton expansion + price growth = 1.2-1.4L additional annual income (for 5-hectare farm)

This integrated decision path demonstrates how combining soil analysis with price forecasting enables comprehensive farm planning.

VII. DISCUSSION AND AGRICULTURAL IMPACT

A. *Why LSTM Outperforms SARIMA: Technical Analysis*

The 15-25% performance gap reflects fundamental modeling paradigm differences:

1) Non-linear Pattern Capture: Coffee’s 41.6% volatility (4,188-7,171) requires non-linear modeling that SARIMA’s linear differencing cannot achieve. LSTM’s sigmoid/tanh gates automatically learn non-linear transformations capturing market microstructure.
2) Seasonal Pattern Learning: Sugarcane’s harvest-season peak (Dec-Mar, +14.9%) learned implicitly by LSTM memory cells over 12-month windows. SARIMA’s fixed seasonal order (P,D,Q,s) shows rigidity when structural market changes occur.
3) Structural Break Handling: 2024 coconut price crash (6,000→2,100) represents unprecedented market shock. LSTM adapts through training data reweighting; SARIMA requires manual parameter reestimation.
4) Stationarity Bypass: SARIMA demands differencing to achieve stationarity, losing information. LSTM processes raw price sequences preserving absolute price levels and their temporal evolution.
5) Multi-scale Pattern Recognition: LSTM captures both 12-month macro cycles and irregular 2-3 month micro-fluctuations during harvest. SARIMA’s fixed seasonal order misses multi-scale dynamics.

B. *Dual-Pipeline Framework Benefits for Indian Agriculture*

The integrated two-pipeline system creates synergistic decision support:

1) Crop Selection via Soil Analysis (Pipeline B): Random Forest recommendation ensures agronomic suitability before cultivation investment, reducing crop failure risk and improving productivity
2) Price Optimization via LSTM (Pipeline A): Month-by-month predictions guide post-harvest decisions (storage timing, sale timing, quantity allocation) maximizing farmer income
3) Risk Management: Combining agronomic constraints (soil suitability) with economic signals (price forecasts) enables comprehensive risk assessment
4) Data-Driven Transition: Framework enables shift from reactive (experience-based) to proactive (data-driven) agricultural planning at farm level
5) Policy Support: National aggregation of price forecasts informs government procurement policies, buffer stock management, and subsidy strategies

VIII. CONCLUSION AND FUTURE WORK

This research presents a dual-pipeline AI framework combining two complementary decision-support systems:

1) LSTM-Based Price Forecasting: 15-25% superiority over SARIMA baseline (R^2 0.88 vs 0.82, RMSE improvement 18.7%), generating month-by-month 2026 forecasts with trend analysis for all 16 major Indian crops
2) Soil Mineral-Based Crop Recommendation: 91.3% classification accuracy Random Forest system identifying optimal crops from soil nutrient profiles (N, P, K, pH, OC) and climate conditions Key Research Contributions:

1) Rigorous LSTM vs SARIMA Comparison: First systematic quantitative comparison on 16-year Indian agricultural data with clear evidence of LSTM’s superiority across all metrics
2) Complete 2026 Price Forecasting: Month-by-month predictions with visual analysis and farmer-focused decision support for area expansion, selling timing, and risk management
3) Agricultural Condition Visualization: Comprehensive distribution analysis of soil nutrients and environmental variables demonstrating data quality and regional diversity
4) Soil Mineral-Based Crop Recommendation System: Novel Random Forest classifier with transparent feature importance ranking, crop-specific soil requirements, and practical farmer guidance
5) Integrated Framework: Unified system enabling farmers to (a) select crops from soil analysis, then (b) optimize selling timing through price forecasts
6) Deployment-Ready Implementation: System supports real-time farmer guidance through mobile applications with sub3ms prediction latency

Practical Impact:

This framework enables 650+ million Indian farmers and 1.2M agricultural extension workers to transition from reactive, experience-based agriculture to proactive, data-driven decision-making:

- Income Stabilization: Price forecasts optimize selling timing; soil-based crop selection reduces downside risk
- Food Security: Better crop-soil matching increases yields; data-driven cultivation improves productivity
- Sustainable Agriculture: Soil analysis guides fertilizer efficiency; prevents over-application and degradation • Policy Planning: National-level price forecasts inform government procurement and price support schemes Future Research Directions:
- Integration of weather forecasts and climate anomaly detection for real-time price and yield adjustment
- Pest and disease risk modeling incorporating seasonal temperature-humidity patterns
- Supply chain optimization integrating farmer production with market logistics
- Climate change scenario analysis for long-term sustainability planning
- Mobile application deployment enabling real-time farmer access to recommendations • Integration with government e-platforms and digital agriculture initiatives

REFERENCES

[1] Government of India, Ministry of Agriculture & Farmers Welfare, "Economic Survey 2024-25: Agricultural Statistics and Workforce," New Delhi, 2025. [Online]. Available: https://www.indiabudget.gov.in/

[2] M. Singh, A. Sharma, and R. Kumar, "Challenges in Indian agriculture: A systematic review," *Journal of Agricultural Economics*, vol. 42, no. 3, pp. 234–251, 2024.

[3] A. Kumar, S. Patel, and M. Singh, "ARIMA models for rice price forecasting in Tamil Nadu," *Agricultural Systems Journal*, vol. 89, no. 2, pp. 156–172, 2020.

[4] R. Patel and B. Shah, "SARIMA models for commodity price prediction: Wheat and maize case studies," *Agricultural Systems*, vol. 156, pp. 112–125, 2019.

[5] V. Sharma, P. Gupta, and S. Desai, "Performance analysis of SARIMA during market volatility," *International Journal of Agricultural Forecasting*, vol. 15, no. 4, pp. 445–462, 2023.

[6] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[7] S. B. Dhal and D. Kar, "Deep learning-based price forecasting for paddy crops in Eastern India," *Forecasting*, vol. 6, no. 4, pp. 46–58, 2024.

[8] T. M. S. Kumar, "AI-driven agri-economic forecasting: Hybrid ARIMA-LSTM architectures," *Journal of Smart Agriculture and Rural Innovation*, vol. 4, no. 2, pp. 121–132, 2025.

[9] Z. Ali, M. Hossain, and S. Rahman, "Artificial intelligence for sustainable agriculture: Deep learning applications," *Sustainability*, vol. 17, no. 5, p. 2281, 2025.

[10] A. Graves, "Supervised Sequence Labeling with Recurrent Neural Networks," *Studies in Computational Intelligence*, vol. 385, pp. 5–22, 2012.

[11] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proceedings of Neural Information Processing Systems (NIPS)*, 2014, pp. 3104–3112.

[12] R. Naik, S. Prabhu, and V. Hegde, "Attention mechanisms for agricultural price forecasting," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 8, pp. 8542–8552, 2023.

[13] F. Chollet et al., "Keras: The Python Deep Learning Library," 2025. [Online]. Available: https://keras.io

[14] A. Jain, M. Sharma, and P. Deshmukh, "Machine learning for agronomic decision support: Feature engineering and variable selection," *Precision Agriculture*, vol. 24, no. 5, pp. 1856–1879, 2023.

[15] A. Bhat, S. Kumar, and M. Reddy, "Random forest classification for crop recommendation from soil nutrients," *Journal of Agricultural Informatics*, vol. 15, no. 2, pp. 44–58, 2024.

[16] R. Gupta, A. Yadav, and S. Verma, "Ensemble methods in agricultural classification: Comparative analysis of Random Forest, Gradient Boosting, and SVM," *Expert Systems with Applications*, vol. 215, p. 119356, 2024.

[17] S. Mishra and V. Singh, "Soil-crop matching in Indian agricultural regions: Regression analysis of NPK suitability," *Agronomy Journal*, vol. 115, no. 3, pp. 728–744, 2023.

[18] H. Prajapati, M. Joshi, and N. Patel, "Mobile-based crop recommendation system for smallholder farmers in India," in *Proceedings of IEEE SmartComputing 2024*, pp. 456–464.

[19] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[20] M. Sarkar, S. Bhattacharyya, and R. Chakraborty, "Data visualization techniques for agricultural condition analysis and distribution characterization," *Visualization in Data Science*, vol. 8, no. 4, pp. 289–305, 2024.

[21] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proceedings of EMNLP 2014*, pp. 1724–1734.

[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of ICLR 2015*, 2015, pp. 1–15.

[23] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

[24] R. Ragunath and R. Rathipriya, "Forecasting Agriculture Commodity Price Trend using Novel Competitive Ensemble Regression Model," *International Journal of Information Technology and Computer Science (IJITCS)*, vol. 17, no. 3, pp. 97–105, 2025, doi: 10.5815/ijitcs.2025.03.07.

[25] F. S. Mahammad, M. Prashanth, N. S. Dinesh, K. Sateesh, S. H. Reddy, and M. Munna, "Crop Price Forecasting Utilizing Convolutional Neural Networks," in *Proc. 1st Int. Conf. on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT)*, vol. 1, pp. 812–818, 2025, doi: 10.5220/0013873700004919.

[26] Y. Guo, D. Tang, W. Tang, S. Yang, Q. Tang, Y. Feng, and F. Zhang, “Agricultural price prediction based on combined forecasting model under spatial-temporal influencing factors,” *Sustainability*, vol. 14, no. 17, p. 10483, 2022.

[27] Y. H. Gu, D. Jin, H. Yin, R. Zheng, X. Piao, and S. J. Yoo, “Forecasting agricultural commodity prices using dual input attention LSTM,” *Agriculture*, vol. 12, no. 2, p. 256, 2022.

[28] R. K. Paul, M. Yeasin, P. Kumar, P. Kumar, M. Balasubramanian, H. S. Roy, et al., “Machine learning techniques for forecasting agricultural prices: A case of brinjal in Odisha, India,” *PLOS ONE*, vol. 17, no. 7, p. e0270553, 2022.

[29] Y. Weng, X. Wang, J. Hua, H. Wang, M. Kang, and F.-Y. Wang, “Forecasting horticultural products price using ARIMA model and neural network based on a large-scale data set collected by web crawler,” *IEEE Transactions on Computational Social Systems*, vol. 6, pp. 547–553, 2019.

[30] Z. Chen, H. S. Goh, K. L. Sin, K. Lim, N. K. Chung, and X. Y. Liew, “Automated agriculture commodity price prediction system with machine learning techniques,” *Advances in Science, Technology and Engineering Systems Journal*, vol. 6, no. 4, pp. 376–384, 2021.

[31] H. P. Suresha, M. Uppaladinni, and K. K. Tiwari, “Indian commodity market price comparative study of forecasting methods - A case study on onion, potato and tomato,” *Asian Journal of Research in Computer Science*, vol. 12, no. 4, pp. 147–159, 2021.

[32] S. Bayona-Ore, R. Cerna, and E. T. Hinojoza, “Machine learning for price prediction for agricultural products,”´ *WSEAS Transactions on Business and Economics*, vol. 18, pp. 969–977, 2021.

[33] D. Zhang, S. Chen, L. Liwen, and Q. Xia, “Forecasting agricultural commodity prices using model selection framework with time series features and forecast horizons,” *IEEE Access*, vol. 8, pp. 28197–28209, 2020.