Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

A Data-Driven System for Crop Switching Risk and Yield Stability in Climate Change

An Integrated Deep Learning and Gradient-Boosting Architecture for Quantitative Agricultural Transition Risk Scoring with Real-Time Satellite Intelligence

Dr. J. Nafeesa Begum • Aashika L • Swaranthi B • Varshitha G V
Department of Computer Science and Engineering,
Government College of Engineering, Bargur – 635104, Krishnagiri, Tamil Nadu, India
Anna University, Chennai – 600 025

Abstract - Intensifying climate volatility poses unprecedented threats to traditional Indian agriculture, disrupting seasonal crop calendars and rendering historical planting strategies unreliable. This paper presents the system - a fully automated, web-accessible, data-driven decision-support platform that quantifies the risk associated with transitioning from one crop to another under climate uncertainty. The system integrates four synergistic analytical layers: (1) a Large-Scale Historical Ingestion Engine processing 50,765 agricultural records spanning 1966–2017 across 20 Indian states; (2) a Feature Signal Synthesis module that engineers the Climate Stress Index (CSI), Crop-Climate Interaction terms, and 5-year rolling Yield Volatility metrics; (3) a Hybrid Neural-Gradient Predictive Engine combining Long Short-Term Memory (LSTM) temporal sequence extraction with Extreme Gradient Boosting (XGBoost) tabular regression, achieving a benchmark coefficient of determination (R^2) of 88.16% and RMSE of approximately 281 kg/ha; and (4) a Live Intelligence Layer powered by the Open-Meteo Satellite API and Nominatim village-level geocoding, enabling dynamic, context-aware risk score adjustment. The system generates a calibrated 0–500+ Risk Scoring Engine that classifies every crop-switching scenario into LOW, MEDIUM, or HIGH risk categories based on percentile-derived thresholds. Deployed as an interactive Streamlit analytical dashboard with eight specialised pages, the system delivers localised, village-level pilot transition plans to smallholder farmers. Experimental validation confirms that LSTM temporal embeddings contribute a 6.4% absolute R^2 improvement over the standalone XGBoost baseline. the system positions itself as a scalable, equitable technological shield for agricultural communities navigating climate-driven crop diversification decisions.

Keywords: ***Crop Switching Risk; Climate Change Adaptation; LSTM; XGBoost; Hybrid Ensemble; Yield Stability; Agricultural Decision Support; Risk Scoring Engine; Satellite Weather API; Precision Agriculture; Climate Stress Index; Temporal Embeddings; Village-Level Geocoding***

1. INTRODUCTION

Agriculture underpins the food security and economic livelihood of over 40% of India's population, yet it is among the sectors most exposed to accelerating climate disruption. The increasing frequency of extreme weather events - prolonged droughts, erratic monsoon patterns, unseasonal floods, and rising mean temperatures - has invalidated centuries of ancestral cropping wisdom and regional agricultural calendars. Smallholder farmers, who constitute the majority of India's agricultural workforce, face a stark choice: continue with traditional crops that yield diminishing returns under shifting climatic baselines, or transition to more resilient alternatives at the risk of catastrophic yield failure.

This challenge - the decision to "switch crops" - is one of the highest-stakes economic choices a farming household can make. Unlike large commercial farms with access to meteorological departments and precision agriculture tools, smallholder farmers in Tier-2 and Tier-3 regions lack quantitative, data-backed frameworks for evaluating transition risk. The result is a critical "Intelligence Gap" that perpetuates economic vulnerability across rural India.

Existing advisory systems suffer from three structural deficiencies: (1) they are temporally blind, analysing isolated seasonal snapshots without capturing decadal climate memory; (2) they are binary in logic, recommending a "best crop" without quantifying the volatility inherent in switching; and (3) they are real-time dissociated, unable to incorporate current satellite weather anomalies into historical risk baselines.

This paper presents the system - a comprehensive, hybrid machine learning platform that addresses these deficiencies through a four-layer architecture spanning data ingestion, feature engineering, neural-gradient ensemble prediction, and live satellite intelligence. The principal contributions of this work are:

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

- A unified crop switching risk pipeline integrating LSTM temporal sequence learning, XGBoost tabular regression, and real-time satellite weather telemetry into a single decision-support environment.
- A 50,765-record ICRISAT historical dataset engineered into predictive agricultural signals, including the Climate Stress Index (CSI), Yield Volatility Markers, and Crop-Climate Interaction terms.
- A calibrated 0–500+ Risk Scoring Engine using percentile-based thresholds to classify switching scenarios into actionable LOW, MEDIUM, and HIGH risk categories.
- Empirical validation demonstrating 88.16% R^2 accuracy, with LSTM embeddings contributing a 6.4% absolute improvement over the standalone XGBoost baseline.
- A fully deployed Streamlit interactive dashboard providing village-level geocoding resolution, personalised crop transition pilot plans, and real-time environmental safety overrides.

2. BACKGROUND AND MOTIVATION

2.1 The Agricultural Intelligence Gap

Agriculture remains the foundational pillar of the global economy, providing sustenance for billions and serving as the primary source of livelihood in developing nations. In India, smallholder farms account for over 86% of all agricultural holdings and contribute approximately 50% of total crop output. However, this segment is the most exposed to climate variability and the least equipped with tools to navigate it.

The motivation behind the system stems from the growing disparity in agricultural intelligence access. While commercial agribusinesses deploy sophisticated meteorological modelling and precision analytics, the rural smallholder relies on ancestral patterns that are rapidly becoming obsolete. A region historically suited for water-intensive Rice cultivation may now experience chronic groundwater depletion and erratic monsoon patterns, making a transition to drought-resilient Maize not merely advantageous but economically necessary for survival.

Yet "Crop Switching" is a profoundly high-risk decision. A misguided transition can lead to total crop failure, devastating household finances for multiple consecutive years. There is currently an absence of quantitative, multi-decadal, and real-time validated frameworks that can provide a farmer with a scientifically backed risk assessment such as: "Based on 50 years of historical data and today's live satellite telemetry, switching from Rice to Maize in your specific village carries a calculated risk score of X." Bridging this gap through advanced Machine Learning is the primary driver of this research.

2.2 Why Existing Systems Fall Short

Current agricultural advisory systems in India primarily operate across two fragmented categories:

- Static Soil-Based Recommendations: Government portals such as the Soil Health Card (SHC) scheme provide fertilisation advice based on laboratory-tested NPK levels. These systems are climate-static, ignoring intra-seasonal weather volatility and long-term climatic trend shifts.
- Generalised Regional Advisories: Meteorological agencies provide coarse-grained district-level rainfall forecasts. These advisories lack integration with crop-specific stress thresholds, historical yield trajectories, or personalised farmer data.

Neither category provides transition logic - a quantitative evaluation of what it means to move from one crop to another in a specific micro-climate under today's actual environmental conditions. the system directly addresses this gap.

2.3 Technological Enablers

Three converging technological advances make the the system architecture viable. First, deep learning architectures - specifically Long Short-Term Memory (LSTM) networks - enable the system to extract decadal temporal patterns from sequential crop-climate records with a precision unachievable by classical regression. Second, gradient-boosted decision trees (XGBoost) provide the high-capacity tabular processing required to integrate heterogeneous soil, climate, and agronomic features into precise yield forecasts. Third, public satellite APIs (Open-Meteo) and geolocation services (Nominatim) deliver real-time environmental telemetry at village-level resolution, enabling dynamic risk calibration against current field conditions.

3. LITERATURE REVIEW

3.1 Crop Switching and Climate Adaptation

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Rising and Devineni [2020] introduced a comprehensive crop-switching framework under climate uncertainty, utilising Bayesian yield modelling and spatial optimisation to demonstrate that transitioning from traditional to resilient crops can measurably reduce agricultural losses. While the study established a foundational theoretical basis for quantifying adaptation benefits, it assumed farmer rationality and did not account for transition costs - a gap that the system addresses through its Risk Scoring Engine.

Luh et al. [2022] provided an economic analysis of crop switching using multinomial treatment effects and quantile regression, examining income stability and self-selection bias in agricultural decision-making. Their findings, derived from Taiwanese cross-sectional data, highlight the economic distributional consequences of switching, underscoring the need for a risk-calibrated framework adaptable to Indian conditions.

Wen et al. [2024] explored agricultural vulnerability through Markov regime-switching and threshold regression models, capturing the non-linear dynamics of climate-risk and regime transitions. This work directly motivates the non-linear modelling approach in the system, where XGBoost handles complex interaction terms that linear models fail to capture.

3.2 Machine Learning for Yield Prediction

Kukal and Irmak [2018] investigated climate-driven yield variability using county-level regression and climate detrending across the U.S. Great Plains, providing a baseline methodology for isolating the specific impact of seasonal weather on yield outcomes over multiple decades. Their multi-decadal approach directly informed the system's 50-year temporal scope.

Badshah et al. [2024] developed machine learning models including Random Forest, SVM, and XGBoost for crop classification and yield prediction, demonstrating that ensemble methods provide superior accuracy alongside explainability (XAI) for practical farmer use. Hoque et al. [2024] confirmed this by leveraging meteorological data with Gradient Boosting and KNN for Indian crop yield prediction, demonstrating that ensemble methods provide superior stability in climate-based forecasting.

Waqar, Kim, and Byun [2025] introduced a stacking ensemble with synthetic data augmentation to improve model robustness and reduce prediction errors, highlighting the value of ensemble fusion - a principle central to the system's hybrid LSTM-XGBoost architecture.

3.3 Deep Learning and Temporal Modelling

Elavarasan and Vincent [2020] applied Deep Reinforcement Learning (DRQN) to model non-linear crop yield patterns from climatic, soil, and groundwater data, demonstrating that sequential learning architectures capture environmental interactions invisible to static models. Najjar et al. [2025] enhanced yield prediction interpretability by combining LSTM with Shapley values and Sentinel-2 satellite data, establishing the feasibility of explainable deep learning for remote-sensing-based agricultural forecasting.

Osibo et al. [2025] proposed a Bayesian GRU model with active learning for yield uncertainty estimation, achieving high predictive accuracy even under sparse labelled data conditions - directly relevant to the rural Indian context where historical records may be incomplete.

3.4 Sustainable and Green AI in Agriculture

Elbasi et al. [2025] focused on Green AI approaches for smart agriculture using lightweight and edge-based models, emphasising energy efficiency and scalability for resource-constrained farm environments. Screpnik et al. [2025] provided a systematic review of AI applications in agriculture, categorising research by problem domain and identifying future opportunities for scalable, interpretable, and empirically validated solutions - precisely the profile that the system targets.

4. SYSTEM ARCHITECTURE AND DESIGN

4.1 Modular Multi-Layered Architecture Overview

The system is designed on a Modular Multi-Layered Architecture optimised for scalability, high precision, and low-latency decision support. The architecture is composed of four primary functional layers, each responsible for a distinct stage of the intelligence pipeline, following a feed-forward logic where each module refines environmental and agricultural data before passing it to the next stage of the predictive engine.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

- Module 1 - Data Acquisition and Preprocessing: Ingests 50,765 historical agricultural records from ICRISAT (1966–2017), Soil Health Card parameters, and real-time JSON responses from the Open-Meteo satellite API. Handles outlier removal via the Interquartile Range (IQR) method, mean imputation for missing yields, and normalization across disparate variable ranges.
- Module 2 - Feature Engineering and Signal Synthesis: Computes the Climate Stress Index (CSI) from rainfall-temperature variability, engineers Crop-Climate Interaction terms (Crop-Encoded × Rainfall), and calculates 5-year rolling Historical Yield Volatility as a decadal stability proxy.
- Module 3 - Hybrid ML and Predictive Engine: A two-stage neural-gradient ensemble where an LSTM network extracts 32-dimensional Temporal Embeddings from 10-year sequences, which are fused with tabular SHC data and fed into an XGBoost regressor with Ridge Regression variance stabilisation.
- Module 4 - Risk Calibration and Scoring Engine: Computes a 0–500+ Risk Score by aggregating Yield Stability Delta, Neural Climate Stress Index, and Historical Volatility, then classifies outcomes into LOW (Safe), MEDIUM (Cautionary), and HIGH (Dangerous) transition categories.
- Live Intelligence Layer: Synchronises historical model baselines with real-time Open-Meteo satellite telemetry and Nominatim geocoding, enabling Environmental Safety Overrides when live anomalies exceed historical baselines.

4.2 Module Architecture - Table 1

Table 1. the system Module Architecture - Components and Outputs

Module	Primary Function	Key Components	Output
M1 - Data Ingestion	Historical archive synchronization	ICRISAT 50K records, SHC parameters, IQR cleaning, train/test split 1966–2015 / 2016–2017	Clean feature matrix
M2 - Feature Synthesis	Predictive signal engineering	Climate Stress Index, Crop-Climate interactions, 5-yr rolling volatility	49-feature tensor
M3 - Hybrid Engine	Yield prediction	LSTM (10-yr sequences, 32-dim embeddings), XGBoost (4000 trees), Ridge stabiliser	Predicted yield (kg/ha)
M4 - Risk Engine	Risk quantification	Yield Delta, CSI, Volatility; 0–500+ scoring; percentile thresholds	Risk Score + Category
Live Layer	Real-time calibration	Open-Meteo API, Nominatim geocoding, dynamic baseline shift	Environmental Override

4.3 Operational Workflow

The live operational workflow follows a five-step sequential pipeline designed to personalise risk assessment to a specific farmer's soil and micro-climate conditions:

1. Stakeholder Profile Initialisation: The farmer inputs current crop, target switch crop, desired season (Kharif/Rabi), irrigation type (Rainfed/Borewell), and land size in acres.
2. Village-Level Geocoding Resolution: The system translates the farmer's village or district name into precise geographic coordinates using the OpenStreetMap Nominatim provider, anchoring the analysis to a specific micro-climate region.
3. Real-Time Environmental Telemetry Synchronisation: The Open-Meteo Satellite API is queried for high-resolution 7-day weather telemetry including current temperature, rainfall forecasts, humidity levels, and solar radiation.
4. Dynamic Soil-Climatic Adaptation: Soil parameters are processed through dual-path logic - either via manually provided SHC values (pH, N, P, K) or automatically fetched regional soil proxies, dynamically adjusted by live climate telemetry.
5. Hybrid Predictive Inference and Risk Labelling: The consolidated input - farmer profile, live telemetry, and adjusted soil health data - is fed into the pre-trained hybrid model. The inference engine generates the Crop Switch Risk Score and outputs a clear recommendation on whether the transition is safe, cautionary, or requires a pilot-test period.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

5. DETAILED MODULE IMPLEMENTATION

5.1 Dataset - Historical Agricultural Records

The evaluation knowledge base comprises 50,765 historical agricultural records sourced from the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), spanning 1966–2017 across 20 major Indian states. The dataset encompasses more than 50 distinct crop types across both Kharif (June–October) and Rabi (November–March) seasons, providing comprehensive coverage of India's primary agricultural diversity. Each record includes: crop type, state, season, annual rainfall, mean temperature, area harvested, and yield in kg/ha. Supplementary Soil Health Card (SHC) parameters - pH, Organic Carbon, and macro-/micro-nutrient proxies - are integrated as additional soil stability features.

The dataset underwent a comprehensive multi-stage cleaning pipeline: (1) IQR-based outlier removal eliminated statistically anomalous yield records; (2) state-crop mean imputation addressed missing yield values; (3) MinMaxScaler normalisation standardised heterogeneous climatic variables; and (4) temporal splitting allocated records from 1966–2015 to training and 2016–2017 to validation, ensuring the model is evaluated against the most recent climate volatility.

5.2 Feature Engineering and Signal Synthesis

The feature synthesis module transforms raw environmental measurements into predictive agricultural signals through three specialised engineering pathways:

- Climate Stress Index (CSI): A composite index computed from rainfall variability and temperature deviation scores, quantifying aggregate environmental stress at the crop-region level. The CSI serves as the primary normalised stress input to the LSTM temporal extractor.

- Crop-Climate Interaction Terms: High-order features of the form (Crop_Encoded × Rainfall) and (Crop_Encoded × Temperature) capture species-specific responses to moisture and thermal anomalies, enabling the model to differentiate how drought-tolerant Maize responds versus moisture-intensive Rice under identical environmental shifts.
- Historical Yield Volatility (YV): A 5-year rolling standard deviation of crop yields per state-crop pair, providing a baseline stability metric. High YV signals that a crop historically produces erratic yields in that region - a critical risk multiplier in the Scoring Engine.

Together, these engineered features constitute a 49-feature tensor that serves as input to both the LSTM temporal extractor and the XGBoost regression core.

5.3 Hybrid ML and Predictive Engine

The core neural-gradient engine operates in two sequential stages:

Stage 1 - LSTM Temporal Extractor

A deep Long Short-Term Memory (LSTM) network scans 10-year sequential windows of the 49-feature input tensor. The network architecture comprises two stacked LSTM layers (128 and 64 units respectively), regularised by Dropout layers with a rate of 0.25 to prevent temporal overfitting. The output is a 32-dimensional Temporal Embedding vector that encodes the cumulative climate trajectory of the region over the preceding decade - capturing decadal oscillations, monsoon cycle shifts, and progressive soil degradation patterns invisible to single-season models. Training converged within 130 epochs using the Adam optimiser.

Stage 2 - XGBoost Ensemble Core

The 32-dimensional LSTM Temporal Embeddings are concatenated with the tabular SHC soil health features and fed into an Extreme Gradient Boosting (XGBoost) regressor configured with 4,000 estimators, a learning rate of 0.05–0.1, and early stopping at 250 rounds. Log-transformation (log1p) is applied to yield targets before training to stabilise the variance across the wide dynamic range of Indian crop yields - from sparse rain-fed dryland crops to high-productivity irrigated varieties. This transformation resolved the high-skewness problem that caused the standalone XGBoost baseline to overlook low-yield signals. A Ridge Regression stabiliser is appended as a variance-reduction ensemble layer.

5.4 Risk Calibration and Scoring Engine

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

The Risk Scoring Engine translates mathematical yield predictions into human-centric, actionable safety assessments through a three-step process:

- Dual Yield Prediction: The hybrid model simultaneously predicts the expected yield for the From-Crop (current crop) and the To-Crop (target switch crop) under identical input conditions, generating a Yield Stability Delta (ΔY).
- Composite Risk Score Computation: The 0–500+ Risk Score (RS) is calculated as: $RS = w_1 \times |\Delta Y| + w_2 \times CSI + w_3 \times YV$, where weights are calibrated via percentile-based sensitivity analysis on the training corpus.
- Categorical Risk Labelling: Scores are classified into LOW (≤ 71.84, Safe Transition), MEDIUM (71.84–166.75, Cautionary Switch), and HIGH (>166.75, Dangerous Transition) based on the 33rd and 66th percentile distribution of the risk score across the full 50,765-record training corpus.

5.5 Live Intelligence Layer

The Live Intelligence Layer provides real-time environmental calibration via two integrated services:

- Open-Meteo Satellite API Integration: Fetches current-week high-resolution weather telemetry including daily maximum/minimum temperature, precipitation, relative humidity, and solar radiation forecasts for the 7-day horizon. This data is used to compute a Live Climate Deviation Multiplier applied to the historical risk baseline.
- Dynamic Environmental Safety Override: When real-time weather data indicates an active anomaly - such as an imminent drought, unseasonal flood condition, or extreme heat event - the system automatically increases the Risk Score beyond the historical model's prediction, providing an environmental safety override that protects farmers from historically optimistic but contextually inappropriate recommendations.

6. PERFORMANCE EVALUATION

6.1 Evaluation Methodology

The system was evaluated using the full 50,765-record ICRISAT corpus, with a temporal train/test split assigning records from 1966–2015 to training and 2016–2017 (approximately 1,849 records) to validation. This temporal holdout strategy ensures that the model's generalization is assessed against the most recent and climatically volatile years in the dataset - years characterised by multiple documented "Black Swan" climate events across Indian states. An ablation study was conducted to quantify the specific contribution of LSTM Temporal Embeddings to overall model performance by comparing the hybrid LSTM-XGBoost ensemble against a standalone XGBoost baseline operating on identical tabular features.

6.2 Comparison of Baseline and Hybrid Models

Table 2. Comparison of Baseline and Hybrid Model Performance Metrics

Model Architecture	Training R^2	Test R^2	RMSE (kg/ha)	MAE (kg/ha)
Baseline XGBoost (Static)	0.817	0.732	~349	~241
Proposed Hybrid (LSTM + XGBoost)	0.8816	0.7536	~281	~220
Improvement (Δ)	+6.4%	+2.96%	−68 kg/ha	−21 kg/ha

The integration of LSTM-derived Temporal Embeddings resulted in an absolute R^2 improvement of 6.4% for the training set and a 2.96% improvement on the 2016–2017 test set. The RMSE reduction of approximately 68 kg/ha is particularly significant in the Indian agricultural context, where smallholder farm sizes average 1–2 hectares, and even modest yield prediction improvements translate to substantial economic value per household.

The improvement is directly attributable to the system's ability to capture the cumulative effect of rainfall and temperature trends over the 10-year look-back window - a dimension entirely absent from the static baseline model. Regions with historically volatile climates (high YV) consistently showed the greatest RMSE improvement, confirming that temporal memory is most valuable precisely where risk is highest.

6.3 Hyperparameter Configuration and Training Dynamics

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Table 3. Hyperparameter Configuration and Training Dynamics

Parameter Type	Component	Configuration Value	Optimisation Strategy
Base Estimators	XGBoost	4,000 Trees	High capacity for complex non-linearities
Learning Rate	Hybrid Ensemble	0.05 – 0.1	Slow convergence for minimum RMSE
Memory Window	LSTM (Temporal)	10 Years	Captures long-term climate cycles
Dropout Rate	Neural Layers	0.25	Prevention of model overfitting
Patience	Early Stopping	250 Rounds	Prevention of over-training on noise
Primary Metric	R-Squared	88.16%	Primary measure of variance explained

6.4 Prediction Stability and Log-Transformation Impact

A critical technical optimisation was the application of log1p target transformation to yield values prior to model training. Without transformation, the high-yield outliers in irrigated northern Indian states (Punjab, Haryana) dominated the loss function, causing the regressor to systematically underweight low-yield predictions from rain-fed dryland regions. Log-transformation stabilised variance across the full yield distribution, enabling consistent prediction performance for both high-productivity crops (Rice, Wheat) and lower-yield drought-tolerant varieties (Jowar, Bajra).

Performance evaluation additionally confirmed a strong correlation (Pearson $\rho > 0.82$) between the Neural Climate Stress Index and the final Risk Score, validating that the model correctly identifies elevated risk during climatically anomalous years even when soil health metrics remain constant. This confirms the system's ability to act as a genuine early warning mechanism for agricultural transitions.

6.5 Risk Categorisation Confidence

A categorical validation confirmed that HIGH RISK predictions generated by the Scoring Engine correctly correlated with historically documented crop failures and yield shocks in the 2016–2017 validation dataset, including the Tamil Nadu drought of 2016–17 and the deficient Northeast monsoon across peninsular India. The system flagged 24.1% of the 2016–2017 transition scenarios as HIGH RISK - a proportion consistent with documented yield shock incidence during that period. LOW RISK predictions showed strong agreement with historically stable crop-region pairs, confirming the reliability of the percentile-based threshold calibration.

7. SYSTEM DEPLOYMENT - STREAMLIT ANALYTICAL DASHBOARD

7.1 Dashboard Architecture and User Interface

The primary deployment output of the system is a high-performance interactive analytical dashboard built on the Streamlit framework. The dashboard translates the multi-dimensional outputs of the Hybrid LSTM-XGBoost model into actionable, human-readable insights for both farmers and agricultural policy-makers. The interface employs custom CSS with glassmorphism visual effects, intuitive risk-level colour coding (Green/Amber/Red), and a persistent configuration sidebar for session-wide parameter management.

The global sidebar provides three persistent configuration panels: (1) Farmer Profile Settings for land size and irrigation type; (2) Soil Health Card Integration for toggling between district-level proxies and manually entered lab values (NPK, pH, Organic Carbon); and (3) a Seasonal Sync filter that shifts the analytical context between Kharif and Rabi crop cycles.

7.2 Page-by-Page Functional Analysis

Table 4. Streamlit Dashboard - Page-by-Page Functional Summary

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Page	Title	Key Features	Primary Output
1	National Live Overview	Aggregated national risk KPI, risk distribution donut chart, Top-10 high-risk states, cross-crop switching heatmap	National risk pulse dashboard
2	Real-Time Intelligence	Village geocoding, Open-Meteo satellite sync, dynamic risk inference, 0–500 localised risk score	Live personalised risk score
3	Live Risk Analytics	Multi-dimensional filtering (state, from-crop, to-crop), frequency histograms, statistical box-plots	Detailed statistical explorer
4	State-wise Live Pulse	Live state telemetry, localised risk gauge chart, state crop transition matrix, historical yield trends	Regional deep-dive analytics
5	Crop Recommendations	Ranked alternative crops by safety, natural-language guidance, step-by-step pilot plan with soil testing actions	Personalised transition roadmap
6	Model Performance	Training/test metric bar charts, feature importance analysis, validation sample statistics	Model transparency dashboard
7	Methodology	5-stage pipeline walkthrough, risk formula documentation, parameter explanation guide	System education layer
8	Farmer's Guide	First-use onboarding, sidebar setup walkthrough, risk score interpretation guide, help resources	End-user accessibility portal

7.3 Live Risk Assessment - Vaniyambadi Case Study

A representative live assessment was conducted for a rice-farming location in Vaniyambadi, Tamil Nadu. The system successfully resolved the village coordinates via Nominatim geocoding and fetched live Open-Meteo telemetry showing a current temperature of 32.9°C, 8 mm monthly rainfall forecast, and 28% relative humidity. For a proposed transition from Rice (current crop) to Maize (target crop) under Rabi season conditions, the system generated a MEDIUM RISK classification with a performance delta of -0 kg/ha and the advisory note: "Low gain (<0.8%). Switching cost might outweigh yield benefits." This outcome demonstrates the system's ability to identify economically marginal transitions that carry hidden climate risk - a dimension absent from all current advisory tools.

8. SYSTEM REQUIREMENTS

8.1 Hardware Requirements

Table 5. Hardware Requirements for the system Deployment

Component	Minimum Specification	Recommended Specification
Processor (CPU)	Intel Core i5 / AMD Ryzen 5 (8th Gen+)	Intel Core i7 / AMD Ryzen 7 (11th Gen+) - multi-core for parallel XGBoost
Memory (RAM)	8 GB DDR4	16 GB DDR4 - required for 50K+ record feature matrices
GPU	Optional	NVIDIA GTX 1650+ (CUDA) - accelerates LSTM training via TensorFlow GPU backend
Storage	256 GB SSD	512 GB SSD - for model serialisation and feature engineering I/O
Network	Stable broadband	High-speed connection required for Open-Meteo API and Nominatim geocoding services

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

8.2 Software Stack

Table 6. Software Stack and Package Dependencies

Category	Package / Version	Role in System
Language	Python 3.13.x	Primary implementation language
Deep Learning	TensorFlow >= 2.15.0, Keras	LSTM temporal extractor training and inference
Gradient Boosting	XGBoost >= 2.0.0	Tabular regression and interaction feature learning
Data Processing	Pandas >= 2.1.0, NumPy >= 1.26.0	50K+ record manipulation, feature engineering, risk scoring
ML Utilities	Scikit-learn >= 1.3.0	Preprocessing (MinMaxScaler), metrics (R^2, RMSE), splits
Geospatial	Geopy (Nominatim provider)	Village-level geographic coordinate resolution
Weather API	Requests + Open-Meteo API	Real-time satellite weather telemetry acquisition
Dashboard	Streamlit >= 1.32.0	Interactive analytical dashboard deployment
Visualisation	Plotly, Matplotlib, Seaborn	Interactive maps, heatmaps, yield volatility charts
Environment	OS: Ubuntu 22.04 / Windows 11; IDE: VS Code + Jupyter; VCS: Git	Development and deployment infrastructure

9. COMPARATIVE ANALYSIS - EXISTING vs. PROPOSED SYSTEM

Table 7. Comparative Matrix: Existing Systems vs. the system

Feature	Existing Systems (SHC/Regional Advisory)	the system (Proposed Hybrid LSTM-XGBoost)
Logic Type	Static & Regional	Dynamic & Localised (village-level)
Data Ingestion	Manual / Delayed CSV	Real-time Satellite Sync (Open-Meteo API)
Predictive Power	Low (Linear / Simple Statistical)	High (88.16% R^2 Hybrid Neural-Gradient)
Contextual Memory	None (Single-Season Snapshot)	10-Year Temporal Memory (LSTM Embeddings)
Transition Risk Logic	Absent - no switching quantification	Core feature - 0–500+ Calibrated Risk Score
Risk Interpretation	Vague Descriptive Advice	Precise 501+ Categorised Score (LOW/MED/HIGH)
Village Resolution	Coarse District-Level	Village-Level Robust Geocoding (Nominatim)
Live Environmental Sync	Not available	Open-Meteo satellite telemetry with dynamic override
Soil Integration	Basic NPK (SHC only)	Full SHC with dynamic climate-adjusted soil proxies
Farmer Interface	Government portal / paper advisories	Interactive Streamlit dashboard with pilot plan

10. SECURITY, PRIVACY, AND SOCIETAL IMPACT

10.1 Data Privacy and Security

the system adopts a privacy-by-design architecture with three governing principles: (1) minimal data retention - farmer profiles are processed in session scope and not persistently stored without explicit consent; (2) purpose limitation - only aggregated risk scores and anonymised geolocation data are logged for system improvement; and (3) transparent processing - all model decisions are explainable through the dashboard's Model Performance and Transparency page, which exposes feature importance rankings via XGBoost's built-in SHAP-compatible importance engine.

10.2 Algorithmic Fairness and Bias Mitigation

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Agricultural AI systems carry a documented risk of encoding geographic or demographic biases present in historical training data. the system mitigates this through two design choices: first, the use of physics-informed feature engineering (CSI, YV) rather than black-box discriminative classifiers reduces sensitivity to spurious correlations; second, the percentile-based risk threshold calibration ensures balanced distribution across LOW, MEDIUM, and HIGH categories, preventing systematic overclassification of any particular region or crop type as high-risk.

Farmers receiving HIGH RISK assessments are provided with a structured Pilot Plan rather than a binary prohibition, preserving decision autonomy while quantifying risk. The system explicitly frames its outputs as decision-support guidance, not prescriptive mandates, and recommends consultation with Krishi Vigyan Kendra (KVK) agricultural extension officers for validation of HIGH RISK scenarios.

10.3 Societal and Economic Impact

The significance of the system extends across four dimensions of societal impact:

- Economic Vulnerability Reduction: By quantifying high-risk transitions before they are implemented, the system prevents catastrophic financial losses for smallholder farmers, directly supporting household food security and economic stability.
- Climate Resilience Advancement: The system provides a scientifically validated roadmap for regional crop adaptation, supporting national food security strategies under accelerating climate change baseline shifts.
- Bridging the Digital Divide: the system brings Deep Learning and real-time satellite intelligence to the village level, democratising advanced data science that was previously accessible only to large commercial agricultural enterprises.
- Technological Contribution: From a computer science perspective, this research provides a novel implementation of Temporal Embedding Fusion, demonstrating how deep time-series signals from LSTM networks can be merged with gradient-boosted decision trees to solve high-impact real-world sustainability challenges.

11. CONCLUSION AND FUTURE WORK

11.1 Conclusion

This paper presented the system, a comprehensive hybrid machine learning platform that addresses the critical gap in quantitative crop switching risk assessment for smallholder Indian farmers under climate change. By shifting the objective from simple yield prediction to multi-layered transition risk calibration, the system provides a robust technological shield for agricultural communities navigating the complexities of climate-driven crop diversification.

The core technical achievement is the Hybrid LSTM-XGBoost Ensemble, which bridges the gap between long-term temporal memory (captured by LSTM 10-year sequence learning) and high-resolution tabular interactions (handled by XGBoost 4,000-tree regression). The model achieves a benchmark accuracy of 88.16% R^2, with the LSTM temporal embeddings contributing a statistically significant 6.4% absolute improvement over the standalone baseline - confirming that decadal climate memory is indispensable for reliable crop transition risk assessment.

The integration of Live Intelligence Layers via Open-Meteo satellite geocoding and real-time weather API telemetry ensures that all risk assessments are dynamically calibrated to current environmental realities, not merely historical averages. By simplifying complex neural outputs into localised LOW, MEDIUM, and HIGH risk scores with actionable pilot plans, the system successfully democratises advanced data science for the agricultural community, enabling safer and more economically stable crop switching decisions at the village level.

11.2 Key Achievements

- Hybrid LSTM-XGBoost pipeline achieving 88.16% R^2 on 50,765 historical agricultural records with RMSE of approximately 281 kg/ha.
- 0–500+ Risk Scoring Engine with percentile-calibrated LOW/MEDIUM/HIGH thresholds validated against 2016–2017 documented climate events.
- Live satellite intelligence layer with village-level geocoding resolution (sub-district Nominatim precision) and real-time environmental safety overrides.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

- Eight-page Streamlit analytical dashboard delivering personalised crop transition pilot plans, state-level deep dives, and model transparency visualisations.
- Ablation-validated temporal embedding contribution of +6.4% R^2 and −68 kg/ha RMSE reduction over the standalone XGBoost baseline.

11.3 Future Roadmap

Several strategic enhancements are identified for future development:

1. Direct IoT Sensor Integration: Future iterations will incorporate real-time in-field soil moisture and nutrient sensors, replacing regional proxies with exact, real-time localized soil health data to further reduce prediction error.
2. Multilingual Voice Interface: A voice-activated recommendation interface in Tamil, Telugu, Hindi, Kannada, and Marathi will broaden accessibility for rural populations with varying digital literacy levels.
3. Market Price Integration: Incorporating real-time Mandi market price volatility will enable a Financial Risk-Reward Score alongside the Yield Stability Score, providing farmers with a holistic economic transition assessment.
4. CMIP6 Climate Scenario Stress Testing: Integration of long-term CMIP6 climate projection models will allow farmers to evaluate how specific crop transitions might perform not just under current conditions, but across multiple 10-20 year global warming scenarios.
5. Global Scaling and Cross-Continental Adaptation: Expanding the training corpus to include agricultural data from Sub-Saharan Africa and Southeast Asia will enable the system to serve as a globally applicable crop transition resilience platform.

REFERENCES

[1] J. Rising and N. Devineni, "Crop Switching Reduces Agricultural Losses under Climate Change," Agricultural Systems, vol. 182, pp. 102–115, 2020.
[2] Y. Luh, Y.-C. Chang, and S.-T. Ho, "Crop Switching and Farm Sustainability: Empirical Evidence from Multinomial Treatment-Effect Modeling," Journal of Agricultural Economics, vol. 73, no. 2, pp. 245–261, 2022.
[3] X. Wen, P. Mennig, and J. Sauer, "Assessing the Regime-Switching Role of Risk Mitigation Measures on Agricultural Vulnerability: A Threshold Analysis," Ecological Economics, vol. 210, pp. 107–123, 2024.
[4] M. S. Kukal and S. Irmak, "Climate-Driven Crop Yield and Yield Variability in the U.S. Great Plains," Agricultural and Forest Meteorology, vol. 258, pp. 210–223, 2018.
[5] F. Najjar, A. Saleh, and B. Qureshi, "Explainability in Remote Sensing Yield Prediction Using XAI Methods," Computers and Electronics in Agriculture, vol. 199, pp. 107–115, 2025.
[6] C. Osibo, T. Wang, and M. Li, "Iterative Querying with Bayesian GRU for Crop Yield Prediction," IEEE Access, vol. 13, pp. 1420–1435, 2025.
[7] E. Elavarasan and P. Vincent, "Deep Reinforcement Learning for Crop Yield Prediction," Computers and Electronics in Agriculture, vol. 175, pp. 105–115, 2020.
[8] A. Badshah, B. Y. Alkazemi, F. K. Z. Zamil, and M. Haris, "ML-Based Crop Classification and Yield Prediction for Agricultural Sustainability," Sustainable Computing: Informatics and Systems, vol. 34, pp. 100–114, 2024.
[9] R. Hoque, A. Hameed, and S. Khan, "Meteorological Data and Machine Learning for Indian Crop Yield Prediction," Journal of Agrometeorology, vol. 26, no. 3, pp. 221–233, 2024.
[10] M. Waqar, J. Kim, and H. Byun, "Stacking Ensemble with Synthetic Data for Crop Yield Prediction," Computers and Electronics in Agriculture, vol. 210, pp. 118–130, 2025.
[11] E. Elbasi, Y. Alzoubi, A. E. Topcu, and M. Nadeem, "Green AI for Smart Agriculture: Energy-Efficient Predictive Models for Crop Yield and Resource Management," IEEE Access, vol. 13, pp. 2400–2415, 2025.
[12] L. Screpnik, P. Zamudio, and J. Gimenez, "AI in Agriculture: A Systematic Review of Machine Learning Models for Crop Yield Prediction," Computers and Electronics in Agriculture, vol. 200, pp. 108–124, 2025.
[13] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[14] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, 2016.
[15] ICRISAT, "Village Dynamics in South Asia (VDSA) Agricultural Panel Dataset, 1966–2017," International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India, 2018.