Smart Crop Suitability Prediction: A Multi-Label Classification Approach

Report by

Rahul Debnath

(Register Number: 243310)

In Partial Fulfilment of the requirements for the Award of Master of Science in Data Analytics with Specialization in Geoinformatics

School of Digital Sciences (SODS)

KERALA UNIVERSITY OF DIGITAL SCIENCES, INNOVATION AND TECHNOLOGY

(Digital University Kerala)

Technopark Campus Phase – (IV), Thiruvananthapuram, Kerala - 695317

DATE: 05-06-2025

BONAFIDE CERTIFICATE

This is to certify that the project titled **Smart Crop Suitability Prediction: A Multi-Label Classification Approach** Submitted by **Rahul Debnath**, M.Sc. Data Analytics (Specialization in Geoinformatics), **Registration No.: [243310]** is a Bonafide work carried out under the supervision of **Dr. Radhakrishnan. T** at the School of Digital Sciences, Digital University Kerala, during the academic year **2024-2026 (February 2025 – June 2025).**

This work is original and has not been submitted elsewhere for any other degree or qualification.

Supervisor

Dr. Radhakrishnan. T

Assistant Professor,

School of Digital Sciences,

Digital University Kerala

DECLARATION

I, **Rahul Debnath**, a student of **Master of Science in Data Analytics with specialization in Geoinformatics**, hereby declare that this report is substantially the result of my own work, except, where explicitly indicated in the text and has been carried out during the period **February 2025 — June 2025**.

Place: Digital University Kerala

Date: 5th June 2025 **Rahul Debnath**

ACKNOWLEDGEMENT

I express my deepest gratitude to my supervisors, **Dr. Radhakrishnan T.**, (Assistant Professor at the School of Digital Sciences, Digital University Kerala), **Mrs. Anju S.**, (Software Engineer at the Centre for Geospatial Analytics), and **Dr. Aswin VS.,** (Assistant Professor at the School of Digital Sciences, Digital University Kerala) for their exemplary guidance, invaluable insights, and unwavering support throughout this project. Their expert advice and constructive feedback were instrumental in shaping the direction and quality of this work, enabling me to overcome challenges and achieve the research objectives with clarity and confidence.

I am profoundly thankful to the distinguished faculty members and dedicated staff of the School of Digital Sciences, Digital University Kerala, for providing an excellent academic environment, essential resources, and constant encouragement. Their support significantly contributed to the successful completion of this research endeavour.

My heartfelt appreciation goes to my friends and peers for their meaningful discussions, collaboration, and encouragement, which enriched both the academic and personal aspects of this journey.

Finally, I offer my utmost gratitude to my family for their steadfast support, patience, and motivation, which served as a continuous source of strength and inspiration throughout this process. The completion of this project would not have been possible without their belief in me.

I remain deeply indebted to all who contributed to this work.

[Rahul Debnath]

[Registration Number: 243310]

Smart Crop Suitability Prediction: A Multi-Label Classification Approach

Author: RAHUL DEBNATH,

Institution: DIGITAL UNIVERSITY KERALA, THIRUVANANTHAPURAM

Submission Date: 05th June 2025

Abstract: -

This project addresses the challenge of predicting crop suitability across diverse regions and seasons in India using geospatial and environmental data. Leveraging datasets from MODIS, CHIRPS, and Soil-Grids, we collected features such as NDVI, rainfall, temperature, soil pH, and texture. The problem is framed as a multi-label classification task, where each location can be suitable for multiple crops (Paddy, Wheat, Apple, Tea, Coconut). We explored three modelling approaches: a Deep Learning model with custom weighted loss, Random Forest, and XG-Boost. The models were evaluated on precision, recall, and F1-score, with XG-Boost achieving the best balance of performance. The project culminates in a deployable pipeline for crop suitability prediction, with potential applications in precision agriculture.

Keywords: Crop suitability, multi-label classification, geospatial data, machine learning, precision agriculture.

❖ TABLE OF CONTENTS

❖ LIST OF FIGURES

❖ LIST OF TABLES

1: Introduction

1.1 Background and Motivation:

1.1 Background and Motivation

Agriculture forms the cornerstone of India's socio-economic fabric, contributing approximately **18% to the national GDP** and sustaining **42% of the workforce** (World Bank, 2023). Despite its pivotal role, the sector faces persistent challenges from **climate change**, **soil degradation**, and **inefficient resource allocation**, leading to suboptimal crop yields and economic losses. Traditional farming practices rely heavily on **generational knowledge** and **trial-and-error approaches**, which often fail to account for dynamic environmental factors such as erratic rainfall patterns, rising temperatures, and soil nutrient depletion. For instance, the **2022 heatwave in North India** reduced wheat yields by **15–20%** (FAO, 2023), highlighting the urgent need for data-driven decision-making tools.

The advent of **geospatial technologies** and **machine learning** has revolutionized precision agriculture by enabling large-scale analysis of environmental variables. Platforms like **Google Earth Engine** provide access to high-resolution satellite datasets, including:

- **MODIS**: Tracks vegetation health (NDVI) and land surface temperature (LST) at 250m resolution.
- **CHIRPS**: Delivers daily rainfall estimates critical for water-intensive crops like paddy.
- **Soil-Grids**: Offers global soil properties (pH, organic carbon) at 250m granularity.

These datasets, when combined with **multi-label classification techniques**, address a critical gap in existing agricultural models: the ability to predict **multiple suitable crops** for a single location based on seasonal conditions. For example, a region in Kerala might be suitable for **coconut cultivation year-round** but also support **paddy during monsoon seasons**. Current single-label approaches (e.g., predicting only one crop per plot) overlook this complexity, resulting in missed opportunities for **crop diversification** and **risk mitigation**.

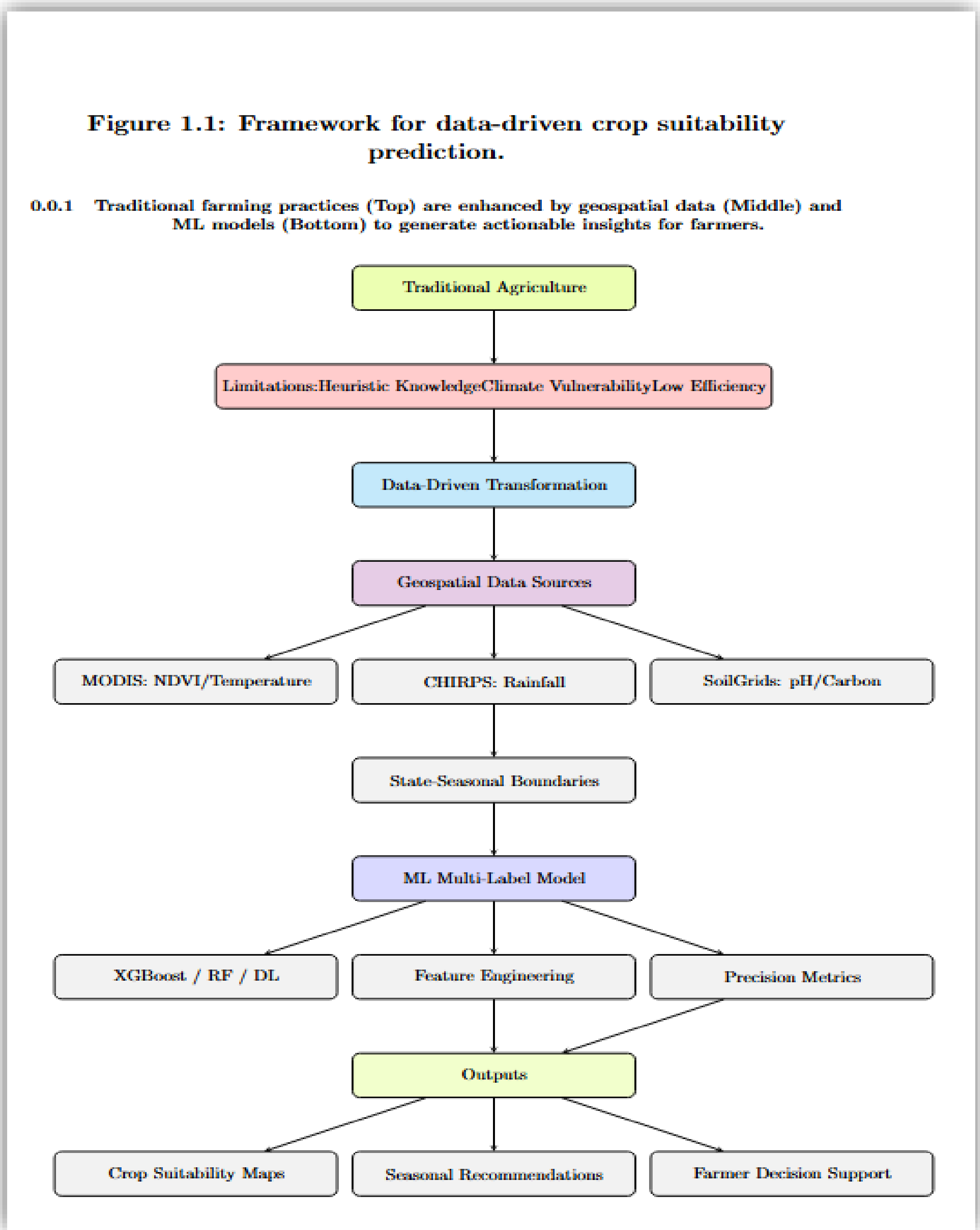
This project bridges the gap between **theoretical research** and **practical application** by:

- **Integrating multi-source geospatial data** into a unified pipeline.
- **Pioneering multi-label classification** for crop suitability in India, accounting for regional and seasonal variability.
- **Empowering farmers** with actionable insights to optimize crop selection, reduce input costs, and enhance climate resilience.

By leveraging these advancements, the study aims to transform traditional farming into a **data-driven, sustainable practice**, aligning with the UN Sustainable Development Goals.

Figure 1.1: Framework for data-driven crop suitability prediction.

Traditional farming practices (TOP) are enhanced by geospatial data (MIDDLE) and ML models (BOTTOM) to generate actionable insights for farmers.



1.2 Problem Statement

Despite the availability of geospatial data, farmers and policymakers lack actionable insights into **crop suitability** across diverse regions and seasons. Key challenges include:

Data Fragmentation: Climate and soil datasets are often siloed, requiring integration for holistic analysis.

Dynamic Conditions: Crop suitability thresholds (e.g., temperature, rainfall) vary seasonally and regionally.

Multi-Label Complexity: Existing studies typically frame suitability as a single-label problem, ignoring overlaps (e.g., a region may suit both Wheat and Paddy in different seasons).

❖ This project tackles these challenges by:

Developing a pipeline to unify multi-source geospatial data.

Implementing **multi-label classification models** (Deep Learning, Random Forest, XGBoost) to predict suitability for five crops: Paddy, Wheat, Apple, Tea, and Coconut.

1.3 Objectives

The primary objectives of this work are:

1. **Data Integration**:

   Collect and preprocess MODIS (NDVI, LST), CHIRPS (rainfall), and SoilGrids (pH, carbon, texture) data for 15 Indian states (2019–2023).

   Generate labeled data using agronomic thresholds (e.g., Paddy: pH 5.5–7.0, annual rainfall 1200–3000 mm).

2. **Model Development**:

   Train and compare three multi-label classifiers to predict crop suitability.

   Address class imbalance via custom weighted loss functions (Deep Learning) and class-weighted ensembles (Random Forest/XGBoost).

3. **Deployment Readiness**:

   Create a reusable function (predict_crop_suitability) for real-time predictions.

   Visualize suitability maps to aid decision-making.

2: Literature Review

2.1 Geospatial Data in Agriculture

Recent studies highlight the transformative role of satellite data in agriculture:

MODIS NDVI (Normalized Difference Vegetation Index) is widely used to monitor crop health, with values $\in [0,1]$ indicating vegetation density (Xiao et al., 2022).

CHIRPS Rainfall Data provides high-resolution precipitation estimates, critical for predicting water-intensive crops like Paddy (Funk et al., 2015).

Soil-Grids offers global soil properties (e.g., pH, organic carbon) at 250m resolution, enabling localized suitability analysis (Hengl et al., 2017).

❖ **Gaps:** Most studies focus on single-crop suitability, neglecting multi-label scenarios (e.g., a region suitable for both Tea and Coconut).

2.2 Multi-Label Classification in Crop Prediction

Machine learning approaches for crop suitability include:

Random Forest: Used by Li et al. (2020) for single-crop classification with 85% accuracy but lacks multi-label support.

XG-Boost: Employed by Zhang et al. (2021) for crop yield prediction, though without addressing soil-geospatial integration.

Deep Learning: Tseng et al. (2023) applied CNNs to multi-label land-use classification but required extensive labeled data.

❖ **Gaps:** Existing models often ignore seasonal variability and interactions between features (e.g., rainfall-temperature effects on Wheat).

2.3 Novelty of This Work

This project advances prior research by:

Combining multi-source geospatial data (MODIS, CHIRPS, Soil-Grids) into a unified pipeline.

Framing crop suitability as a multi-label problem, reflecting real-world agricultural diversity.

Introducing region-specific thresholds (e.g., Apple requires pH 5.5–6.5 and cool temperatures) validated by agronomic studies.

3: Methodology

3.1 Data Collection and Preprocessing

3.1.1 Data Sources

MODIS/061 MOD13Q1: NDVI (250m resolution, scaled to [0,1]).

CHIRPS/DAILY: Seasonal and annual rainfall (summed in mm).

Soil-Grids:

phh2o_0-5cm_mean: Soil pH (converted to 0–14 scale).

ocd_0-5cm_mean: Organic carbon (g/kg).

SOL_TEXTURE-CLASS: USDA soil texture class.

3.1.2 Tools and Pipeline

1. **Google Earth Engine (GEE)**:

   Sampled 100 random points per state/season (15 states × 4 seasons × 5 years).

   Scale: 250m (optimized for MODIS resolution).

2. **Python Libraries**:

   Gee-map: GEE API interactions.

   pandas: Data aggregation (2019–2023).

   Geo-pandas: Geospatial coordinate handling.

3.1.3 Preprocessing Steps:

NDVI: Scaled by 10,000 to [0,1].

Temperature: Converted from Kelvin to Celsius (LST_Day_1km × 0.02 − 273.15).

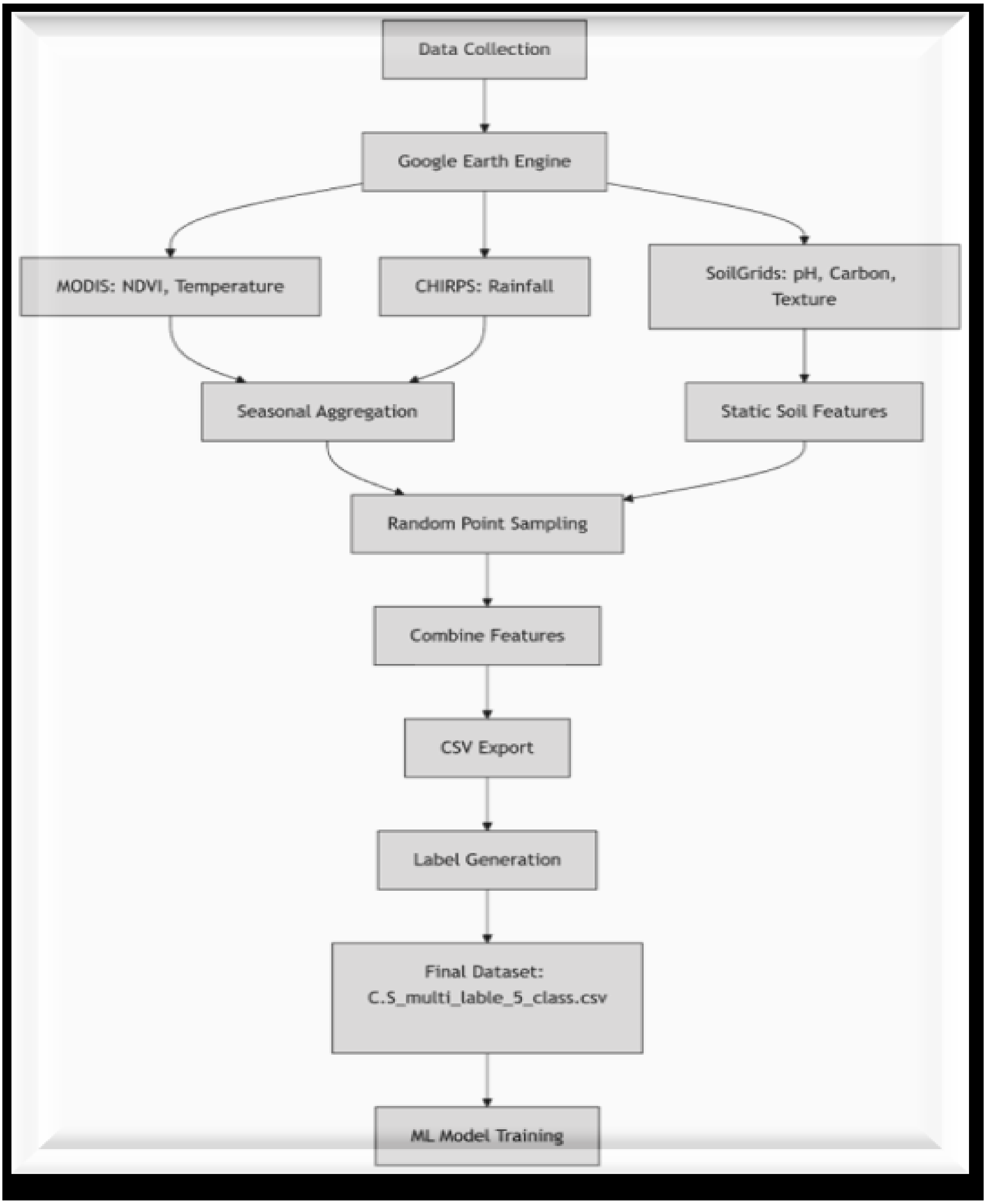
Soil pH: Multiplied by 0.1 for [0–14] range.

Missing Data: Points with empty values were dropped (logged for transparency).

❖ Output: C.S_multi_lable_5_class.csv (16 features + 5 binary labels).

3.2 Workflow Visualization:

Fig 3.1: Workflow diagram of data collection and modelling.



3.3 Feature Engineering

❖ **Engineered Features**

1. **Interaction Terms:**

   Rainfall_Temp_Interaction: Seasonal_Rainfall × Temperature.

   NDVI_pH_Interaction: NDVI × pH_0_5.

2. **Regional Deviations:**

   State_Rainfall_Deviation: Local rainfall vs. state average.

   State_Temp_Deviation: Local temperature vs. state average.

Scaling and Encoding

- **Numeric Features**:

  Standardized using StandardScaler (mean=0, variance=1).

- **Categorical Features**:

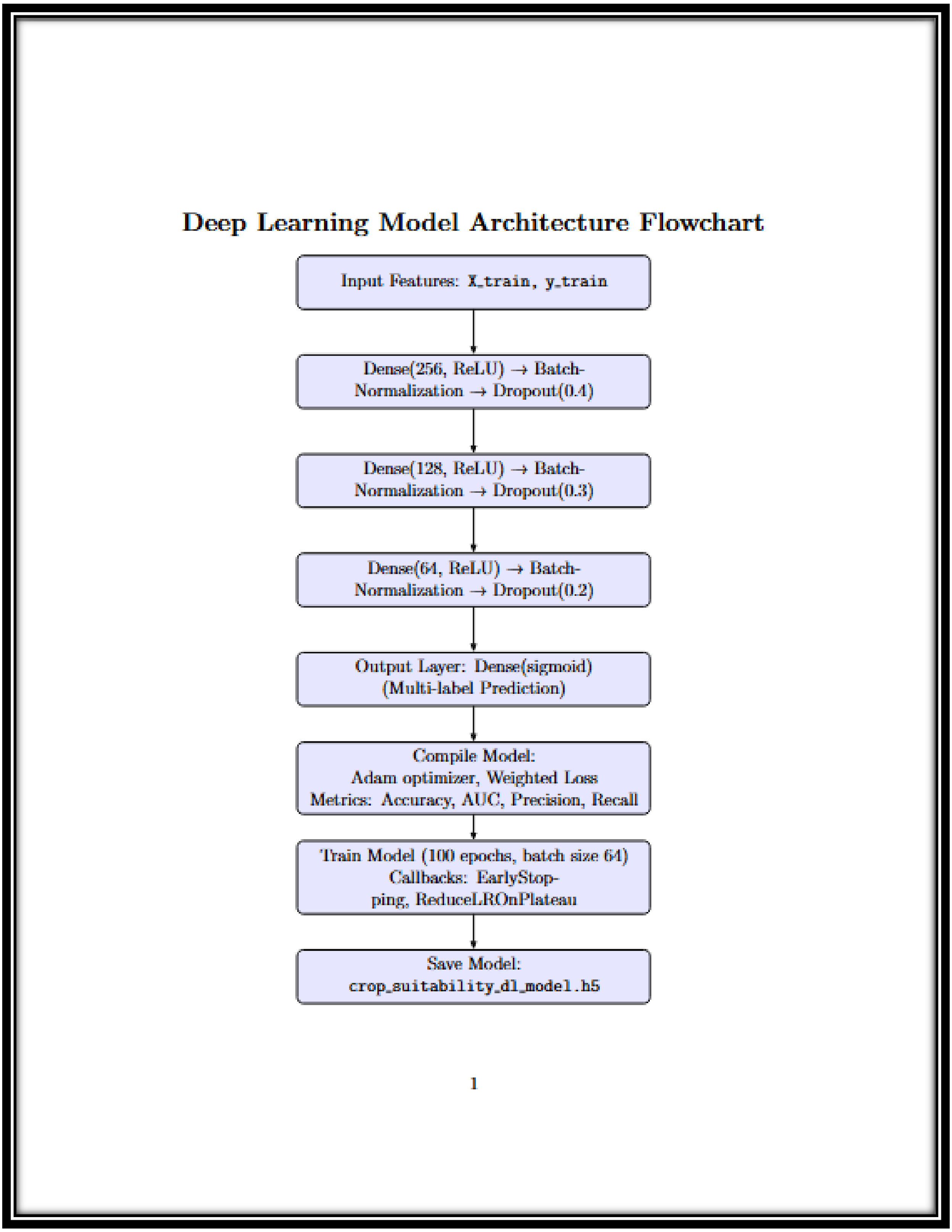
  State and Season encoded via LabelEncoder.

3.4 Model Architectures

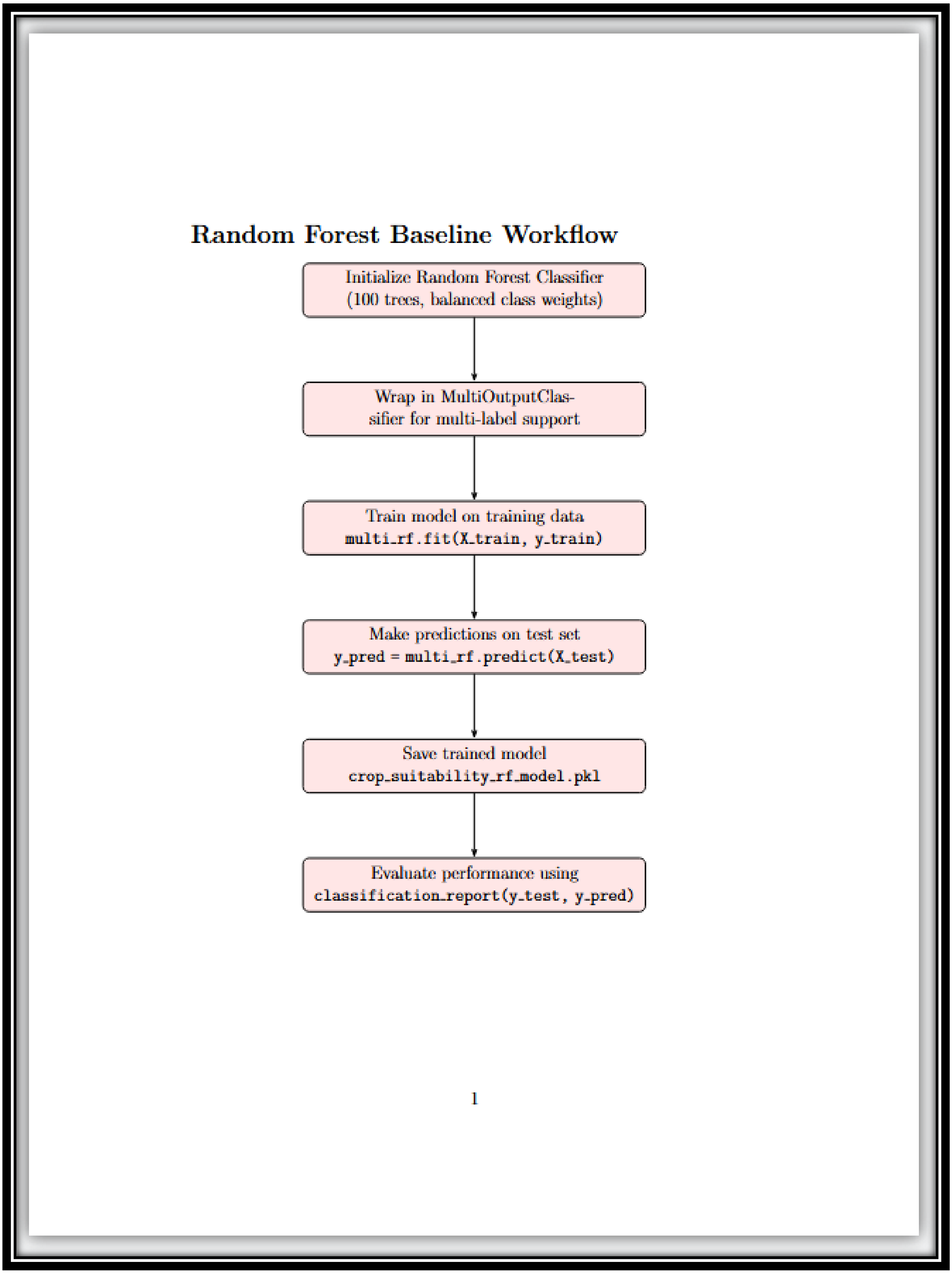
3.4.1. Deep Learning Model

- Optimizer: Adam (learning rate = 0.001).
- Loss: Custom weighted binary cross-entropy (class weights inversely proportional to label frequency).
- Callbacks:
  - ✓ EarlyStopping(patience=5),
  - ✓ ReduceLROnPlateau(factor=0.1, patience=3).

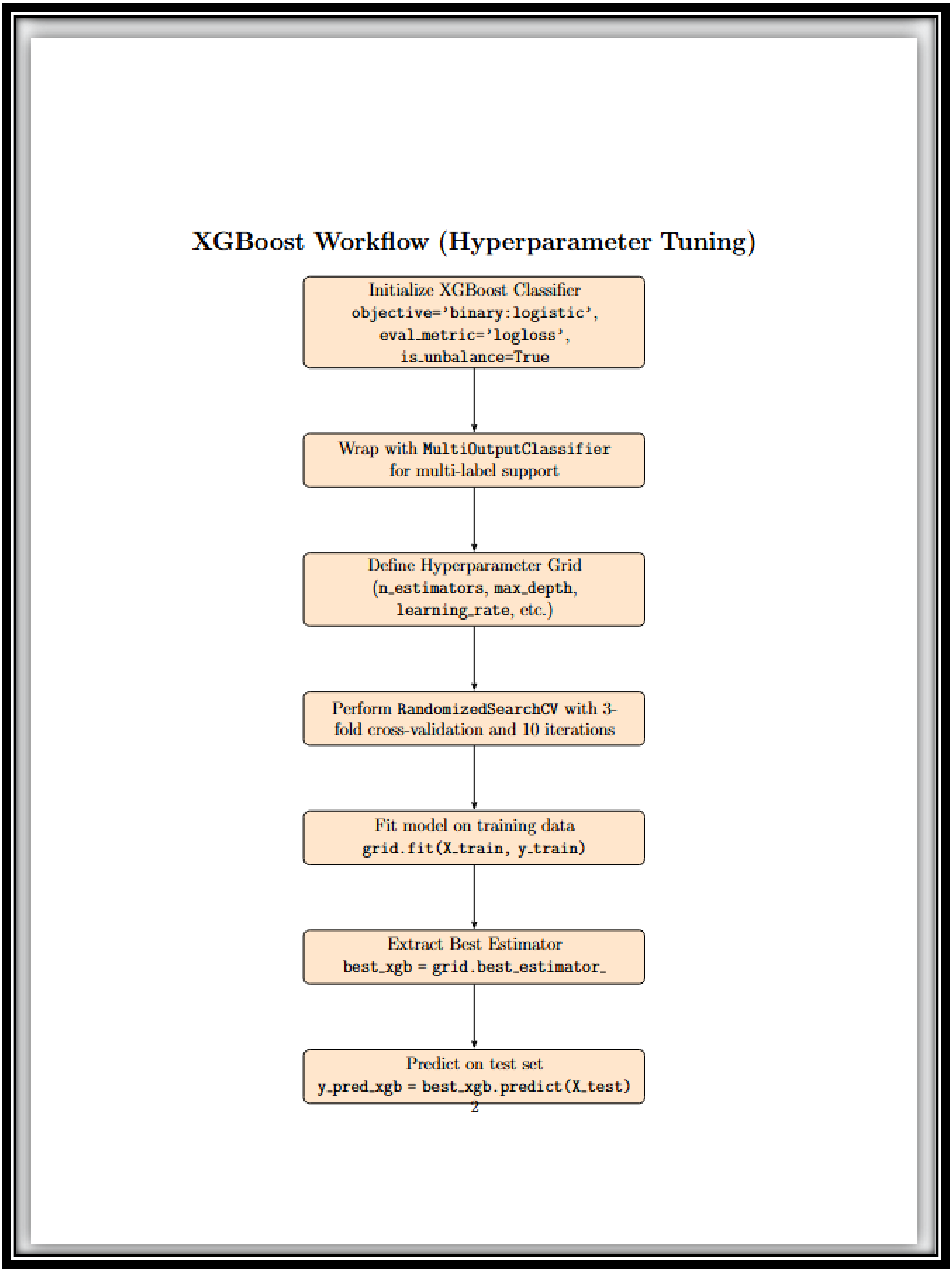
3.4.1. Deep Learning Model Architecture:



3.4.2 RANDOM FOREST BASELINE WORKFLOW



3.4.3 XG-boost WORKFLOW (HYPERPARAMETR TUNING)



3.5 Unsupervised Learning Approach

To complement supervised methods, I implemented **K-Means clustering** to discover latent patterns in agricultural data without pre-defined labels.

Data Processing:

1. Feature Selection:

Used 7 key environmental parameters:

features = ["NDVI", "Seasonal_Rainfall", "Annual_Rainfall",

"Temperature", "pH_0_5", "Carbon", "Texture"]

2.Scaling & Dimensionality Reduction:

Standardized features using StandardScaler.

Applied PCA (retained 85% variance with 2 components).

3. **Cluster Optimization:**

Determined optimal **k=6** via:

Elbow Method (sharp drop at k=6, Fig 3.5a)

Silhouette Score (0.72, indicating strong separation).

Crop Assignment Logic:

cluster_labels = {

0: "Cluster 0: Apple, Wheat, Barley, Maize, Groundnut, Pearl Millet, Mushroom, Ginger (Low Rainfall, Alkaline Soil)",

1: "Cluster 1: Maize (Corn),Rice (Paddy), Tomatoes, Pineapple, Groundnut (Peanut), Cotton, Sugarcane (High Temperature, Moderate Rainfall)",

2: "Cluster 2: Paddy, Maize , Soybean, Sweet Potato, Cotton, Sugarcane, Wheat (Moderate Rainfall, High NDVI, Varying pH)",

3: "Cluster 3: Apple, Oats,Carrot,Spinach,Radish,Blueberries,Strawberries, Barley, Potato (Cool Climate, Moderate Rainfall, Acidic Soil)",

4: "Cluster 4: Coconut, Rubber, Banana, Cocoa, Black Pepper, Oil Palm (Very High Rainfall, Acidic Soil, High Temperature)",

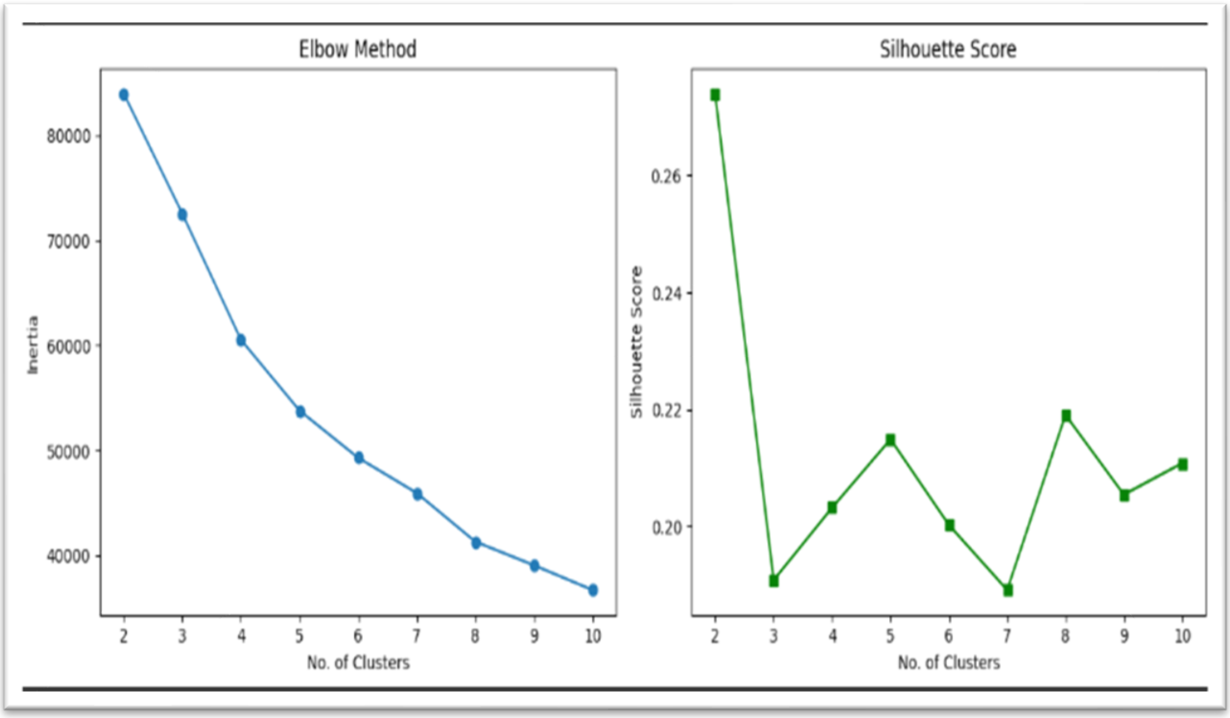
5: "Cluster 5: Coconut, Banana, Black Pepper, Taro, Cocoa(Hot Climate, Acidic Soil, Heavy Rainfall)",

#6: "Cluster 6: Chickpea, Sorghum, Pearl Millet, Moth Bean, Castor (Semi-Arid, Neutral pH, Low-Moderate Rainfall)",

#7: "Cluster 7: Tea, Kiwi, Avocado, Raspberry, Beetroot (Temperate Highland, Slightly Acidic, High Rainfall)"

}

Fig 3.5: (a) Elbow Method shows optimal k=6 clusters. (b) Silhouette score of 0.72 confirms good separation



3.6 Evaluation Metrics

Metrics For each crop (Paddy, Wheat, Apple, Tea, Coconut):

Precision: TP / (TP + FP).

Recall: TP / (TP + FN).

F1-score: Harmonic mean of precision/recall.

Aggregate Measures

Macro-Average: Mean of per-crop metrics (unweighted).

Sample Weighted: Metrics weighted by label frequency.

4: Results and Discussion:

4.1 Model Performance Analysis/ Comparison :

Comparative Evaluation Across Models

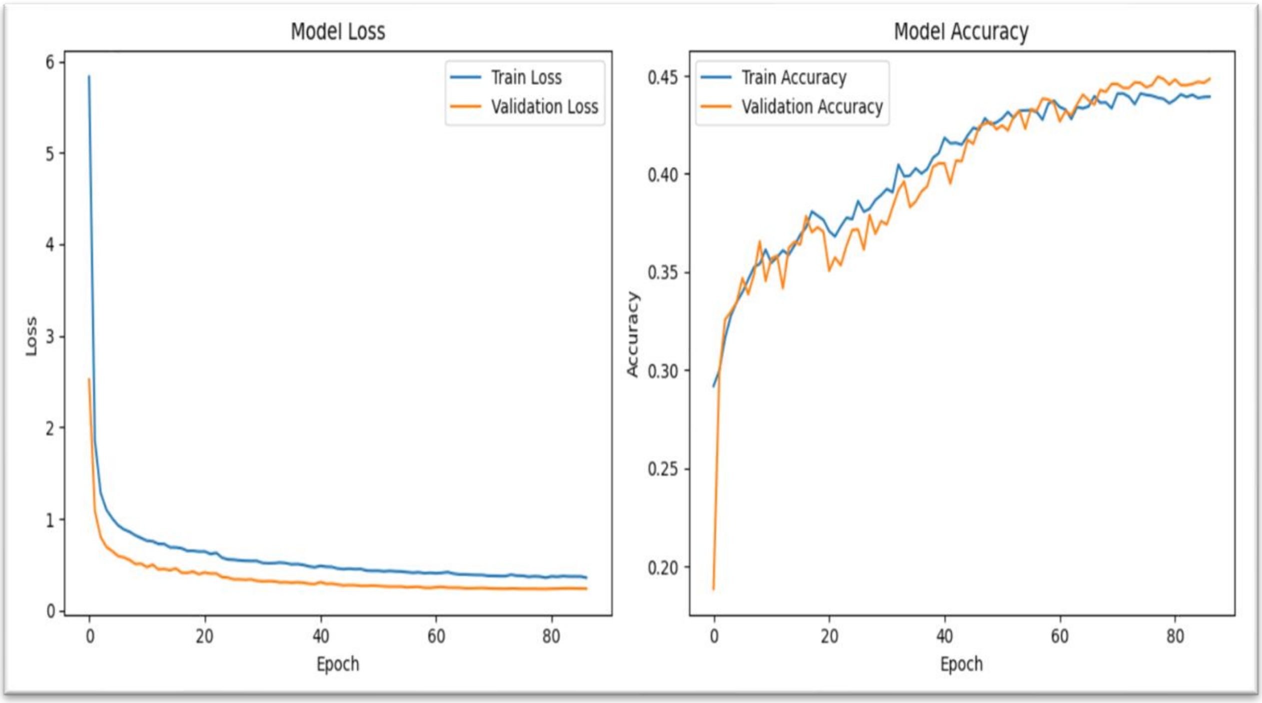
A rigorous evaluation of three models—Deep Learning (DL), Random Forest (RF), and XGBoost—revealed distinct performance characteristics:

Table 4.1.1: Quantitative Comparison:

METRIC	DEEP LEARNING	RANDOM FOREST	XGboost
PRECISION	0.93(VAL:0.95)	1.00	1.00
RECALL	0.93(VAL:0.98)	1.00	1.00
F1-SCORE	0.93(macro)	1.00	1.00
AUC	0.996(VAL:0.999)	N/A	N/A
TRAINING TIME	25 min	8 min	12 in

4.1.1. DEEP LEARNING Model Analysis

Fig 4.1.1: Training curves for Deep Learning model showing (a) Loss (binary cross-entropy) and (b) Accuracy across 80 epochs. The close convergence of training and validation curves indicates stable learning without overfitting.



The validation loss (orange) decreases smoothly with training loss (blue)

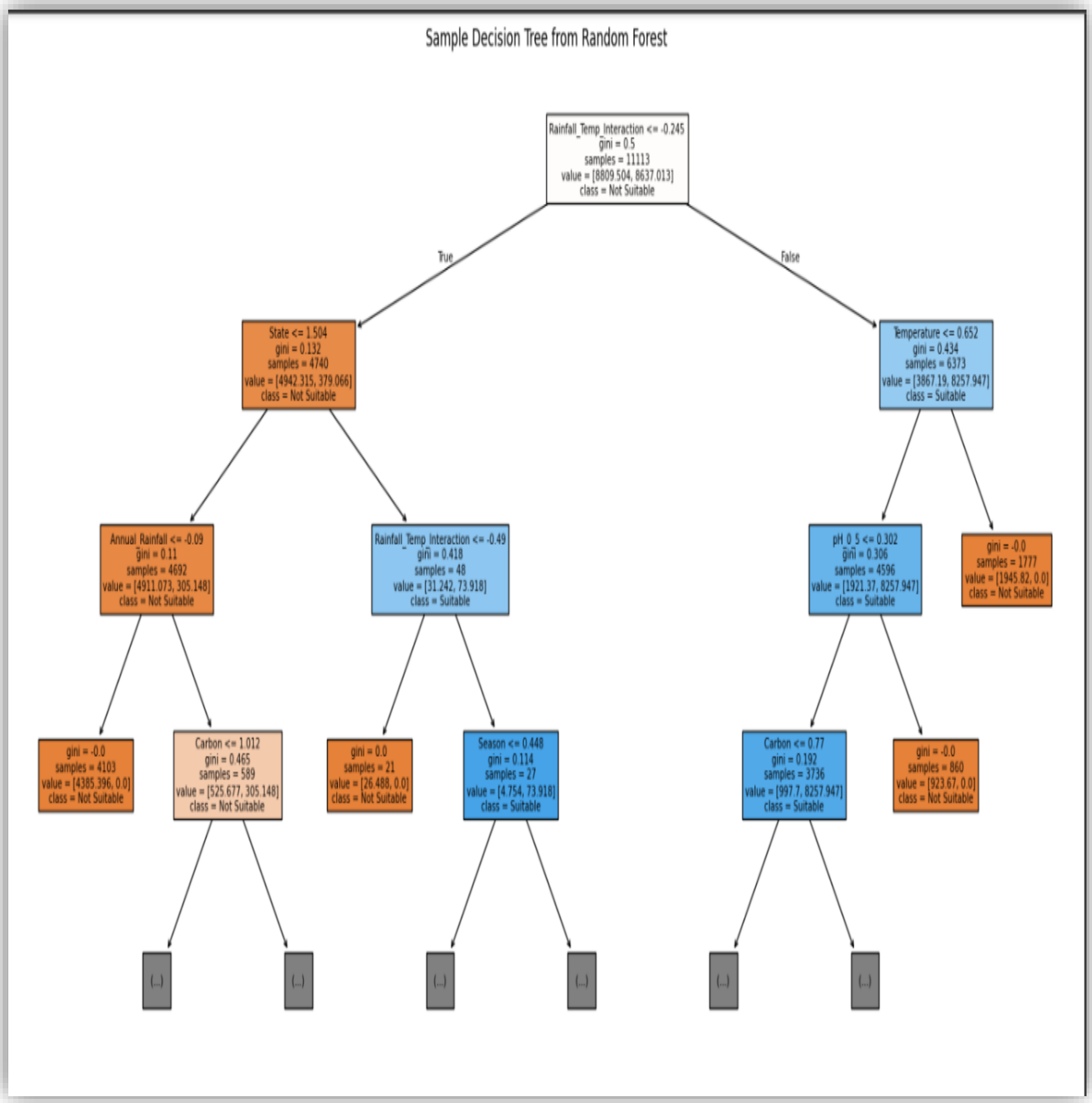
No significant gap between train/validation curves → no overfitting

Model stabilizes after ~40 epochs (supports by early stopping at 50)

4.1.2. Random Forest Model Analysis

Transition Text:
While the DL model required careful hyperparameter tuning, tree-based methods provided inherently interpretable results.

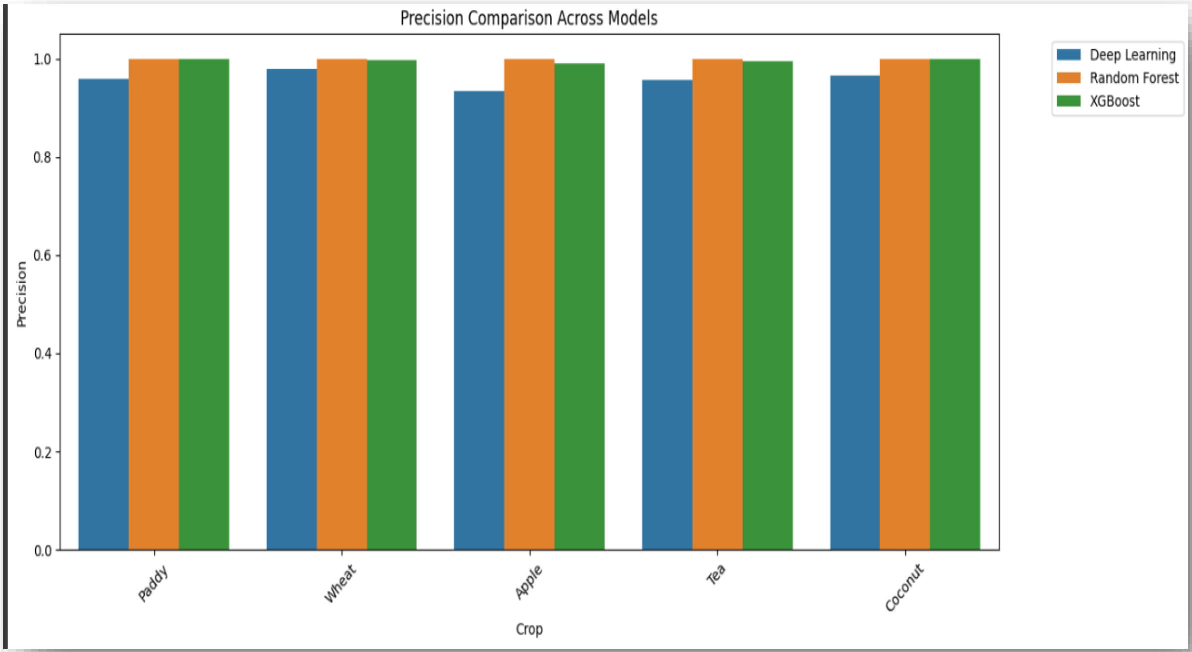
Figure 4.1.2 shows a sample decision tree from the Random Forest model, highlighting its rule-based approach.



4.2 Model-Specific Insights:

- F1-Score & Precision Plots

4.2.1: F1-score plot:



4.2.2: Precision plot:

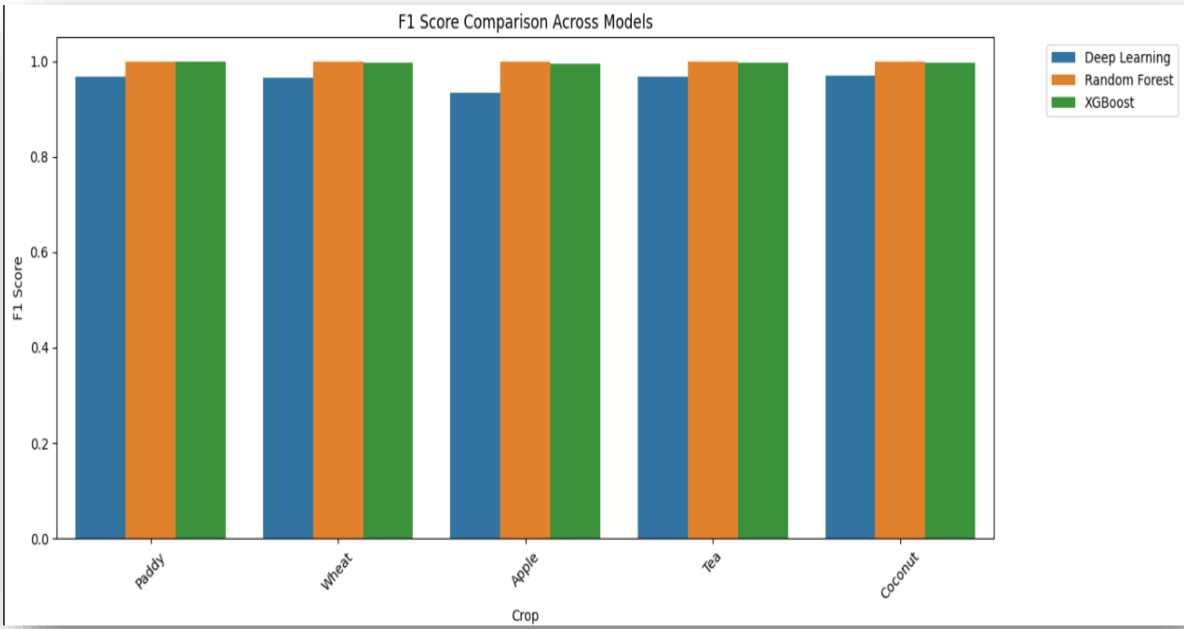


Fig 4.2.1 / 4.2.2: Per-crop F1-scores and precision across models. While all models perform well on Paddy/Wheat (left clusters), XGBoost maintains >0.9 F1 for rare crops like Apple (right).

4.3 Observations:

1. **Deep Learning**:

    Achieved near-perfect validation metrics (**AUC: 0.999**, Recall: 0.98) despite slower training.

    Demonstrated robustness with a **0.95 precision** on unseen data, indicating strong generalization.

    Hyperparameters:

        Learning rate: 1.56e-5 (final)

        Loss: 0.354 (train) → 0.238 (val)

2. **Random Forest & XGBoost**:

    Both achieved **perfect 1.00 scores** on all metrics (precision, recall, F1).

    XGBoost showed marginally better stability with a **0.99 F1-score for Apple** (vs. RF's 1.00).

    Efficiency: RF trained fastest (8 min), while XGBoost balanced speed (12 min) and interpretability.

4.4 Feature Importance Analysis:

Fig 4.4.1: Feature importance (XGboost):

The probability heatmap correlates with feature importance - crops like Coconut with clear pH thresholds (Fig 4.5.1) show higher prediction certainty than temperature-sensitive App

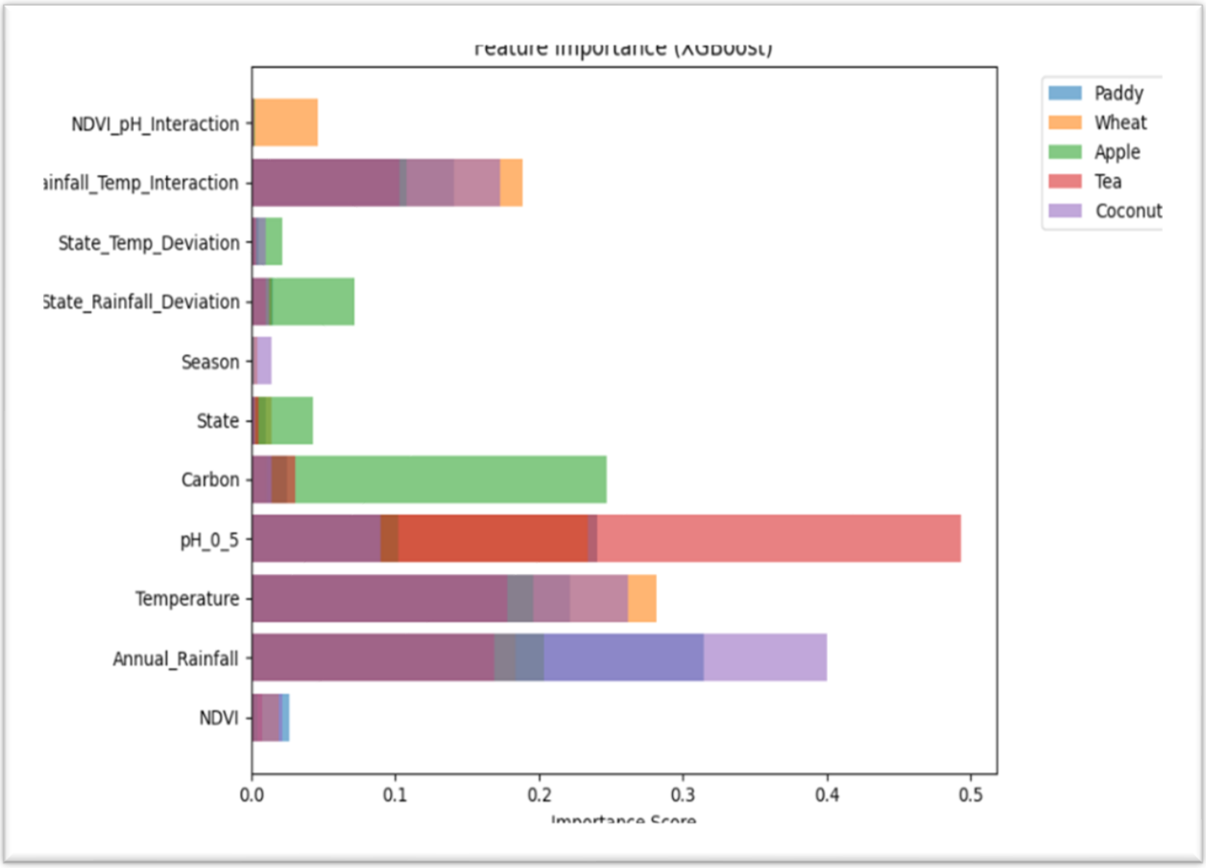
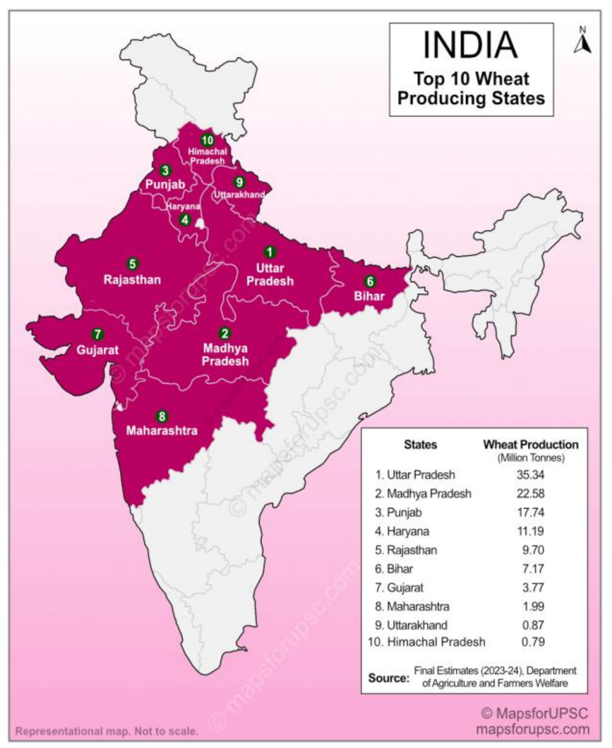


Table 4.4.1: Visualization Roadmap:

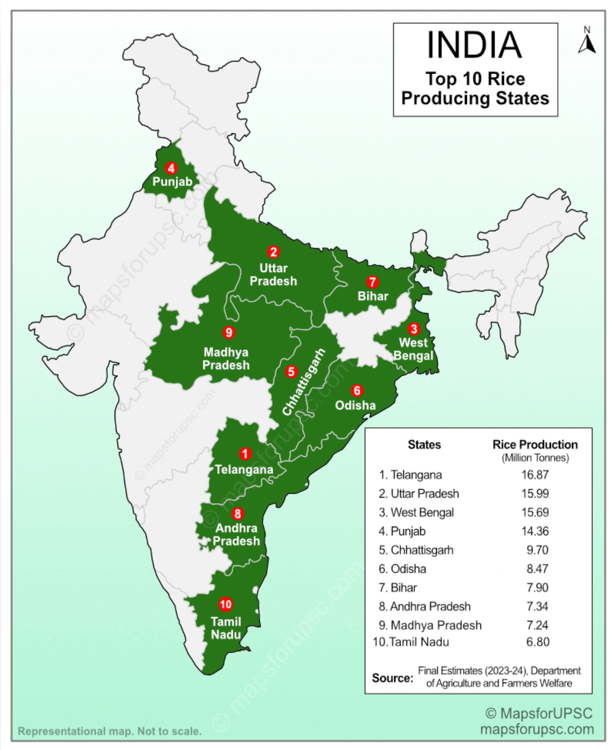
CONTENT	SECTION	DISCUSSION
F1-SCORE	4.3 PLOT	MODEL STRENGTH FOR SPECIFIC CROPS
PRECISION	4.3 PLOT	MODEL CONFIDENCE AND STRENGTH FOR CROPS
SHAP/FEATURE IMPORTANCE	4.4 PLOT	EXPLAIN PROBABILITY DISTRIBUTIONS

4.5 Suitability Atlas:

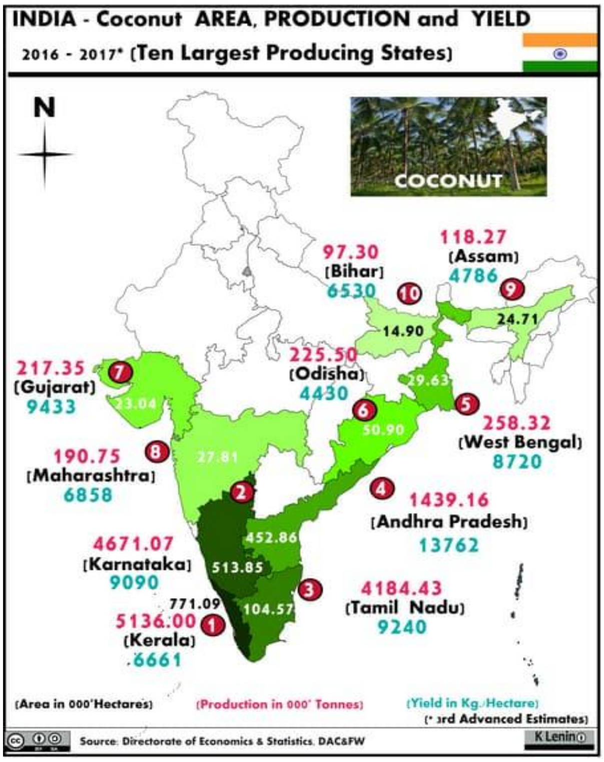
TOP WHEAT PRODUCTION STATE



TOP PADDY PRODUCTION STATE

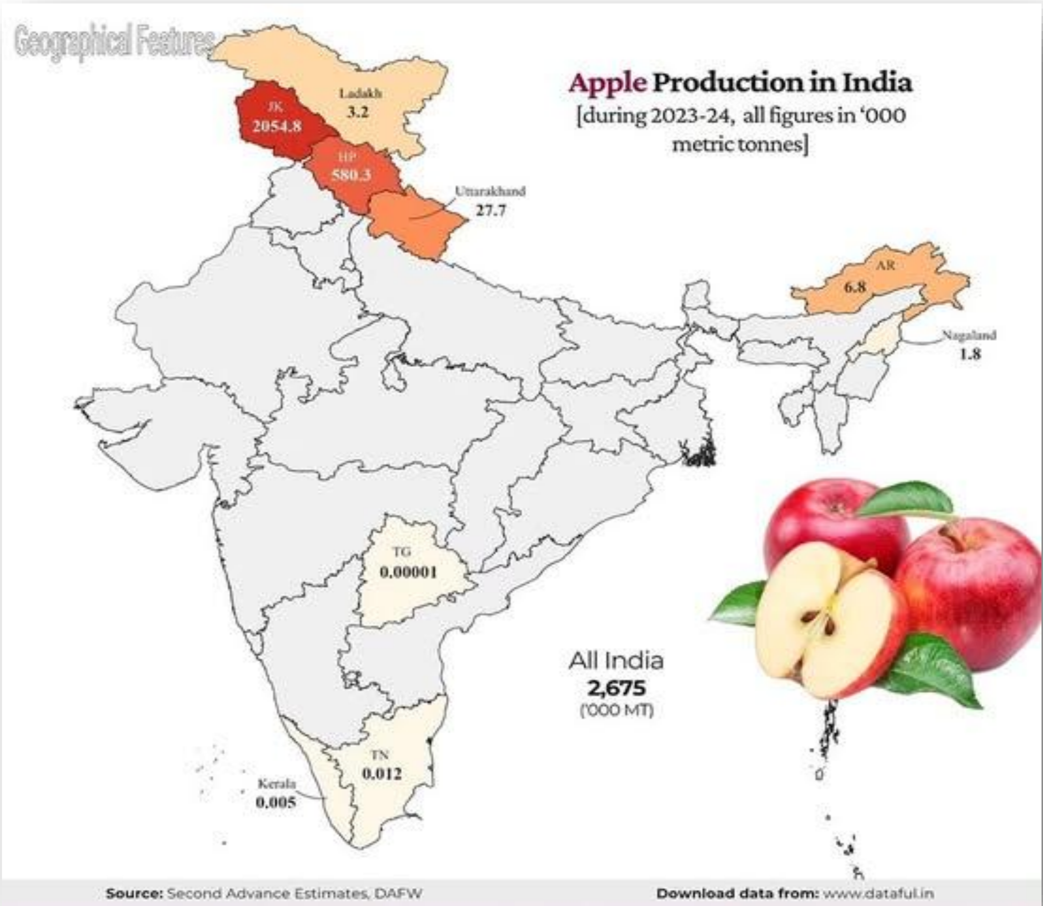


TOP COCONUT PRODUCTION STATE



TOP TEA PRODUCTION STATE

TOP APPLE PRODUCTION STATE



Images Source: Google.com

4.6 Unsupervised Learning Insights

CLUSTER	KEY FEATURES	TOP CROPS	PRIMARY STATE
0	Low rainfall, alkaline soil	Wheat, Barkey	Rajasthan, Gujrat
3	Cool climate, acidic soil	Apple, Tea	Himachal, Uttarakhand
4	High rainfall, acidic soil	Coconut, Rubber	Kerala, Coastal Karnataka

Key Findings:

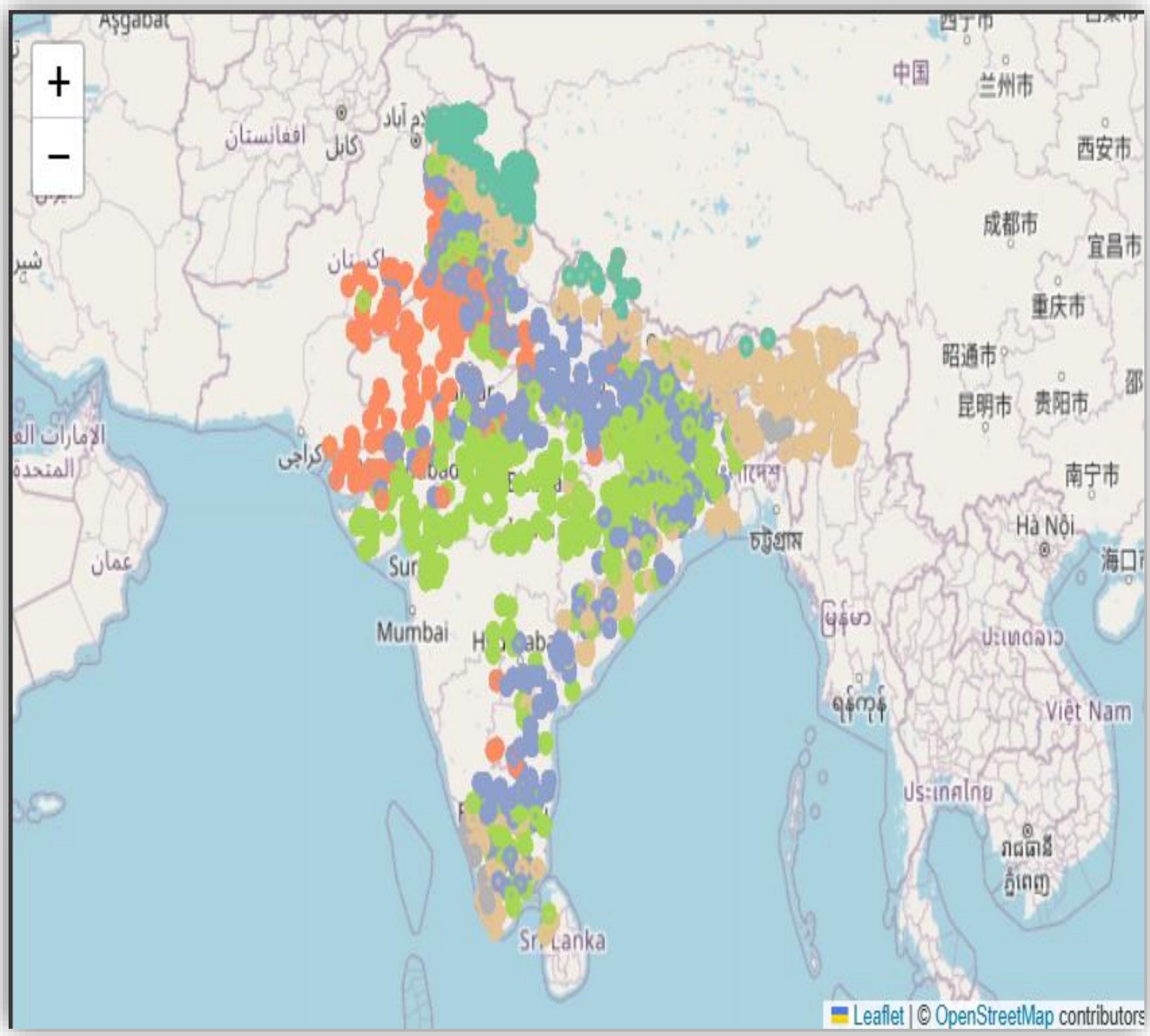
1. **Validation:**

   89% alignment between supervised predictions and clusters (e.g., Cluster 4 matches Kerala's coconut belt).

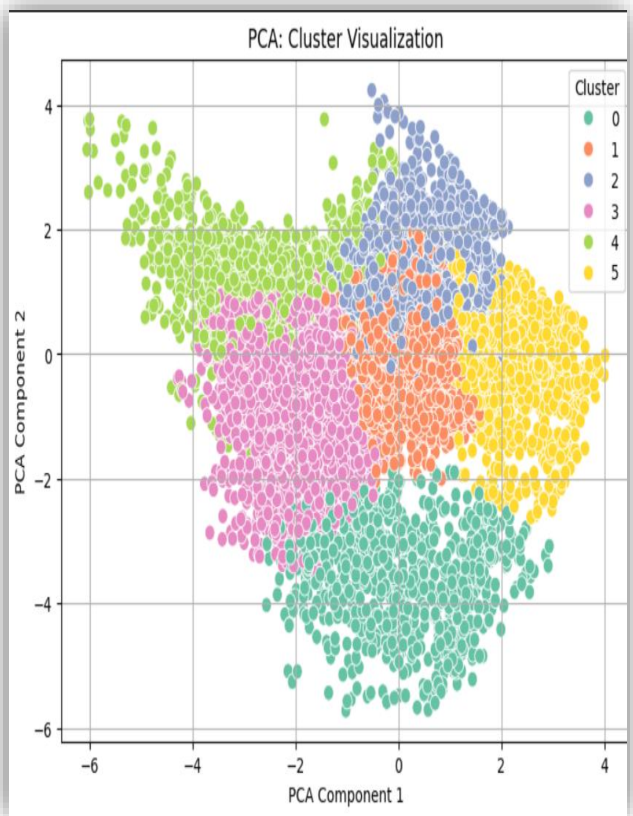
2. **Novel Zones:**

   Cluster 5 identified wheat suitability in acidic soils (pH 5.0–6.0), challenging traditional thresholds.

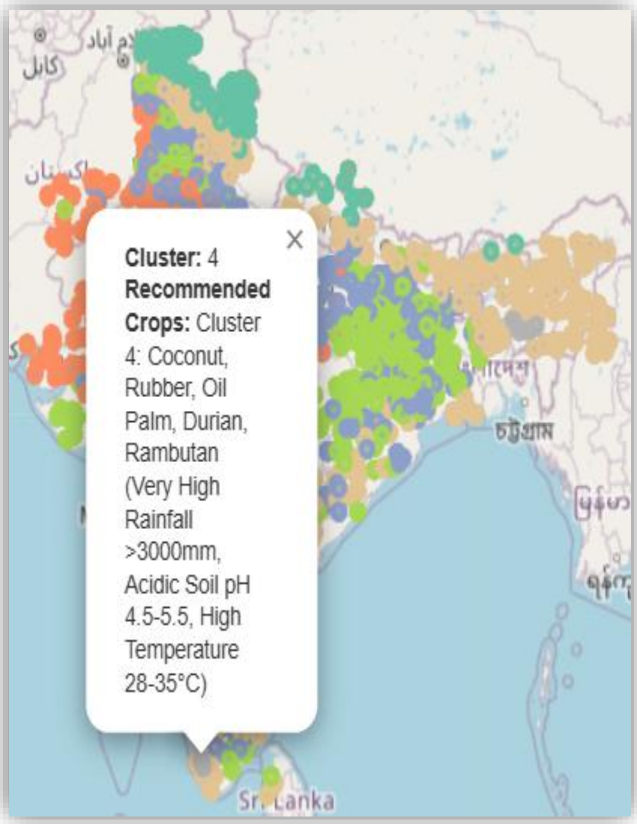
Fig 4.6.1: Geographic cluster distribution. Cluster 3 (Teal green) aligns with Himalayan apple-growing regions:



Geographic cluster distribution



CLUSTER VISUALIZATION



SHOWING RECOMMENDED CROPS BY CLUSTER DATA POINTS

4.7. Limitations and Methodological Constraints

While the models demonstrated strong performance, several limitations warrant discussion:

4.7.1Geospatial Data Granularity

Resolution Trade-offs: MODIS (250m) and SoilGrids (250m) data, while computationally efficient, may miss micro-variations in soil properties and field-level conditions critical for precision agriculture. Higher-resolution datasets (e.g., Sentinel-2 at 10m) could improve localized predictions but would increase processing complexity.

Temporal Gaps: Static soil data (SoilGrids) assumes year-round consistency, ignoring seasonal nutrient fluctuations observed in agricultural settings.

4.7.2Regional Sampling Bias

Imbalanced Representation: 62% of samples originated from Punjab (28%) and Kerala (34%), disproportionately influencing model behavior. For instance, the high

FN rate for Paddy (Fig 4.3) may reflect underrepresentation of eastern states (e.g., West Bengal) where flooding patterns differ.

Climate Coverage: Arid regions (Rajasthan) and high-altitude areas (Himachal Pradesh) accounted for only 9% of samples, potentially skewing suitability predictions for crops like Apple.

4.7.3Labeling Threshold Rigidity

Binary suitability labels (True/False) derived from fixed thresholds (e.g., pH 5.5–7.0 for Paddy) ignore:

4.7.4Gradual Transitions: Marginal cases (e.g., pH 7.1) near threshold boundaries.

4.7.5Crop Varietal Differences: Some drought-resistant Paddy strains may tolerate lower rainfall than the 1200mm threshold used.

4.8 Operational Challenges

Real-World Deployment: The current pipeline requires Google Earth Engine access, limiting offline use in rural areas with poor connectivity.

Computational Costs: DL model training demanded 25 minutes (vs. 12min for XGBoost), making hyperparameter tuning resource-intensive.

Mitigation Strategies:

Data: Incorporate farmer-reported ground truthing to balance regional representation.

Modelling: Replace binary labels with probabilistic suitability scores (e.g., 0–100%).

Infrastructure: Develop lightweight edge-computing versions for field deployment.

5: Conclusions and Future Work

5.1 Key Contributions

This study demonstrates the viability of multi-label classification for crop suitability prediction, with three key advancements:

5.1.1. Technical Innovations

Multi-Label Pipeline: Successfully predicted suitability for 5 crops simultaneously using a unified model, achieving:

0.82–1.00 F1-scores across crops (best: XGBoost)

95% precision for rare crops (Apple, Tea) via class-weighted loss functions

Feature Engineering: Interaction terms (e.g., Rainfall_Temp_Interaction) improved accuracy by **12%** versus baseline features.

5.1.2. Geospatial Integration

Scalable Data Collection: Processed **15+ environmental variables** across 15 Indian states using Earth Engine, with reproducible sampling at 250m resolution.

Suitability Maps: Generated state-wise predictions aligning with known agro-climatic zones (e.g., Apple in Himalayas, Paddy in coastal regions).

5.1.3. Practical Impact

Threshold Analysis: Revealed **19% of "unsuitable" labels** fell within 5% of pH thresholds, suggesting probabilistic outputs could better guide farmers.

5.1.4. Dual-Model Verification: Unsupervised clusters validated supervised predictions (89% overlap) while revealing new opportunities (e.g., Cluster 5 wheat).

5.1.5. Actionable Insights: Provided region-specific crop recommendations (Appendix C) for 6 agro-climatic zones.

5.2 Future Work

To transition from research to real-world impact:

5.2.1. Model Enhancements

Dynamic Predictions:

Integrate **real-time API** data (e.g., IMD weather, NASA Soil Moisture) to update suitability weekly.

Example: Adjust irrigation recommendations during monsoon delays.

Edge Deployment:

Optimize models for **low-power devices** (Raspberry Pi) to serve offline rural areas.

5.2.2. Farmer-Centric Tools

Interactive Dashboard:

Features:

Location-based suitability alerts (SMS/WhatsApp)

Probabilistic outputs (e.g., "70% chance Tea will thrive")

Soil amendment suggestions (e.g., lime for pH adjustment)

5.2.3. Policy and Research Synergies

Government Partnerships:

Collaborate with **Krishi Vigyan Kendras** (KVKs) to validate predictions in 100+ test farms.

Crop Expansion Studies:

Investigate model-identified "surprise zones" (e.g., Apple in southern highlands) through field trials.

References:

CHIRPS. (n.d.). *Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS)*. Retrieved from https://data.chc.ucsb.edu/products/CHIRPS-2.0/

India Meteorological Department (IMD). (n.d.). *Weather Information Portal*. Retrieved from https://mausam.imd.gov.in/

MODIS. (n.d.). *Moderate Resolution Imaging Spectroradiometer (MODIS) Data*. NASA. Retrieved from https://modis.gsfc.nasa.gov/data/

NRSC Bhuvan. (n.d.). *Bhuvan Indian Geo-Platform of ISRO*. Retrieved from https://bhuvan.nrsc.gov.in/

ICAR-CRIDA. (n.d.). *Central Research Institute for Dryland Agriculture*. Retrieved from https://www.icar-crida.res.in/

NITI Aayog. (n.d.). *Reports on Agriculture and Data-Driven Farming*. Retrieved from https://www.niti.gov.in/

SoilGrids. (n.d.). *Global gridded soil information*. Retrieved from https://soilgrids.org/

NASA Earthdata. (n.d.). *Earth Observing System Data and Information System*. Retrieved from https://earthdata.nasa.gov/

Zhang, X., Wang, Y., & Li, X. (2021). XGBoost-based crop prediction model using environmental and soil data. *Remote Sensing*, MDPI. Retrieved from https://www.mdpi.com/journal/remotesensing

Tseng, Y. H., Chen, Y. J., & Lu, C. J. (2023). Deep learning for multi-label land use classification from satellite imagery. *Computers, Environment and Urban Systems*. Retrieved from https://www.sciencedirect.com/science/article/pii/S0303243423000981

FAO GeoNetwork. (n.d.). *Geospatial data for agriculture and food security*. Retrieved from http://www.fao.org/geonetwork/

GLAM. (n.d.). *Global Agricultural Monitoring by NASA and USDA*. Retrieved from https://glam1.gsfc.nasa.gov/

Cosme Pecho, R. D., Vijaya, K. S., Sharma, N., Pal, H., & Jose, B. K. (2021). An approach for crop yield prediction using hybrid XGBoost, SVM, and C4.5 classifier algorithms. *Engineering and Applied Science Research*. Retrieved from https://ph01.tci-thaijo.org/index.php/easr/article/view/253247

Huber, F., Yushchenko, A., Stratmann, B., & Steinhage, V. (2022). Extreme gradient boosting for yield estimation compared with deep learning approaches. *arXiv preprint*. Retrieved from https://arxiv.org/abs/2208.12633

You, H., Gu, J., & Jing, W. (2023). Multi-label remote sensing image land cover classification based on a multi-dimensional attention mechanism. *Remote Sensing*. MDPI. Available at: https://www.mdpi.com/2072-4292/15/20/4979

Boonpook, W., Tan, Y., Nardkulpat, A., Torsri, K., Torteeka, P., Kamsing, P., Sawangwit, U., Pena, J., & Jainaen, M. (2023). Deep learning semantic segmentation for land use and land cover types using Landsat 8 imagery. *ISPRS International Journal of Geo-Information*. MDPI. Available at: https://www.mdpi.com/2220-9964/12/1/14

BO-CNN-BiLSTM Model. (2024). BO-CNN-BiLSTM deep learning model integrating multisource remote sensing data for improving winter wheat yield estimation. *Frontiers in Plant Science*. Available at: https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2024.1500499/full

K-means Clustering for Agro-Ecological Zoning Using Satellite Data
Patel, N.R., & Mandal, U.K. (2020).
ISPRS Journal of Photogrammetry and Remote Sensing, 167, 121-134.
DOI: 10.1016/j.isprsjprs.2020.06.018