

Machine Learning Applications for Precision Agriculture

Thesis Submitted

By

Abhirup Paria

Doctor of Philosophy (Engineering)

Department of Computer Science and Engineering
Faculty Council of Engineering & Technology
Jadavpur University
Kolkata, India

2025

Machine Learning Applications for Precision Agriculture

Thesis Submitted

By

Abhirup Paria

Doctor of Philosophy (Engineering)

Under the supervision of

Dr. Sarmistha Neogy

Department of Computer Science and Engineering

Jadavpur University

Kolkata, West Bengal–700032, India

Dr. Arindam Giri

Department of Computer Science and Engineering

Haldia Institute of Technology

Haldia, West Bengal–721657, India

2025

JADAVPUR UNIVERSITY

KOLKATA-700032

Reg No. 1022204010

Index No. 112/22/E

1. Title of the thesis:

Machine Learning Applications for Precision Agriculture

2. Name, Designation and Institution of the Supervisors:

(a) Dr. Sarmistha Neogy

Professor

Department of Computer Science and Engineering

Jadavpur University

Kolkata, West Bengal, India, 700032

(b) Dr. Arindam Giri

Associate Professor

Department of Computer Science and Engineering

Haldia Institute of Technology

Haldia, West Bengal, India, 721657

3. List of Publications:

(a) Journal:

Published

- (i) **A. Paria**, A. Giri, S. Dutta, and S. Neogy “Smart irrigation for coriander plant: Saving water with AI and IoT.” *Water Resources Management* (2025): 1-17.
- (ii) **A. Paria**, S. Neogy, A. Giri, S. Dutta, P. Sureephong and S. D. Ravana, “Eco-Friendly and Explainable Hybrid Machine Learning Models for Sustainable Crop and Fertilizer Recommendation System.” *IEEE Transactions on Sustainable Computing*, vol. 11, no. 3, pp. 291–301, May–June 2026.

Communicated

- (i) **A. Paria**, S. Neogy, A. Giri and S. Dutta “Enhancing Farmer Decision-Making: An Explainable Ai Framework for Reliable Crop Yield Prediction.” *IEEE Transactions on Emerging Topics in Computational Intelligence*.

(b) Conference:

- (i) **A. Paria**, A. Giri, S. Dutta, and S. Neogy “AcoNeural: A Hybrid Neural Network Model for Predicting Daily Rainfall”. *International Conference on Systems and Technologies for Smart Agriculture* . Singapore: Springer Nature Singapore, 2023.
- (ii) **A. Paria**, S. Neogy, A. Giri and S. Dutta “Enhanced Soil Fertility Classification Using ACO-Optimized XgBoost and Explainable AI.” *6th International Conference on Frontiers in Computing and Systems (COMSYS-2025)*. (Accepted)

4. List of Patents:

- (i) **A. Paria**, S. Neogy, A. Giri, S. Dutta and C. Ghorai “Evapotranspiration based Smart Irrigation System for Optimized Water Conservation” *Indian Design Patent*, Date of Filing: 27.08.2025, Application no: 202531081924, Date of Publication: 05.09.2025, Publication No: 36/2025.

5. List of Presentations in National / International Conference:

- (i) **A. Paria**, A. Giri, S. Dutta, and S. Neogy “AcoNeural: A Hybrid Neural Network Model for Predicting Daily Rainfall”. *International Conference on Systems and Technologies for Smart Agriculture*. Kolkata, December 19-20, 2023.
- (ii) **A. Paria**, S. Neogy, A. Giri and S. Dutta “Enhanced Soil Fertility Classification Using ACO-Optimized XgBoost and Explainable AI.” *6th International Conference on Frontiers in Computing and Systems (COMSYS-2025)*. Poland, September 25-27, 2025.

Statement of Originality

I, **Mr. Abhirup Paria** registered on **17th May, 2022** do hereby declare that this thesis entitled **“Machine Learning Applications for Precision Agriculture”** contains literature survey and original research work done by the undersigned candidate as part of Doctoral studies.

All information in this thesis have been obtained and presented in accordance with existing academic rules and ethical conduct. I declare that, as required by these rules and conduct, I have fully cited and referred all materials and results that are not original to this work.

I also declare that I have checked this thesis as per the **“Policy on Anti Plagiarism, Jadavpur University, 2019”**, and the level of similarity as checked by iThenticate software is **2%** .

Signature of Candidate: *Abhirup Paria*

Date : 14-07-2026

Certified by Supervisor(s):

(Signature with date, seal)

1. *Sarmistha Neogy, 14.07.2026*

(Dr. Sarmistha Neogy,

Professor,

Department of Computer Science and Engineering

Jadavpur University.)

Professor
Computer Sc. & Engg. Department
Jadavpur University
Kolkata-700032

2. *Arindam Giri*

(Dr. Arindam Giri,

Associate Professor,

Department of Computer Science and Engineering

Haldia Institute of Technology.)

Associate Professor
Dept. of Computer Science & Engineering
HALDIA INSTITUTE OF TECHNOLOGY

CERTIFICATE FROM THE SUPERVISOR/S

This is to certify that the thesis entitled "Machine Learning Applications for Precision Agriculture" submitted by Mr. Abhirup Paria, who got his name registered on 17th May 2022, for the award of Ph.D.(Engineering) degree of Jadavpur University, is absolutely based upon his own work under the supervision of Dr. Sarmistha Neogy and Dr. Arindam Giri and that neither his thesis nor any part of the thesis has been submitted for any degree/diploma or any other academic award anywhere before.

Sarmistha Neogy.

14.07.2026

Dr. Sarmistha Neogy,
Professor,
Department of Computer
Science and Engineering
Jadavpur University.
(Supervisor)

Professor
Computer Sc. & Engg. Department
Jadavpur University
Kolkata-700032

Arindam Giri *14.07.26*

Dr. Arindam Giri,
Associate Professor,
Department of Computer
Science and Engineering
Haldia Institute of Technology.
(Co-Supervisor)

Associate Professor
Dept. of Computer Science & Engineering
HALDIA INSTITUTE OF TECHNOLOGY

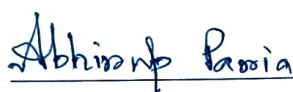
Acknowledgment

First and foremost, I extend my deepest gratitude and heartfelt thanks to my esteemed supervisors, Dr. Sarmistha Neogy, Professor at Jadavpur University, and Dr. Arindam Giri, Associate Professor at Haldia Institute of Technology. Their unwavering guidance, invaluable insights, dedicated assistance, and continuous encouragement have been instrumental in steering and completing this research, contributing significantly to the fulfillment of my Doctor of Philosophy. Without their enthusiasm, profound perspectives on research, and unwavering support, this journey toward my PhD would not have been possible.

I am sincerely grateful to the other members of my thesis committee for their invaluable and constructive feedback, which has substantially enhanced the quality of this thesis. Additionally, I wish to acknowledge the commendable contributions made by my wife, Dr. Ruma Das, whose work has had a significant impact on this research.

I express my deepest gratitude and thanks and to my parents, Sri Jadupati Paria and Mrs. Manjusree Paria for their love, encouragement and support. I am indebted to my parents for their sacrifice to make me educated. I strongly recognize my father and mother's effort put towards possessing my mastery in all subjects. I express my gratitude to my sister, Mrs. Pamita Das Paria for love and best wishes. I could never forget my childhood with my sister.

I extend my thanks to the Department of Computer Science and Engineering at Jadavpur University for providing me with this exceptional opportunity and for their exceptional support and guidance throughout my research journey.

 14-07-2026

Abhirup Paria
Jadavpur University

Abstract

Agriculture is facing increasing challenges due to global climate shift, water scarcity, and the global need for sustainable food production. Traditional farming methods are often unable to respond efficiently to these dynamic environmental conditions. The thesis presents an integrated system that combines Internet of Things (IoT), Hybrid Machine Learning (HML), and Explainable Artificial Intelligence (XAI) to enhance decision-making in every stage of modern agriculture. The research moves from understanding environmental conditions to building intelligent, sustainable, and transparent systems that assist agriculturalists in making evidence-based decisions.

The work begins with the development of an intelligent rainfall prediction model, recognizing that understanding the environment is the foundation of effective farming. A novel Ant Colony Optimization-based Neural Network (ACO-NN) is presented to forecast daily rainfall in Alipurduar, West Bengal. By integrating key weather parameters and optimizing the neural architecture through ACO, the method achieves high prediction accuracy and significantly reduces errors. Accurate rainfall forecasting enables farmers to plan irrigation schedules and crop management strategies well in advance, marking a major step toward precision agriculture.

Building on this environmental understanding, the work presents an intelligent irrigation system that combines IoT-based sensors with HML models optimized using a Genetic Algorithm (GA). The system predicts the precise water requirements of coriander plants across different growth stages using solar radiation data captured by the BH1750 sensor to estimate evapotranspiration (EV_{T0}). The best-performing hybrid model achieves an accuracy of 0.98%, ensuring efficient irrigation and water conservation. A mobile-based monitoring system further allows farmers to control irrigation remotely, reducing both water wastage and manual labor. The work then focuses on improving soil health through soil fertility classification. A hybrid ACO-XGBoost model is designed to classify soil fertility based on micro and macronutrients, achieving 98% accuracy. The integration of XAI techniques helps to explain the model's decisions, allowing farmers to understand which nutrients influence soil fertility the most. This interpretability turns complex soil data into practical guidance for fertilizer management and soil improvement.

To promote environmentally responsible farming, an eco-friendly hybrid machine learning (EHML) system is developed for sustainable crop and fertilizer recommendation. Models like CatBoost and XGBoost are optimized with nature-inspired algorithms, achieving accuracies above 98%. XAI methods reveal the reasoning behind each recommendation, while environmental indicators such as CO_2 emissions and energy use are analyzed to promote green agricultural practices. A web application is also built to deliver real-time, interpretable crop and fertilizer recommendations to farmers.

Finally, an Explainable AI framework for a crop yield forecasting system is proposed to build transparency and trust in AI-driven decisions. The framework combines global, local, and counterfactual explanations to show how various factors affect yield. Random Forest and AdaBoost models achieve the best results, while a new XAI-based feature selection method highlights the most influential factors. This framework enables farmers to understand and trust AI predictions.

Contents

Acknowledgment	ii
Abstract	iii
List of Figures	ix
List of Tables	xi
Abbreviations	xiii
1 Introduction	1
1.0.1 Artificial Intelligence and IoT in PA	2
1.0.2 ML Algorithms	5
1.1 Applications of ML in PA	6
1.1.1 Forecasting Soil Characteristics and Climate Conditions	7
1.1.2 Crop Yield Forecasting	9
1.1.3 Drip Irrigation	13
1.2 Application of IoT in PA	16
1.2.1 Use of IoT Sensors in PA	18
1.3 Motivation of the Work	19
1.4 Aims and Objectives	20
1.4.1 Main Aim	20
1.4.2 Specific Objectives	20
1.5 Research Queries	21
1.6 Extent of the Research	21
1.6.1 Geographical and Time-Based Extent	21
1.6.2 Crop-Specific Focus	22
1.6.3 Model Development and Evaluation	22
1.6.4 XAI Framework	22
1.6.5 Data and Technology Stack	22
1.7 Thesis Structure	23
2 The Foundation of Precision Agriculture – Understanding the Environment	24
2.1 The Role of Rainfall Prediction in Agriculture	25
2.2 AI and ML in Weather Prediction	26
2.3 Practical Applications in Precision Agriculture	27
2.4 Background of the Study	28

2.5	Case Study: Alipurduar, West Bengal	30
2.6	Methodology	30
2.6.1	Artificial Neural Network (ANN)	30
2.6.1.1	Narrow Neural Network(NNN)	32
2.6.1.2	Medium Neural Network(MNN)	32
2.6.1.3	Wide Neural Network(WNN)	32
2.6.1.4	Bilayered Neural Network(BNN)	33
2.6.1.5	Trilayered Neural Network(TNN)	33
2.6.2	Ant Colony Optimization	33
2.6.2.1	State Transition Rule(STR)	33
2.6.2.2	Global Updating Rule (GUR)	34
2.6.2.3	Local Updating Rule (LUR)	34
2.6.3	ACO-ANN Hybrid Model	35
2.6.4	Data Pre-processing	35
2.6.4.1	Feature Selection Using ACO	36
2.6.5	Experimentation	37
2.7	Result Analysis	38
2.8	Chapter Summary	40
3	From Knowledge to Action – Smart Irrigation for a Sustainable Future	42
3.1	Challenges of Traditional Irrigation Systems	43
3.2	Background of the Study	44
3.3	Contributions of the present work	46
3.4	The Estimation of EV_{T0} using Climate-Based Methodologies	46
3.4.1	PM Model	47
3.4.2	Proposed Method to Estimate Solar Radiation	47
3.5	Crop Water Requirement Estimation	49
3.6	Artificial neural networks (ANNs)	49
3.7	Genetic Algorithm (GA)	49
3.8	ML Hybridization using GA and ANN	50
3.9	Proposed Framework for Intelligent Irrigation	50
3.9.1	Data Acquisition Module	51
3.9.2	EV_{T0} Prediction Module	53
3.9.3	Real-Time Monitoring using Android Application Module	55
3.9.4	IoT-based Irrigation Pump Controller Module	55
3.10	Data Collection	55
3.10.1	Experimental Testbed and System Configuration	56
3.11	Experimentation	57
3.12	Performance Evaluation	58
3.13	Results Comparison	60
3.14	Real-time Smart Irrigation System	61
3.15	Chapter Summary	63
4	The Soil Beneath – Classifying Soil Fertility for Better Crop Choices	67
4.1	Reasons for Soil Fertility Classification	69
4.2	Methodology	70
4.2.1	Dataset Description	70

4.2.2	Machine Learning Algorithm	70
4.2.2.1	SHapley Additive exPlanations (SHAP)	70
4.3	Results and Discussion	73
4.3.1	Explanation of Correct and Incorrect Classifications Using XAI	75
4.3.1.1	Correct Classification:	75
4.3.1.2	Incorrect Classification:	76
4.4	Chapter Summary	77
5	Building for the Future – Sustainable Crop and Fertilizer Recommendations	78
5.1	Literature Survey	79
5.2	Predictive Analysis	81
5.3	Hybrid ML Models	81
5.4	Proposed Crop Recommendation Model	82
5.4.1	Workflow of Proposed Recommendation System	82
5.4.2	Eco-friendly AI Metrics and CodeCarbon	83
5.5	Crop-Wise Soil Nutrients-Based Fertilizer Recommendation System	83
5.6	Experiment Design	84
5.6.1	Dataset Description & Data Pre-Processing	84
5.6.2	Experimental Environment	85
5.6.3	Evaluation Metrics	86
5.7	Results	86
5.8	Results Comparison with Existing Work	88
5.9	Sustainable AI Assessment	89
5.10	The Significance of Features Towards Recommendation Through XAI	90
5.10.1	Accurate and Inaccurate Recommendation Explained with XAI	90
5.10.1.1	Accurate Recommendation	90
5.10.1.2	Inaccurate Recommendation	91
5.11	A Farmer-Friendly Web Application for Crop and Fertilizer Recommendation	92
5.12	Chapter Summary	93
6	The Final Frontier – Enhancing Farmer Decision-Making with Explainable AI	96
6.1	Literature review	98
6.2	Methodology	98
6.2.1	Proposed framework	98
6.2.2	A novel feature selection method using SHAP	99
6.2.3	Dataset description	99
6.2.3.1	Dataset 1	100
6.2.3.2	Dataset 2	100
6.2.3.3	Dataset Preprocessing	101
6.2.4	Model selection and prediction	101
6.2.5	Model explainability	101
6.2.5.1	Local Interpretable Model Explanation (LIME)	101
6.2.6	Evaluation metrics	106
6.3	Performance evaluation	106

6.3.1	Performance of various blackBox AI models	106
6.3.2	Performance comparisons with existing XAI frameworks	107
6.3.3	Effects on various combinations of features	107
6.3.4	Effects on tuning various hyper-parameters	109
6.3.5	Comparison with various benchmark feature selection techniques across various models	110
6.3.6	Quantitative Validation of the Proposed XAI Framework	110
6.4	The significance of features through XAI	112
6.4.1	Global summary plot (GSP)	113
6.4.2	Beeswarm summary plot (BSP)	113
6.5	Comparing XAI models with black-box AI methods	114
6.5.1	Dataset 1	114
6.5.1.1	Accurate prediction	114
6.5.1.2	Inaccurate prediction	115
6.5.2	Dataset 2	115
6.5.2.1	Accurate prediction	115
6.5.2.2	Inaccurate prediction	116
6.6	Counterfactual analysis of influential features for prediction adjustment .	117
6.7	Chapter Summary	118
7	Conclusion	121
7.1	Principal Contributions	122
7.1.1	Methodological Contribution	122
7.1.2	System-level Contribution	122
7.2	Implications of the Research	123
7.2.1	For Farmers	123
7.2.2	For Environment	123
7.3	Limitations of the Work	123
7.4	Recommendations for Future Research	123
	Bibliography	125

List of Figures

1.1	Precision Agriculture	2
1.2	AI techniques for PA	3
1.3	Workflow of a ML model	6
1.4	Various kinds of ML algorithms	7
1.5	(a) Conventional agriculture (b) PA	7
2.1	Experimentation block diagram.	31
2.2	ACO-ANN hybrid model flowchart.	36
2.3	Abstract ANN model with Selected Features and output	38
2.4	Result Comparison of various Hybrid Models.	39
3.1	Sunshine duration with respect to the five cloudiness classifications. . . .	47
3.2	GA Workflow: Key Steps and Operations.	50
3.3	Integration of GA and ANN.	51
3.4	The Schematic Arrangement of Sensor Nodes in the Crop field.	51
3.5	(a) Actual vs HML4 Predicted Water Requirement for Vegetative Growth Cycle, (b) Actual vs HML4 Predicted Water Requirement for Bolting Cycle, (c) Actual vs HML4 Predicted Water Requirement for Flowering and Seed development Cycle	60
3.6	(a) Correlation between the forecasted EV_{T0} and the observed EV_{T0} for training, (b) Correlation between the forecasted EV_{T0} and the observed EV_{T0} for validation, (c) Error Histogram, (d) Correlation between the forecasted EV_{T0} and the observed EV_{T0} for testing, (e) Correlation between the forecasted EV_{T0} and the observed EV_{T0} for all data, (f) MSE between the predicted and actual values for the training, validation, and testing data	61
3.7	Experimentation block diagram	62
3.8	(a) Sensor node at Region 1, (b) Sensor node at Region 2, (c) Sensor node at Region 3, (d) Gateway(Raspberry Pi), (e) In-Depth view of each Sensor Node, (f) Irrigation Pump with power supply and relay module . .	63
3.9	(a) Home Screen, (b) Graph for Temperature, (c) Graph for Atmospheric Pressure, (d) Graph for Specific Humidity, (e) Graph for Relative Humidity, (f) Graph for Net Radiation, (g) Screen for Pump on, (h) Screen for Pump Off	65
4.1	Workflow for Soil Fertility Classification	69
4.2	SHAP Waterfall Plots for Feature Contributions: (a) Correctly Classified Sample, (b) Incorrectly Classified Sample	77

5.1	Workflow of the sustainable crop & fertilizer Recommendation system using hybrid ML, XAI, and eco-friendly AI	82
5.2	Confusion matrices for best model in seven districts, illustrating the true vs. predicted class distributions	87
5.3	A comparison of electricity consumption, CO_2eq emissions, emission rate, and CPU energy usage for the best performing model (XgBoost-PPSO) across seven districts	89
5.4	SHAP summary plots: feature importance analysis for crop recommendations using XAI for each seven districts	91
5.5	LIME and SHAP interpretations for model predictions in Hooghly district: (a), (c) LIME feature importance explanations for correctly and incorrectly classified instances, and (b), (d) Waterfall plots showing feature contributions to the model predictions	92
5.6	(a) Breadboard setup with NPK sensor, RS485 module, OLED display and ESP32 micro-controller unit for data acquisition. (b) NPK sensor inserted near sunflower roots for real-time soil nutrient monitoring. (c) Distributed sensor nodes deployed along the sunflower plantation rows. (d) Enhanced sunflower growth observed after applying NPK nutrients based on sensor data feedback.	93
5.7	(a) AI-powered crop recommendation web application: analyzing macro and micro-nutrient levels to suggest optimal crop selection, (b) AI-based fertilizer recommendation system: web application for optimized crop fertilizer management	94
6.1	Proposed XAI framework for crop yield prediction	99
6.2	Global summary plot of feature importance for each AI models (generated Using SHAP) for dataset 1: (a) RF, (b) AdaBoost, (c) CatBoost, (d) LGBM, (e) KNN, (f)DT	115
6.3	Global Summary Plot of Feature Importance for each AI models (Generated Using SHAP) for dataset 2: (a) RF, (b) AdaBoost, (c) CatBoost, (d) LightGBM, (e) KNN, (f)DT	116
6.4	Beeswarm summary plot of feature importance for each AI models (generated using SHAP) for dataset 1: (a) RF, (b) AdaBoost, (c) CatBoost, (d) LGBM, (e) KNN, (f) DT	117
6.5	Beeswarm Summary Plot of Feature Importance for each AI models (Generated Using SHAP) for dataset 2: (a) RF, (b) AdaBoost, (c) CatBoost, (d) LightGBM, (e) KNN, (f) DT	118
6.6	LIME local and waterfall Plot for local explanations on dataset 1: (a, b) Accurate Prediction, (c, d) Inaccurate Prediction	119
6.7	LIME local and waterfall Plot for local explanations on Dataset 2: (a, b) Accurate Prediction, (c, d) Inaccurate Prediction	119
6.8	(a) Counterfactual explanations plot for Dataset 1, (b) Counterfactual explanations plot for Dataset 2	120

List of Tables

1.1	Various ML Algorithms	4
1.2	Required nutrients for plants	8
1.3	Overview of ML Applications in Soil and Climate Parameter Prediction	11
1.4	Overview of ML Applications in crop yield prediction	14
1.5	Overview of ML Applications for drip irrigation	17
1.6	Various IoT sensors used in PA	18
2.1	Best Architecture Configuration	37
2.2	Results comparisons of various Regression Models.	39
2.3	Results of comparison of various hybrid models.	39
2.4	Results comparisons of various Wrapper Based Feature Selection Methods	40
3.1	Feature comparisons of proposed work with existing works	45
3.2	Summary of various Sensors and Gateway Used	52
3.3	Acronyms used in the Algorithm	52
3.4	Dataset of Vegetative Growth Cycle[August, 2023]	55
3.5	Dataset of Bolting Cycle[September,2023]	56
3.6	Dataset of Flowering and Seed Development Cycle[October, 2023]	56
3.7	Hardware and Software Configuration of the Smart Irrigation Testbed	57
3.8	GA Variables and their Values	57
3.9	Results for Vegetative Growth, Bolting, Flowering and Seed Development Cycle	59
3.10	Results Comparisons with Existing Work	60
3.11	Models & their Best Genes	66
4.1	Evaluation of ML models' performance for different classes	73
4.2	Results of ACO optimized ML models' performance for different classes	75
4.3	Results of comparison with other implemented work.	76
5.1	Comparative feature analysis of proposed work with previous literature	80
5.2	Nutrient Factors for Fertilizer Recommendation for various crops	85
5.3	The performance of the CatBoost model optimized with SSA are reported, as CatBoost-SSA consistently delivers the best results across all districts	88
5.4	Performance of CatBoost and XgBoost models with benchmark optimizers GA and PSO	88
5.5	Results comparison with existing work	89
5.6	Summary of NPK sensor data collected in a sunflower garden from 30th January 2025 to 5th April 2025, indicating levels of NPK in kg/ha.	94

6.1	Result of black-box AI techniques using all features from Dataset 1	106
6.2	Result of black-box AI techniques using all features from Dataset 2	106
6.3	Performance comparisons of proposed XAI-framework with existing frame- works	107
6.4	Performance of AI models on Dataset 1 using top selected features through proposed SHAP-based feature selection and ranking method	108
6.5	Feature ranking for Dataset 1	108
6.6	Feature ranking for Dataset 2	109
6.7	Performance of AI models on Dataset 2 using top selected features through proposed SHAP-based feature selection and ranking method.	109
6.8	Effect of hyper-parameter tuning for the AI models on dataset 1	111
6.9	Effect of hyper-parameter tuning for the AI models on dataset 2	112
6.10	Performance comparisons of proposed Shap-based feature selection with benchmark feature selection technique using all features from Dataset 1 .	113
6.11	Performance comparisons of proposed Shap-based feature selection with benchmark feature selection technique using all features from Dataset 2 .	114

Abbreviations

Acronym	What (it) Stands For
AI	A rtificial I ntelligence
ACO	A nt C olony O ptimization
ARIMA	A utoregressive I ntegrated M oving A verage
AdaBoost	A daptive B oosting
AWS	A mazons W eb S ervices
AICRP	A ll I ndia C oordinated R esearch P roject
AD-PO	A daptive D ynamic P uma O ptimizer
ANN	A rtificial N eural N etwork
ANFIS	A daptive N euro F uzzy S ystem
B	B oron
BO	B ayesian O ptimization
BSE	B ovine S erum E xpenditure
BNN	B ilayered N eural N etwork
BRT	B oosted R egression T ree
CatBoost	C ategorical B oosting
CNN	C onvolutional N eural N etworks
CGSAAN	C ombined G raph S ample and A ggregate A ttention N etwork
C	C arbon
Ca	C alcium
Co	C obalt
Cu	C opper
Cl	C hlorine
Cr	C hromium
DL	D eep L earning
DT	D ecision T ree
CO₂eq	C arbon D ioxide E quivalent
DNN	D eep N eural N etwork
ELM	E xtrême L earning M achine
EV_{T0}	E vapotranspiration (Reference)
EN	E lastic N et
EC	E lectrical C onductivity

EDA	Exploratory Data Analysis
F1-score	F1 Score
FS	Feature Selection or Forward Selection
FA	Firefly Algorithm
Fe	Iron
FF-BP	Feedforward Backpropagation
FL	Fuzzy Logic
FAO	Food and Agriculture Organization
GBM	Gradient Boosting Machines
GLM	Generalized Linear Model
GA	Genetic Algorithm
GPR	Gaussian Process Regression
GUR	Global Updating Rule
GEP	Gene Expression Programming
GWO	Grey Wolf Optimizer
HGSO	Hunger Games Search Optimization
HML	Hybrid Machine Learning
HHO	Harris Hawks Optimization
H	Hargreaves (Method)
HML	Hybrid Machine Learning
IoT	Internet of Things
IG	Integrated Gradients
I	Iodine
KNN	K-Nearest Neighbours
KNORA	K-Nearest Oracles
K	Kalium (Potassium)
LSTM	Long Short-Term Memory
LaRC	Langley Research Center
LR	Logistic Regression
LS-SVM	Least Squares Support Vector Machine
LRP	Layer-wise Relevance Propagation
LoRa	Long Range
LIME	Local Interpretable Model-Agnostic Explanations
LUR	Local Updating Rule
LGBM	Light Gradient Boosting Machine
MAE	Mean Absolute Error
MNN	Medium Neural Network
MQTT	Message Queuing Telemetry Transport
ML	Machine Learning
Mn	Manganese
Mg	Magnesium
MLP	Multi-Layer Perceptron

MLR	M ulti- L inear R egression
MSE	M ean S quare E rror
MODIS	M oderate R esolution I maging S pectroradiometer
MKGPCNN	M ulti-level K ronecker guided P elican CNN
Mo	M olybdenum
NLP	N atural L anguage P rocessing
NDVI	N ormalized D ifference V egetation I ndex
Ni	N ickel
N	N itrogen
Na	S odium
NB	N aive B ayes
NDRE	N ormalized D ifference R ed E dge
NASA	N ational A eronautics and S pace A dministration
OC	O rganic C arbon
PA	P recision A griculture
PCA	P rincipal C omponent A nalysis
PM Model	P enman-Monteith M odel
POWER	P rediction O f W orldwide E nergy R esources
PR	P attern R ecognition
PNN	P robabilistic N eural N etwork
P	P hosphorus
PSO	P article S warm O ptimization
PPSO	P hasor P article S warm O ptimization
pH	potential of H ydrogen
RF	R andom F orest
R²	C oefficient o f D etermination
PLSR	P artical L east S quares R egression
Si	S ilicon
RBFNN	R adial B asis F unction N eural N etwork
RFE	R ecursive F eature E limination
RF	R andom F orest
RMSE	R oot M ean S quare E rror
RNN	R ecurrent N eural N etworks
SVM	S upport V ector M achine
SHAP	S Hapley A dditive ex P lanations
SSA	S parrow S earch A lgorithm
S	S ulfur
SI	S warm I ntelligence
SMOTE	S ynthetic M inority O versampling T echnique
SVR	S upport V ector R egression
SCM	S tructural C ausal M odel
SOC	S oil O rganic C arbon

STCR	S oil T est C rop R esponse
SPEI	S tandardized P recipitation and E vapotranspiration I ndex
T-LSTM	T ransductive L ong S hort- T erm M emory
TNN	T rilayered N eural N etwork
TN	T otal N itrogen
TCN	T emporal C onvolutional N etwork
UAV	U nmanned A erial V ehicle
V2X	V ehicle-to- E verything
VRI	V ariable R ate I rrigation
VPD	V apour P ressure D eficit
V	V anadium
WNN	W ide N eural N etwork
WOA	W hale O ptimization A lgorithm
WSN	W ireless S ensor N etwork
XAI	e Xplainable A rtificial I ntelligence
XGBoost	e Xtreme G radient B oosting
Zn	Z inc

Chapter 1

Introduction

The worldwide population is projected to increase to nearly 9.1 billion people by the year 2050, an increase of approximately 34% compared with today's figures. This population growth will cause a $\approx 70\%$ rise in food demand [2]. Due to urbanization, the availability of agricultural land is reducing day by day. In the preset date India is struggling with insufficient food production. Also, India is expected to reach the most populated country by 2050 [3]. The key factors for the downfall of food production include poor planning, unpredictable weather, outdated harvesting and irrigation methods. Due to climate change, significant disruptions in weather patterns are experienced in recent years [4]. Global warming causes repeated droughts and massive rainfall that create substantial problems for small-scale farmers. According to annual economic survey done by the government of India [5], such unfavorable environmental conditions reduce farmers' incomes by $\approx 20\text{-}25\%$.

Precision Agriculture (PA) offers a viable approach for safeguarding worldwide food availability [6]. PA is an innovative approach for eco-friendly farm management. By integrating cutting-edge technologies, software applications, and intelligent devices, PA can make informed decisions for farming, as shown in Fig. 1.1. Modern PA represents a continuation of earlier agricultural transitions, rooted in the era of mechanized cultivation and the high-yield breakthroughs of the Green Revolution [7]. These two stages are widely recognized as the first and second revolutions of agriculture, while PA now stands at the core of the third revolution within the farming sector.

The primary focus of PA is to boost farm profitability by reducing production costs and environmental impact. John Deere in 1990 introduced a precision farming technology [8][9]. GPS-controlled tractors for seed planting and fertilizer spraying. Modern techniques like AI, IoT, cloud-edge-fog computing, and block-chain are the main components of precision farming [10]. In PA, IoT-driven intelligent sensors are installed in farmland

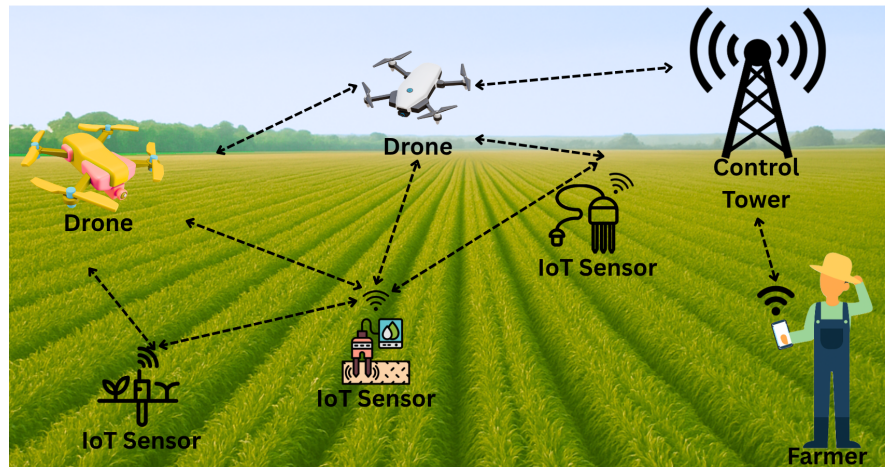


FIGURE 1.1: Precision Agriculture

for gathering data on soil characteristics, water, manures, and crop growth. Fully and partially self-operating devices use computer vision in UAVs [11] and robots to detect weeds and plant diseases. Satellite imagery is also used for field monitoring and plant health issues. Information gathered through sensors is operated and assessed through the application of ML algorithms to optimize regular farming practices [12]. These algorithms also help to predict weather and rainfall patterns. Additionally, ML algorithms can mitigate the risks pertaining to crop failure due to unpredictable weather and prevent tragic outcomes like farmer suicides due to crop failure. Technology-enabled herd management is another key aspect of PA, which helps to monitor animal health, performance, and breeding over the course of their lifespan well-being, as well as breeding across the entire life span. Sensors and cameras track livestock health, while computer vision helps to control the spread of animal-related diseases. Self-operating tractors and smart irrigation systems are the modern tools for efficient farming practices. The global adoption of PA is driven by the availability of machine learning techniques, fast internet connectivity, and advanced computing systems. Studies have explored the role of ML in enhancing sustainable agriculture supply chain (ASC) performance [13][14], frameworks for academics and professionals to analyze the influence of digital innovations on agriculture. It is expected that governments will introduce policies to encourage broad adoption of PA.

1.0.1 Artificial Intelligence and IoT in PA

The idea of AI was initially planned by John McCarthy during the Dartmouth Conference in the year 1956 [15][16]. John McCarthy defined AI is described as the branch of science and engineering that aims at producing intelligent machines, with focus on computer programs. AI-powered systems are capable to acquire knowledge, interpret, and

respond to the situations with computational intelligence. It has several sub fields, such as ML [17], DL [18], NLP [19], SI [20], Expert Systems [21], FL [22], Computer Vision [18] etc., as illustrated in Fig. 1.2. AI can be found in various applications in diverse fields of life, such as medical care, agriculture, financial services, robotics, e-trade, and automated industries. Now a days, giant tech companies such as Samsung and Apple are using AI in all their innovative devices [23].

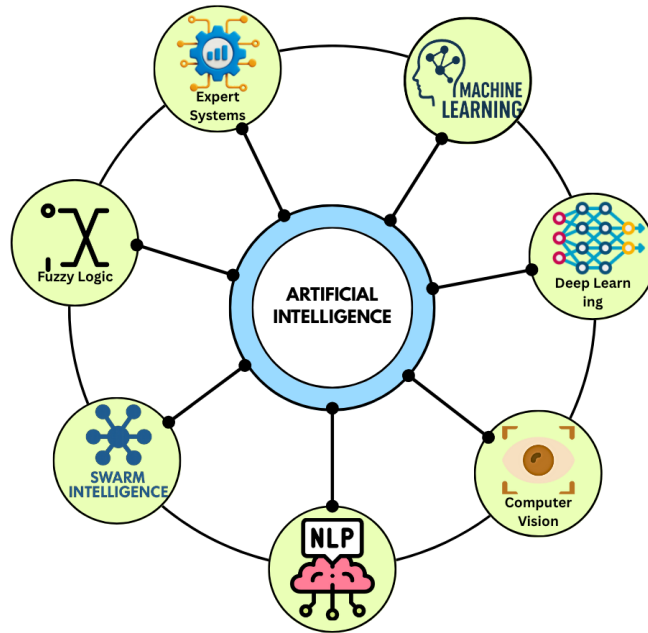


FIGURE 1.2: AI techniques for PA

The IoT [24] is a valuable technology enabling the interconnectedness of intelligent sensors and products via the cyber space. To enhance resource utilization, these sensors are utilized by several industries, such as solar power, agriculture, disaster management, and manufacturing. An increase in population leads to heightened scarcity of food. However, limited land availability and a declining interest in agriculture among younger generations have created challenges in meeting the food requirements for billions of people. To address this issue, the farming sector is progressively integrating smart tools such as IoT and AI to grow crops in smaller spaces and overcome traditional farming challenges.

IoT-enabled PA systems monitor soil conditions, including nutrients and moisture, via sensors. Machine learning algorithms may assist in identifying the most suitable fertilizer [25] amounts before planting the crops. Camera-enabled drones are revolutionizing agriculture by performing tasks such activities include observing crops and fields, spraying chemicals, and controlled drip irrigation. The photos taken by drones throughout

the crop life cycle can be used to identify diseases and weeds using DL algorithms. After proper identification, drones can be further used for spraying pesticides on affected crops.

Uncertainty of weather is a big concern for farmers. The increasing frequency of localized climatic anomalies, such as sudden dry spells during crucial vegetative growth phases or untimely heavy precipitation, exposes traditional rain-fed systems to catastrophic crop yield drops. This unpredictability fundamentally reduces agricultural economic predictability and severely risks farm investments, rendering standard seasonal crop planning obsolete. Smart drip irrigation technologies [26], trained on environment patterns are helping to alleviate the shortage of water by dynamically buffering crops against these short-term weather fluctuations. AI-enabled robots are also transforming agriculture by enhancing rapidity and effectiveness of crop harvesting and thus minimizing dependence on manual labor. NLP-based chat-bots keep farmers updated on the latest agricultural advancements and provide solutions to their related agricultural problems. This technology enables farmers to integrate the latest techniques into their practices, which improves field productivity.

TABLE 1.1: Various ML Algorithms

Algorithm	Description
LR [27]	A supervised ML method applied to forecast a continuous target variable by fitting a proportional association linking the input and output variables.
DT [27]	A tree-like model employed for both predictive and categorical modeling purposes, where data is divided according to feature values, forming decision points and terminal leaves.
RF [27]	An ensemble method that builds an ensemble of DTs and merges their results, providing improved accuracy and generalization by reducing over-fitting.
SVM [27]	A classification algorithm selects the optimal separating plane that divides samples belonging to different categories within a multi-dimensional feature space.
KNN [27]	A simple learning algorithm relying on instances, where classification is decided by the predominant category among the closest neighbors within the feature space
NB [27]	A probability-based classifier grounded in Bayes' theorem, frequently applied in agriculture to identify crop diseases and soil characteristics using likelihood measures of features.

Sustained on succeeding page

Table 1.1 – Sustained from preceding page

Algorithm	Description
K-Means Clustering [28]	An unsupervised method for splitting datasets into K clusters guided by attribute similarity, commonly used for segmentation tasks like soil types or plant varieties, etc.
ANN [28]	A family of models inspired by the human brain, used for complex tasks like crop yield forecasting, disease identification, and environmental monitoring, among others.
DL [28]	A subset of ANN that includes CNN for image recognition and RNN for time-series analysis, helpful in plant disease detection and yield prediction.
GBM [27]	An ensemble technique that builds DT in a sequential manner to minimize error and improve predictive performance. It is often used in agriculture for prediction tasks like weather forecasting and pest detection.
XgBoost [27]	A highly optimized and efficient gradient boosting algorithm, widely used for predictive tasks like crop yield forecasting and optimizing resource management in agriculture, and others.
PCA [28]	A dimensionality reduction approach applied to reshape high-dimensional information into a compact form, often used to analyse soil quality and plant growth features.
LR [27]	A classification algorithm used for handling binary categorization tasks, such as predicting whether a crop will survive under specific weather conditions.
BO [28]	A probabilistic optimization technique used for hyper-parameter tuning, often applied to optimize irrigation systems and fertilization schedules in PA.

1.0.2 ML Algorithms

ML represents a subfield of computer science that allows systems to derive insights from datasets and enhance their performance without relying on explicitly coded instructions. The idea of ML was initially proposed by Arthur Samuel in 1959 [29]. In 1950, Alan Turing presented the notion of machines capable of learning in one of his papers "Computing Machinery and Intelligence" [30], where he acquaints with the Turing Test an approach used to assess whether a machine is able to exhibit behavior indistinguishable from human intelligence. ML models learn from data by identifying patterns within data and developing models capable of generating outcomes in the form of predictions or decisions. The ML workflow generally is composed of three primary phases: data acquisition, model construction, and generalization. (Fig. 1.3). Generalization, which describes how well a trained system can extend its predictive capability to previously unseen datasets that were not part of the learning stage.

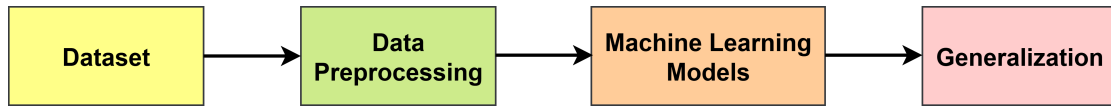


FIGURE 1.3: Workflow of a ML model

ML models have proven their ability in solving difficult tasks where human experience may fail, such as weather forecasting, spam detection, plant illness diagnosis, and feature identification, etc. [31]. The easy availability of large datasets and ML algorithms, combined with internet-connected technologies, has increased the use of ML in both industry and academia. DL, as an advanced branch within ML, is particularly effective in handling massive volumes of data to derive insights and support complex agricultural decision-making. DL models have demonstrated remarkable performance in domains such as computer vision (e.g., object and image analysis), NLP, and autonomous systems.

ML algorithms are typically categorized into three fundamental paradigms: guided learning, unsupervised learning and reinforcement learning (Fig. 1.4). Supervised learning trained using labeled datasets, where the input-output mapping is explicitly defined. The system attempts to identify the fundamental linkage connecting inputs and outputs, and generalizes the learned information to forecast results for unfamiliar or previously unseen data. Supervised learning supports both classification tasks and regression analysis to predict categorical and numerical values. Unsupervised learning, in contrast, operates on unlabeled data and focuses on discovering hidden patterns or grouping similar kinds of data. Unsupervised learning is more challenging to implement than supervised learning. Reinforcement learning involves the training of systems through interaction with the environment by getting continuous feedback in the form of incentives and sanctions. A notable example of reinforcement learning is AlphaGo by DeepMind [32], which used the approach to defeat top human players and the best computer-based chess engines. Different ML techniques applied within PA, which are presented in Table 1.1.

1.1 Applications of ML in PA

In numerous nations, farmers still depend on conventional farming practices which is often guided by the advice and experience of older generations. However, this approach leaves them vulnerable to unpredictable climate conditions, which are becoming more frequent due to global temperature increase and irregular precipitation patterns. Manual application of xenobiotics often leads to resource misuse and environmental harm. AI and IoT-powered PA eliminate this unpredictability, enabling modern farmers to enhance all stage of the cultivation process. Fig. 1.5(a) and 1.5(b) visually illustrate the contrast between traditional farming and technology-driven farm management systems.

Carlos F [33] presented a structured study on the impact of intense weather events like hailstorms, cold surges, and heat surges on agriculture. This study identified that inundations, prolonged aridity, cold snaps, hail, hot spells, and pest infestations are all more likely to be influenced by

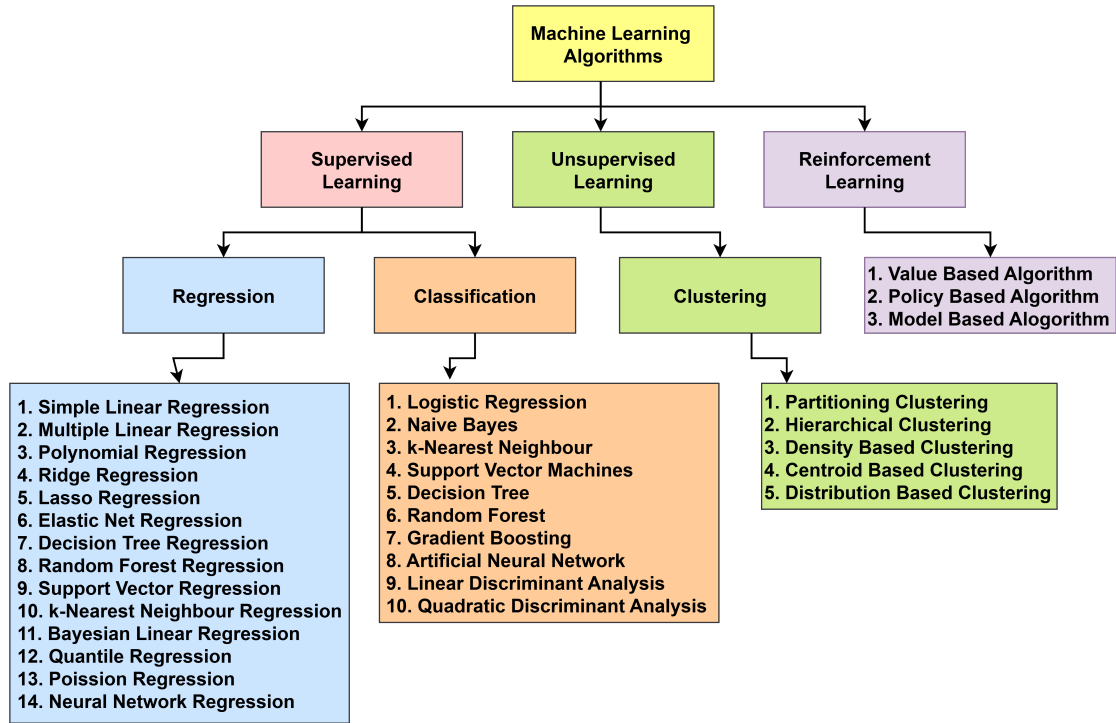


FIGURE 1.4: Various kinds of ML algorithms

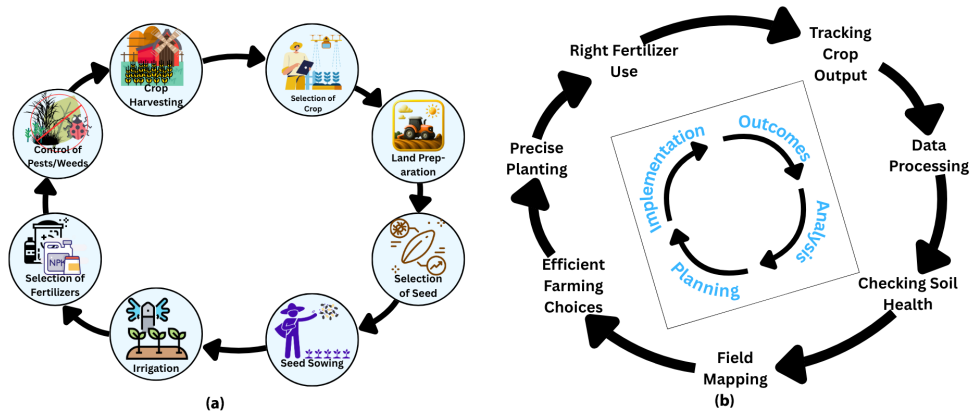


FIGURE 1.5: (a) Conventional agriculture (b) PA

climatic changes. AI is now integrated into various farming operations, and some even extend beyond traditional practices.

1.1.1 Forecasting Soil Characteristics and Climate Conditions

Forecasting soil characteristics is a foundation step within the agricultural sector that affects crop determination, field preparation, seed choice, fertilizer application, and ultimately, crop yield. Soil properties are closely dependent on regional and climatic characteristics of the land. Prediction typically involves assessing essential soil minerals, topsoil moisture, and environmental climate factors throughout the entire crop growth cycle [34]. Human interventions have greatly modified soil characteristics, which affect the ability to cultivate crops effectively [35]. There

are 20 required nutrients are listed in Table 1.2 that play vital role in vegetation growth [36]. The presence of these nutrients in the soil has a direct effect on crop development. Essential soil minerals are often observed using electrical and radio-frequency sensors [37], which guide farmers in choosing suitable crop varieties for their farmland. Although fertilizers and manure can add nutrients but they come at an added cost and may also have negative environmental impacts.

TABLE 1.2: Required nutrients for plants

Nutrient	Significance
N	Crucial for foliage development and photosynthetic activity; supports vigorous green plant growth.
P	Vital for root formation, energy movement, and bloom production.
K	Controls moisture regulation, enzyme stimulation, and resilience to stress.
Ca	Reinforces cellular walls and aids in root formation and cell multiplication.
Mg	Primary constituent of chlorophyll; crucial for photosynthetic processes and enzyme initiation.
S	Necessary for amino acid and protein formation, boosts resistance to diseases.
Fe	Essential for the formation of chlorophyll and electron transport during photosynthesis.
Mn	Participates in enzyme activation and contributes to photosynthetic activity.
Zn	Aids enzyme activity, protein formation, and regulation of plant growth.
Cu	Required for photosynthetic functions and enzyme stimulation.
B	Crucial for maintaining cell wall integrity and reproductive growth.
Mo	Needed for nitrogen fixation and the conversion of nitrate within plants.
Cl	Plays a role in osmotic regulation and ionic balance, assisting in photosynthesis.
Ni	Significant for activating enzymes in nitrogen-related pathways and contributing to seed sprouting.
Na	Improves tolerance to drought and aids photosynthetic functions in certain plant species.
Si	Reinforces cell walls, increases resistance to insects and disease-causing agents, and boosts drought endurance.
Co	Necessary for nitrogen fixation in leguminous plants and supports metabolic enzyme systems.
I	Contributes to plant stress tolerance and the formation of antioxidants.
V	Participates in nitrogen-fixing processes in specific plants, particularly legumes.
Cr	Influences plant resilience to stress, especially when exposed to heavy metal pollution.

Analysis of nutrient profile, hydration status, and pH is essential for identifying soil characteristics. As an example, Mehmet et al. [38] applied an ELM-based regression approach to estimate soil surface moisture. Their study, conducted on two terrains at Dicle University, involved real-time ground data collected using Radarsat-2, which was initially processed using the SNAP tool

and supported by on-site measurements. The ELM model achieved the lowest RMSE greater than two using "sine" kernel method.

Xiaogang et al. [39] used ELM-driven inferential sensors to estimate nutrient solution composition in soil-less cultivation, which requires continuous monitoring of parameters like pH, temperature, and nutrient concentrations. They used conductance, acidity–alkalinity level, flow rate, and heat condition data to predict the concentration of sulfate and dihydrogen phosphate ion and achieved an RMSE of 1.2414 and 0.8892 for sulfate and dihydrogen phosphate ion, respectively.

Seonyoung et al. [40] used ML algorithms to forecast soil humidity from MODIS data [41]. The models, including RF and cubist algorithms, showed a high correlation ($R^2 = 0.96$) and low RMSE (0.06), outperforming traditional statistical methods, which produced an R^2 value of 0.47 along with an RMSE of 0.16.

Rabie et al. [42] explored ML methods to forecast SOC and TN using near-infrared spectroscopy data from four agricultural lands in Morocco. The ELM outperformed other traditional regression models, showing an R^2 of 0.96 for SOC and 0.94 for TN.

Antonios et al. [63] used infrared spectroscopy to forecast SOC, TN and humidity content in Premslin, Germany. The LS-SVM got the best result for SOC and moisture content, while Cubist was optimal for TN prediction.

Emrullah et al. [38] and Xiaogang et al. [39] presented the role of advanced ML techniques like ELM and soft sensors to predict soil properties. The results revealed that the used technologies offered a cost-effective and accurate alternative compared with traditional methods. Table 1.3 provides a relative analysis of various ML algorithms for predicting soil profile and climate conditions.

1.1.2 Crop Yield Forecasting

Forecasting crop yield along with crop improvement is crucial for farmers. The influencing key factors in crop yield are pH level, soil classification, climatic variations (including temperature, precipitation, moisture, and daylight duration), nutrient inputs, and crop-harvest timings [64]. Traditional farming methods can be used a feedback system, in which adjustments are made only once crop development is monitored. The efficiency of using resources directly impacts the crop yield. However, if any issues go unnoticed in the early stages, they can severely affect yield.

Saxena et al. [65] reported the influence of hailstorm events on wheat output in India. Every year, February-March hailstorms led to an 8.4 percentage decline in the country's wheat production. In countries like India, where post-harvest storage capacity is restricted, weather forecasts could significantly benefit farmers. Unlike conventional methods, ML models operate as a feed-forward control system, proactively predicting factors that may affect yield before they cause any damage. This allows farmers to take preventive actions.

Elisa et al. [66] applied several ML approaches to detect wheat yield-gap zones and produced high-resolution yield maps by integrating multi-source data from 2009–2015. The dataset comprised NDVI time-series data, temperature and rainfall records along with machine-recorded grain yield observations. The authors tested nine distinct ML techniques, such as RF, XGBoost, Cubist, MLP, and SVR. Among them, SVR with RBFNN achieved the highest performance, yielding an R^2 of 0.77 and an RMSE of 0.55 using 10-fold cross-validation.

Kentaro et al. [68] used SVR to predict corn yield in Illinois, utilizing a five-year rolling average of crop yield records, MODIS satellite data for the enhanced vegetation index, and historic climate data. The model achieved an RMSE of 8.207 and a R^2 of 0.645 after conducting 10-fold cross-validation.

Hossein et al. [67] reported sophisticated regression methods for forecasting silage maize yield in Iran, analyzing a dataset from the Moghan Agro-Industrial and Livestock Company. Results showed the boosted regression tree delivered superior results, showing an R^2 above 0.87 and an RMSE ranging from 8.5 to 11.1.

Shruti et al. [81] utilized DL models with feeding data of soil characteristics and nutrient data gathered over a period of 31 years, along with historical precipitation records. RNN was used with several activation functions and linear for the enhancement of prediction precision.

Zheng et al. [82] developed a rice yield forecasting system for China. The system used RNN for both warm-season and cold-season rice yields prediction. The RNN framework attained minimal error levels, producing MAE and RMSE values of 0.0045 and 0.0058 for warm-season rice, and 0.0075 and 0.0193 for cold-season rice.

Puyu et al. [83] used a hybrid method with the incorporation of agricultural simulation tools (APSIM), NDVI, and the SPEI for forecasting yield of wheat in New South Wales, Australia. The APSIM-RF hybrid model demonstrated better predictive performance relative to other approaches.

Using climate and satellite information spanning 14 years in Australia, Yaping et al. [84] employed a range of ML technologies (including SVM, RF, and NN) to predict wheat yield. They reported that the climate data was more helpful than satellite data in their yield forecasting, obtaining an R^2 close to 0.75. Precise scheduling of crop planting are crucial for improving crop productivity. Abdulkadir (Gumuscu) et al. [85] applied supervised ML approaches such as kNN, SVM, and DT for the prediction of early, regular, and late planting dates for wheat crops in Turkey. Among all the models, kNN provided the most accurate results.

In other regions, such as in coffee production in Vietnam, Louis et al. [86] used the ELM model for forecasting coffee yield based on soil fertility information. The ELM approach exceeded the performance of LR and RF models, yielding an RMSE of 497.10 and an MAE of 327.10.

TABLE 1.3: Overview of ML Applications in Soil and Climate Parameter Prediction

Reference	Parameter	Dataset Used	Methods Used	Results
Vahid et al. [43]	Soil temperature	Annual dataset includes eight hourly climatic features	CNN, MLP	CNN model performed better than MLP model
Priyanka et al. [44]	Soil moisture evaporation rates	Twenty-one ground-based parameters	MLR, KNN, MLP, SMOreg, RF, KNORA ensemble	KNORA with 7.54 RMSE and MLP with 25.1 RMSE
Jinyang et al. [45]	Soil hydration	Multi-layer soil hydration data from Montana Mesonet	Satellite-driven machine-learning approach	Strong performance in soil moisture predictions ($R^2 \geq 0.91$)
Yulan et al. [46]	Soil loss rates under rainfall events	Field observation data from rainfall events	RF, SVM, ANN	RF outperformed SVM and ANN
Lei et al. [47]	Soil moisture	ERA5 data [48], C3S land cover data [49] for Europe areas	SVR, RF, ConvLSTM	Hybrid model shows substantial improvements over conventional methods
Lu et al. [50]	Soil moisture	SMAP Level-4 soil moisture product	AttConvLSTM	Got $R^2=0.82$, RMSE=0.02
Xiaojie et al. [51]	Soil temperature and moisture	Real data from Heihe River Basin, China	LSTM, ARIMA, SVR	ARIMA and SVR achieved low RMSE and MAE
Attaur et al. [52]	Daily rainfall	12 years of historical weather data from Lahore	DT,NB, KNN, SVM, Fuzzy logic for ML fusion technique	Fusion-based framework outperforms other models
Renze et al. [53]	Weather	Not included	hybrid data assimilation method (HDA-ML), Bilinear neural network, CNN	HDA-ML improves assimilation performance and reduces time cost

Continued on next page

Table 1.3 – continued from previous page

Reference	Parameter	Dataset Used	Methods Used	Results
Yu et al. [54]	Daily air temperature	1097 data from Tabriz city, Iran, 2017 to 2019	LSTM and DLSTM	DLSTM algorithm has RMSE = 0.08 and R-Square = 0.99
Venkatachalam et al. [55]	Weather	HHWD and Jena Climate dataset [56]	LSTM and T-LSTM model	T-LSTM method achieved 98.2% accuracy
Fu et al. [57]	Short-term rainfall	Cloud images and meteorological data	CNN with ResNet-152 and RNN with LSTM	Combination of CNN and RNN achieved 82% accuracy
Li et al. [58]	Soil properties prediction	Different 97 zones of Singapore island	RF, ANN	RF achieved lowest regression error
Rivera et al. [59]	Soil aggregate stability	109 soil samples from Chile	ANN, GLM	ANN achieved better result than GLM with $R^2=0.82$
Padarian et al. [60]	Soil organic carbon prediction	LUCAS soil database [61]	PLS, Cubist regression tree, CNN	CNN reduced the error by 87% relative to Cubist
Khanal et al. [62]	Spatial prediction of soil properties	Multi-spectral aerial images and topographic data of Molly Caren, London, Ohio	RF, NN, SVM, GBM, CU	RF outperformed other models ($R^2=0.53$, RMSE=0.97)

Andrea et al. [87] used SVM and Generalized Linear Model (GLM) to predict cocoa yield in Santander, Colombia. This is just one example of how researchers are increasingly relying on statistical and probabilistic models for the yield prediction. Table 1.4 shows a relative analysis of various ML methods for crop yield forecasting.

1.1.3 Drip Irrigation

In modern farming, irrigation structures have evolved over time, with drip irrigation emerging as a key advancement [88]. Thin plastic tubes are positioned either in or above the soil along plant rows for direct water supply in this irrigation system.

By managing drip irrigation effectively, farmers can optimize water usage while improving crop yields. Drip irrigation has become particularly valuable for high-value crops like vegetables and fruits, and it is appreciated for its socio-economic and environmental benefits. Compared to sprinkler systems, drip irrigation operates at a lower pressure, making it more energy-efficient [89].

Drip irrigation offers several advantages over other irrigation systems, including sub-irrigation and sprinklers. These features reduced water consumption, enabled to direct apply of soluble fertilizers, decreased soil erosion, suppressed weed growth, and supported double cropping. Precision irrigation, a pioneering approach in PA (intelligent farming), optimizes water use, helping farmers to achieve a higher yield with minimal water. This approach ensures precise water application by applying the precise quantity at the optimal time and place often through Variable Rate Irrigation (VRI) systems using either drip or sprinkler methods. [90].

The integration of field-based sensor technologies, climate forecasting, IoT-enabled vegetation sensors, and advanced precision irrigation systems generates huge data. These data assists farmers to optimize water use, enhance crop yields, and maximize overall profits. [91].

Michael et al. [92] highlighted the challenges of sensor-dependent control systems, noting that reliance on process-based crop models can diminish decision-support tools reliability and may require costly calibration, resulting in uncertain representations of soil-plant-atmosphere processes. ML has been effectively used for hydrological protection analysis, such as estimating soil moisture content and groundwater depth. Jiusheng et al. [93] applied an ANN-based approach to predict nitrate distribution in soils within a drip irrigation setup.

Nivas et al. [94] reported an automatic drip irrigation system with an ARM9 processor and various sensors for soil pH and nitrogen content monitor to ensu adequate irrigation. Abioye et al. [95] reported a data-centric framework for drip irrigation in a Malaysian greenhouse, using IoT devices to monitor soil moisture, irrigation volume, and evapotranspiration. These data were analyzed using multiple predictive models, including Autoregressive with eXogenous Inputs (ARX), BJ, and state-space approaches and the ARX framework delivered the most favorable MSE and response time.

TABLE 1.4: Overview of ML Applications in crop yield prediction

Reference	Objective	Dataset Used	Methods Used	Results
Bian et al. [69]	Wheat yield prediction	UAV multispectral images of winter wheat fields	GPR, SVR	GPR model achieved highest accuracy
Gupta et al. [70]	Crop yield prediction	Not included	Relief algorithm for feature selection, PSO-SVM, KNN and RF	PSO-SVM achieved 94% accuracy
Kumar et al. [71]	Corn yield prediction	UAV-based multispectral observations gathered during the V5 and R6 growth stages	UAV-based Vegetation Indices, LR, KNN, RF, SVR, DNN	SVR and KNN outperform other ML models
Sajid et al. [72]	Corn yield prediction	Daily weather data from Iowa Environmental Mesonet (IEM), Crop management data from NASS	Coupling crop modeling to developed five new ML models along with their ensemble versions	Ensemble model achieved 9% RRMSE
Sadenova et al. [73]	Crop yield prediction	Processed images from Sentinel-2 and Landsat-8 satellites	Various ML algorithms	MLP and PR achieved the accuracy of 99%
Pavani et al [74]	Crop yield prediction	Datasets collected for jowar crops	SVM-RF model	Achieves 90% accuracy
Vadav et al.[75]	Crop yield prediction	USDA datasets [75] from 2003 to 2013	EDA, RF, DT, KNN, Bagging Regressor	RF, DT, Bagging Regressor models showed high accuracy
Bouni et al. [76]	Crop yield prediction	Extensive IoT secondary dataset from online repository	LightGBM, DT, RF	92% accuracy achieved with LightGBM

Continued on next page

Table 1.4 – continued from previous page

Reference	Objective	Dataset Used	Methods Used	Results
Shuaibu et al. [77]	Crop yield prediction in Nigeria	Data from the Agricultural Performance Survey of Nigeria	Linear Regression, SVR, K-Nearest neighbor, DT	DT achieved 72% prediction accuracy
Jhajharia et al. [78]	Predict wheat crop yield in Rajasthan, India	Integrated satellite and climate data for crop yield	Lasso regression, SVM, RF, linear regression	SVM outperformed other prediction approaches
Abbas et al. [79]	Potato tuber yield prediction	Soil and Crop properties collected using proximal sensing method	LR, EN, k-NN, SVR	SVR method outperformed all other models with RMSE from 4.67 to 6.68
Kuradusenge et al. [80]	Maize yield prediction	environment and yield data for Musanze, Rwanda	RF, PR, SVR	RF was given the best model with R^2 0.875 and 0.817

Using four different ML algorithms, Amin et al. [96] refined the emitter flow amount of dribble irrigation systems under different temperature scenarios. Their findings indicated that LSSVM delivered the finest performance with the minimal error. Yaoqi et al. [97] built an optimized model in China using back-propagation neural networks, integrating soil and weather parameter for irrigation forecasting. Their model recorded an MSE value of 0.00857724.

Jalal et al. [98] discovered soft computing methods like GEP and RF to model moistening front magnitudes for superficial and sub- superficial irrigation with enhanced correlation metrics and decreased MSE and MAE values. M. N et al. [99] used ANN to predict wetting front dimensions, achieving a correlation range of 0.93 to 0.99.

Yu-Chuan et al. [100] established a smart irrigation system with ML and the LoRa network to gather irrigation-related knowledge from experienced farmers. Gursimran et al. [101] reported an ML-IoT combined framework for forecasting soil wetness on irrigation, while Roque et al. [102] introduced a decision-depend framework for irrigating citrus crops in Spain. With the use of smart sensors, this system monitored water supply, weather, and soil moisture and RF model provided the best prediction accuracy in this case. Other advancements include the PLSR and ANFIS for irrigation management [111], and real-time intelligent irrigation management systems using wireless sensors and actuators in greenhouses [112]. Zhijun et al. [113] used a temporal convolution network (TCN) to estimate crop evapotranspiration, achieving better correlation and reduced error. Table 1.5 shown a relative evaluation of many ML models applied in drip irrigation.

1.2 Application of IoT in PA

The IoT has a crucial role in daily agricultural practices [114]. Using various sensors, weather and soil parameters can be collected and sent to an IoT cloud platform for remote supervision, which reduces the need for farmers' direct involvement. The IoT cloud can also be used for detecting and preventing animal intrusion in agricultural fields [115]. Sensors are essential in PA, as they help effective monitoring and control. Beyond monitoring and control, IoT also serves as a data storage solution. Key parameters such as soil characteristics, crop productivity, seasonal changes, and temperature levels can be efficiently saved and administered within the IoT cloud. The stored data can again be used for analysis, prediction, and making informed agricultural decisions.

TABLE 1.5: Overview of ML Applications for drip irrigation

Reference	Objective	Dataset Used	Methods Used	Results
Jamei et al. [103]	Predict wetting patterns in layered soil profiles	1217 sets of experiments conducted on two-layered soil profiles	Boruta-XGB, Bi-GRU, MLP, AdaBoost	Bi-GRU model outperformed MLP and AdaBoost in accuracy
Trabelsi et al. [104]	Groundwater irrigation suitability estimation	Seventy-two groundwater samples collected from Medjerda River Basin, Tunisia	RF, SVR, ANN, AdaBoost	AdaBoost model predicts irrigation quality parameters most accurately
Mokhtar et al. [105]	Forecasting irrigation water requirement	Climate data from Cairo University platform	SVR, RF, DNN, CNN, LSTM, Hybrid CNN-LSTM	Hybrid LSTM & CNN model showed highest accuracy
Togneri et al. [106]	Irrigation water requirement estimation	Soil moisture data from eight crop types across twelve fields in Brazil	LightGBM, LR, DT, RF, MLP, LSTM, StemGNN	LightGBM outperforming others
Wei et al. [107]	Estimate irrigation water amounts	NDVI, NDWI, Greenness Index, ET products	RF	RF model estimates irrigation water with 82% accuracy
Kurtulmucs et al. [108]	Proximal sensing system for smart irrigation	$\approx 12,214$ images of soil texture from Bursa Uludag University	AlexNet, GoogleNet, ResNet, VGG16, SqueezeNet	AlexNet got 0.9978
He et al. [109]	ET _o estimation	EV_{T0} estimation in Hetao Irrigation District, China	Optimized ELM with GWO, MFO, PSO, WOA	GWO-ELM model achieved highest accuracy
Tang et al. [110]	Water status monitoring from vegetation indices	Proximal sensor observations and UAV imagery data	RF	Proximal sensor model: R^2 of 0.74, MAE of 0.63, UAV imagery model: R^2 of 0.81, MAE of 0.67

1.2.1 Use of IoT Sensors in PA

The IoT is the interconnection of various devices, like sensors. These sensors, working in groups, communicate among themselves and share information with a central control. IoT has extended the scope of agriculture, animal rearing, and smart irrigation. Sensors are the fundamental component of IoT architecture, which are also act as transducers, converting sensed parameters into electrical signals. In IoT, WSN [116] play a vital role in improving agricultural practices by enhancing production efficiency, increasing crop yield quality, detecting and preventing plant-eating pests, and identifying fire risks on farms. Sensors can be classified into two categories: analog and digital. Analog sensors need conversion of their output to digital form before being processed by an IoT system, while digital sensors can be directly used by the IoT system. Table 1.6 lists important sensors used in PA. For the purposes of this thesis, the present list includes those commonly used sensors in agriculture. Sensors and actuators can collect data on crucial agricultural parameters [117] and trigger predefined actions when needed. IoT ensures atmost immediate action occur quickly with minimal delay. The speed of decision-making can be affected by several features such as sensor tolerance, communication protocol, etc. The optimized operating temperature of the sensors influences the tolerance. An increase in temperature can cause deviations in the measured values. The protocol for communication is used to send sensor data to the micro-controller also impacts data transfer speed, with higher data rates resulting in faster communication.

TABLE 1.6: Various IoT sensors used in PA

Sensor Type	Sensor Name	Specification(Measurement Range, Accuracy, Output, Power Consumption)
Soil Moisture Sensor	Decagon Devices EC-5 [118]	0 to 100% (Relative), $\pm 2\%$, Analog/Digital, Low (typically 20-30mA)
Temperature Sensor	DHT22 [119]	-41°C to 81°C , $\pm 0.6^{\circ}\text{C}$, Digital, Low (typically 1-5mA)
Humidity Sensor	DHT22 [119]	0% to 100%, $\pm 2\%$, Digital, Low (typically 2-5mA)
Light Intensity Sensor	BH1750 [120]	1 to 65535 lux, $\pm 5\%$, Digital, Low (typically 3-6mA)
CO_2 Sensor	MH-Z19 [121]	0 to 5000 ppm, ± 30 ppm, Analog/Digital, Low (typically 15-25mA)
pH Sensor	DFRobot SEN0161 [122]	0 to 14 pH, ± 0.1 pH, Analog, Low (typically 1-2mA)
Soil Temperature Sensor	DS18B20 [123]	-56°C to 126°C , $\pm 0.6^{\circ}\text{C}$, Digital, Low (typically 1-2mA)
Nutrient Sensor	Ion Selective Electrode (ISE) [124]	Customizable depending on nutrients, $\pm 2-5\%$, Analog/Digital, Low (typically 15-30mA)

Continued on next page

Table 1.6 – continued from previous page

Sensor Type	Sensor Name	Specification(Measurement Range, Accuracy, Output, Power Consumption)
Rainfall Sensor	RG-11 Rain Gauge [125]	0 to 5000 mm, $\pm 1\%$, Analog/Digital, Low (typically 5-10mA)
Wind Speed Sensor	Anemometer (Vantage Pro2) [126]	0 to 100 m/s, ± 0.5 m/s, Analog/Digital, Low (typically 5-15mA)
Leaf Wetness Sensor	LI-COR LI-250A Light Meter [127]	0 to 100%, $\pm 3\%$, Analog/Digital, Low (typically 1-5mA)
GPS Sensor	Ublox NEO-6M [128]	2D/3D, ± 2 -10 meters, Digital (GPS NMEA output), Low (typically 20-40mA)
EC Sensor (Electrical Conductivity)	SoilEC Sensor (DFRobot)	0 to 10 dS/m, ± 0.1 dS/m, Analog/Digital, Low (typically 2-4mA)
Acoustic Sensor	Microphone Array	10 to 2000 Hz, $\pm 1\%$, Digital (frequency count), Low (typically 10-20mA)
UV Radiation Sensor	Vernier UV-B Sensor	0 to 1200 W/m ² , $\pm 5\%$, Analog/Digital, Low (typically 5-10mA)
Ammonia Sensor	MiCS-5524 [129]	0 to 100 ppm, ± 2 ppm, Analog/Digital, Low (typically 10-20mA)
Nitrogen Dioxide Sensor	MiCS-4524 [129]	0 to 10 ppm, ± 0.1 ppm, Analog/Digital, Low (typically 15-25mA)
Soil Compaction Sensor	Soil Compaction Tester	0 to 2000 psi, $\pm 2\%$, Analog, Low (typically 5-10mA)
Temperature and Humidity Sensor	SHT31 [130]	Temperature: -41°C to 126°C , Humidity: 1% to 101%, Accuracy: $\pm 0.4^{\circ}\text{C}$, $\pm 3\%$, Output: Digital (I2C)
Barometric Pressure Sensor	BMP180 [131]	300 to 1100 hPa, ± 1 hPa, Digital (I2C), Low (typically 5-10mA)

1.3 Motivation of the Work

In the present date, the expanding world population, projected to hit 9.1 billion by 2050, coupled with the rising need for food imposes substantial strain on agricultural systems worldwide. In particular, countries like India face challenges with insufficient food production, exacerbated by the decreasing availability of arable land, poor planning, and erratic climatic patterns. Climate change adds further complexity to the situation, causing irregular weather events that affect crop yields and farmers' incomes. This situation is especially challenging for smallholder farmers. Climate-related disruptions in farming practices, including unpredictable weather patterns, have led to a considerable loss in agricultural income, estimated at 20-25% by the Indian government's annual economic survey.

In this context, PA emerges as a potential solution for sustainable farming. PA leverages data-driven, technology-based approaches to enhance agricultural practices, lower expenses, and boost farm profitability. Through the incorporation of contemporary technologies like the IoT, AI, ML, and data analytics, PA offers the potential to tackle the difficulties encountered by farmers in today's agricultural environment. These technologies support live tracking of soil quality, climatic conditions, crop development, and livestock management, facilitating more informed decisions and reducing the risks associated with traditional farming methods.

This research is motivated by the need to develop innovative, data-driven solutions that support farmers in adapting to a rapidly changing agricultural environment. The application of AI and IoT in PA offers a chance to increase agricultural productivity, improve resource efficiency, and reduce environmental impact. Moreover, by developing XAI frameworks, this research aims to build trust and confidence in these technologies among farmers, ensuring the widespread adoption of PA practices.

By focusing on predictive models for rainfall, smart irrigation systems, soil fertility classification, and crop-fertilizer recommendations, the main aim of this thesis is to support the current agricultural transformation by utilizing digital technologies to advance food security, environmental sustainability, and the economic welfare of farmers.

1.4 Aims and Objectives

1.4.1 Main Aim

The primary purpose of this study is to formulate, construct, and assess an integrated, interpretable framework for PA that assists farmers in taking well-grounded decisions throughout the full crop-production cycle.

1.4.2 Specific Objectives

The main goal will be realized through the following detailed objectives:

1. To develop a rainfall prediction model using a hybrid algorithm to support initial agricultural planning. This objective is addressed in Chapter 2, where an ACO-ANN model is developed and validated for daily rainfall forecasting. The model's performance is evaluated against other regression models, showing its better accuracy in predicting rainfall.
2. To design an intelligent IoT-based irrigation system which optimizes the usage of water with AI predictions. Based on the foundation of environmental prediction, Chapter 3 shows the advancement of a intelligent irrigation system. The system incorporates real-time data from IoT sensors with a hybrid model GA-ANN to predict reference evapotranspiration (EV_{T0}) and determine the optimal water for crops. The objective is to illustrate a decrease in water usage while preserving ideal crop health.

3. To enhance soil fertility and support improved crop selection. Chapter 4 focuses on the essential component of soil health. The hybrid ML model, ACO-XGBoost, is formulated to categorize soil fertility using macro and micro-nutrient characteristics. This objective offer farmers with a powerful tool for understanding the composition of the soil and informed decisions for soil classification and crop selection.
4. To build a sustainable recommendation system for crops as well as fertilizers that considers both crop yield and environmental impact. Chapter 5 demonstrates the predictive capabilities by developing a system for recommending suitable crops and fertilizers. This system employs ML models (CatBoost and XGBoost) fine-tuned using metaheuristic techniques to determine the most appropriate crops and the ideal fertilizer quantities for given soil conditions. The objective is not only to maximize yield but also to promote sustainable farming by minimizing fertilizer wastage and the ecological impact.
5. To establish an XAI framework is established to ensure the transparent, interpretable, and trustworthy predictions for the end-user (farmer). Recognizing the fact that adoption of AI in agriculture depends heavily on trust, Chapter 6 is dedicated to developing an XAI framework. This framework uses techniques like SHAP and LIME to deliver clear and understandable elucidations for the predictions with various models within the framework. The aim is to explain the application of AI clearly and enable farmers to make well-informed decisions.

1.5 Research Queries

This thesis aims to address the following queries:

1. How can accuracy of daily rainfall prediction be improved?
2. To what extent can an AI and IoT-driven smart irrigation system reduce water consumption while maintaining crop health?
3. How effectively can soil fertility be classified based on macro and micro-nutrient data?
4. What are the key factors for a recommendation system that successfully balances crop yield with sustainable fertilizer use?
5. How can XAI be integrated into agricultural models to enhance farmer trust and improve decision-making quality?

1.6 Extent of the Research

1.6.1 Geographical and Time-Based Extent

The studies are mainly concentrated on particular areas within West Bengal, India. The places are chosen for their agricultural significance and availability of data. The rainfall prediction

model developed in Chapter 2 uses a long-term dataset from 1981 to 2023 for Alipurduar, West Bengal. The intelligent irrigation system in Chapter 3 is physically tested in a private garden in Haldia, with real-time data from August to October, 2023. The crop and fertilizer advisory system discussed in Chapter 5 is built upon soil health card data from seven distinct districts of West Bengal. The crop yield prediction framework in Chapter 6 expands its scope by using both global dataset from the World Data Bank (1990-2013) and a more localized dataset of West Bengal from 1958 to 2022.

1.6.2 Crop-Specific Focus

To ensure the depth and practical relevance, certain chapters of the thesis focus on specific crops. The intelligent irrigation system developed in Chapter 3 is validated to meet the precise water requirements of the coriander crop across its vegetative, bolting, and flowering stages. The recommendation system in Chapter 5 contains a portfolio of crops that are both economically and culturally significant for West Bengal, such as tomato, cumin, rice, sunflower, cabbage, wheat, soybeans, etc.

1.6.3 Model Development and Evaluation

The main goal of this work is to create and evaluate various hybrid ML models created for specific agricultural tasks. Hybrid ML models like the ACO-ANN model for rainfall forecasting, the GA-ANN model for smart irrigation, the ACO-XgBoost model for soil fertility classification, and the CatBoost-SSA model are used here for the recommendation of crop and fertilizer. The effectiveness of these models is evaluated against an extensive set of traditional ML algorithms (including RF, SVM, DT, etc.).

1.6.4 XAI Framework

The XAI framework developed in Chapter 6 offers a comprehensive toolkit for model transparency. It specifically utilizes SHAP for generating global explanations of model behavior and identifying key feature importance towards the model output. LIME providing intuitive and instance-specific local explanations. Additionally, Counterfactual Explanations is to explore "what-if" scenarios.

1.6.5 Data and Technology Stack

The research uses data from a diverse set of credible and dependable sources, such as the NASA POWER Project for meteorological data, real-time IoT sensors (DHT22, BMP180, BH1750) connected to ESP32 micro-controllers and Raspberry Pi gateways for actual field data collection, soil nutrient data from the Government of India's official soil health card portal, along with publicly accessible datasets from the World Data Bank. The implementation is carried out in Python environment, using its extensive libraries of data science and ML.

1.7 Thesis Structure

The thesis is explained with seven chapters. From problem identification to final solution, each chapter is logically arranged to guide the readers for this research. Each chapter is based on the findings of the previous one, presenting a comprehensive and cohesive solution for PA.

1. Chapter 2: **The Foundation of Precision Agriculture – Understanding the Environment** shows the importance of environmental prediction in PA. It presents an ACO-ANN hybrid ML model for daily rainfall forecasting and shows its superior accuracy in a case study of Alipurduar, West Bengal.
2. Chapter 3: **From Knowledge to Action – Smart Irrigation for a Sustainable Future** builds directly on the environmental understanding by developing a smart irrigation system. This chapter details how predictive knowledge of evapotranspiration, derived from an AI model and real-time IoT data, can be translated into practical, resource-saving action in the field, thereby optimizing water use for a specific crop.
3. Chapter 4: **The Soil Beneath – Classifying Soil Fertility for Better Crop Choices** shifts the focus from the atmosphere and water to the soil, identifying a critical element of the agricultural ecosystem. It presents a classification model (ACO-XgBoost) to assess soil fertility based on a comprehensive nutrient profile. This work provides the essential basis for informed soil management and crop selection.
4. Chapter 5: **Building for the Future – Sustainable Crop and Fertilizer Recommendations** culminates the predictive work of the thesis by developing an eco-friendly crop and fertilizer advisory system. This chapter integrates the knowledge gained from the previous components to provide holistic, data-driven guidance to farmers, making a balance between high productivity and environmental stewardship.
5. Chapter 6: **The Final Frontier – Enhancing Farmer Decision-Making with XAI** focuses on making these systems usable and trustworthy for farmers by developing an XAI framework, which provides transparent and interpretable explanations for the system's predictions.
6. Chapter 7: **Conclusion** concludes the thesis with the major outcomes, highlighting the principal the significant inputs of the study, discussing the consequences for farming communities and the environment. This section also recognizes the constraints of the study and offers suggestions for further investigation.

Chapter 2

The Foundation of Precision Agriculture – Understanding the Environment

PA is a cutting-edge farming technique that maximizes crop yields and minimizes resource consumption by utilizing technological tools and data-centric approaches [132]. PA aims to enhance the effectiveness of farming practices by incorporating AI and IoT technologies. These technologies facilitate the monitoring and regulation of soil quality, crop growth, water usage, and weather-related conditions. Precision agriculture utilizes advanced tools including GPS, sensors, drones, and satellite images helping farmers to support data-guided decision-making. This approach boosts crop output and overall productivity while reducing the detrimental environmental impact in farming. To meet the food demand of an expanding global population while safeguarding natural resources, precision agriculture offers sustainable a viable means of overcoming the limitations associated with traditional farming methods [133].

Rainfall is a key determinant of agricultural performance, as it influences crop growth, irrigation needs, and overall farm productivity. To make informed decisions regarding irrigation scheduling, crop planting, and harvesting, it is essential to have precise rainfall forecasting [134]. Inadequate or excessive rainfall can lead to water shortages or flooding [135], which can cause crop damage, financial losses, and food security concerns. Predicting rainfall patterns becomes increasingly crucial in areas with highly variable rainfall. Early rainfall forecasting can prepare farmers for extreme weather events and manage water resources efficiently. As a result, accurate rainfall forecasting is essential for maintaining food security and sustainable farming practices.

Reliable rainfall prediction is extremely important for real-world applications, including agricultural planning, disaster readiness, and efficient water resource management. Rainfall prediction is a crucial aspect of weather prediction. Most agricultural decisions depend on rainfall. Due to climate change, rainfall patterns have undergone significant modifications over the last few

decades. For this reason, flood and drought-like situations are continuously occurring. Considering the mainland of West Bengal, numerous heavy rainfall cases have been reported throughout the state. For example, in the year 2000, a flood occurred in Kolkata due to cloudburst rain [136]; in 2007, floods happened in the region of North 24 Parganas, South 24 Parganas and Howrah due to heavy rainfall [137]. The sections of Darjeeling, Jalpaiguri, and Alipurduar flooded in 2016 [138]. In recent years, specifically in 2022, it was found that over 1 million people were affected by floods due to a depression developed in the Bay of Bengal [139]. Local meteorological factors, climatic conditions, and atmospheric circulations are uncertain factors that affect daily rainfall.

Traditional methods [140] for rainfall prediction have numerous limitations. Traditional methods need detailed information about physical processes directly involved in rainfall. AI and ML models have been shown as practical tools that can find complex phenomena from raw data. Nowadays, we find the use of hybrid models in various complex prediction tasks, such as rainfall forecasting. These models leverage the complementary capabilities of machine learning and optimization approaches to improve prediction performance. The fusion of multiple methodological approaches enables hybrid models to recognize a wider variety of underlying data patterns, thereby increasing their adaptability to the intricate nature of rainfall forecasting. Hybrid models are a type of AI models well-suited for rainfall predictions. Hybrid models can autonomously derive meaningful features directly from unprocessed data. Several prior research has demonstrated that hybrid models outperform conventional approaches.

The work in this chapter demonstrates the potential of mixed models for predicting daily rainfall. There are several advantages of using Hybrid models over conventional techniques.

2.1 The Role of Rainfall Prediction in Agriculture

Climate change has significantly disrupted traditional rainfall patterns, resulting in more intense and unpredictable weather [141]. Regions that were once known for dependable monsoons or continuous rainfall cycles are now experiencing variable rainfall, with some parts facing prolonged droughts and others experiencing extreme rainfall and inundation. For irrigation, crop planting, and harvesting, farmers depend on consistent rainfall. Consequently, fluctuations in weather patterns also disrupt their agricultural operations. Unpredictable weather poses direct threat to food security, as it can cause crop failures, water shortages, and flooding. Therefore, agronomist, farmer must adapt to these new conditions because accurate prediction of rainfall has become an essential component of sustainable farming.

Traditional methods of predicting rainfall often rely on past weather records and simple statistics [142]. These methods usually involve manual observations to estimate future weather conditions. Linear regression and simple moving averages are two strategies that work well in stable climate conditions. However, they are unable to deal with the complications that climate change has imposed. These models often fail to capture sudden and severe events, such as droughts or flash floods, and struggle to predict specific weather patterns accurately. Predictions may not accurately reflect local circumstances due to the limited geographic coverage of weather stations.

Traditional weather models frequently find it difficult to predict the increasing number of severe weather occurrences with precision [143]. Consequently, farmers often have no choice but to depend on outdated or incorrect predictions, leading to poor decision-making, thereby affecting crop production.

because conventional forecasting techniques have inherent shortcomings, there is a growing demand for sophisticated, data-driven models that provide real-time predictions with higher accuracy. ML and AI are highly effective tools for improving rainfall predictions [144]. These models can evaluate massive large volumes of information collected from multiple channels, including weather stations, satellites, and environmental sensors, to better identify patterns and forecast rainfall. AI and ML models continue to improve as they learn from new data, enabling them to adapt more accurately to evolving weather patterns. Furthermore, they can provide more precise predictions, which is critical for agriculture due to the significant variance in weather conditions from local to regional levels. Furthermore, these complex models support real-time forecasting, enabling farmers to take well-grounded decisions about disaster preparedness [145], crop production, and irrigation planning. Using AI-driven models, farmers may improve agricultural output and sustainability by optimizing water usage [146], reducing waste, and managing the hazards linked to severe weather events.

2.2 AI and ML in Weather Prediction

AI has revolutionized agriculture by giving sophisticated tools for increasing productivity, resource management, and decision-making. Environmental prediction is a crucial application of AI in agriculture, where AI models are utilized to forecast variables including factors like temperature, precipitation, and soil characteristics. Accurate forecasts are essential for agricultural practitioners because they enable the optimization of irrigation schedules, crop seeding, and harvesting decisions. AI tools, such as ML, can analyze large datasets from diverse channels, such as satellite systems, meteorological stations, and IoT-based sensors, to detect trends and make accurate predictions. This data-driven technique is far more versatile and precise than traditional methods, which are sometimes limited by static models and manual observations. AI helps farmers take better-informed and prompt decisions, improving sustainability, minimizing unnecessary resource expenditure, and enhancing agricultural productivity.

Hybrid models overcome the shortcomings of conventional prediction methods by combining several methodologies to maximize the benefits of each approach. Hybrid models enhance rainfall prediction accuracy and adaptability by employing a range of techniques. Conventional approaches frequently struggle to capture the intricate nature of weather behaviour and are ill-equipped to cope with the ever-changing impacts of climatic variations. Hybrid models enhance forecasting performance by combining multiple methods. For example, combining ACO with ANN provides a powerful way to improve prediction accuracy by optimizing feature selection and the learning process. Hybrid models excel in predicting complicated environmental processes because they solve difficulties like model over-fitting, data sparsity, and nonlinear correlations. Specifically, metaheuristic optimization algorithms like Ant Colony Optimization (ACO) address

data sparsity and over-fitting by dynamically pruning non-informative or redundant inputs during the feature selection phase, ensuring the base learner evaluates only a high-contributing parameter space. Concurrently, the multi-layered feed-forward architecture of the Artificial Neural Network (ANN) map complex non-linear correlations through specialized activation functions, preventing the computational bottlenecks or local minima traps common to isolated standalone estimators.

The ACO-ANN (Ant Colony Optimization-Artificial Neural Network) hybrid model is a cutting-edge technique for enhancing the accuracy of rainfall forecasting. This model serves as a reliable framework for forecasting rainfall behaviours that incorporates the strengths of ACO-ANN.

ACO is an optimization technique modeled on the search patterns exhibited by ants. It is employed in the ACO-ANN model to optimize the selection of input features for the neural network. ACO assists in identifying the most pertinent variables that impact rainfall patterns, thereby ensuring that the neural network learns from the most essential subset of data. ACO enhances the predictive capabilities of the model and reduces its complexity by optimizing the feature selection process.

The data is processed, and predictions are created using ANN based on the features chosen by ACO. With its capacity to absorb and analyze very large datasets, neural networks are practical tools for mimicking non-linear interactions and improving prediction accuracy. The combination of ACO and ANN yields a model that is both efficient and precise, capable of predicting daily rainfall with greater accuracy than traditional forecasting methods.

To find the most relevant input features (including parameters like temperature, moisture levels, wind behaviour, and past rainfall records), the ACO-ANN model first uses ACO. Then, a neural network is trained on these features to anticipate rainfall. This hybrid method not only enhances prediction accuracy but also makes the model more adaptable to changes in weather patterns over time. As a result, the ACO-ANN model presents a promising approach for enhancing rainfall forecasts and facilitating more informed agricultural decision-making.

2.3 Practical Applications in Precision Agriculture

Farmers are able to take well-judged and prompt actions based on accurate rainfall estimates. Farmers can plan irrigation schedules, decide the best time to plant crops and prepare for potential disasters by estimating the amount and timing of expected rainfall. For instance, in preparation for an anticipated period of severe rain, farmers may delay planting or alter irrigation to avoid water logging. In contrast, if a drought is expected, they may implement water conservation measures and expand irrigation. Farmers can take preventive actions, such as protecting crops or adjusting harvesting schedules if they receive early warnings of adverse weather events like droughts or storms. Finally, these decisions increase productivity, optimize resources, and reduce risks, thus improving the sustainability and resilience of agriculture.

The ACO-ANN model improves irrigation and water management by making more accurate rainfall predictions. When farmers have precise forecasts, they can adapt their irrigation procedures

to prevent both insufficient and excessive watering of crops. This helps reduce unnecessary water loss. This is especially important in locations where water availability is constrained. Farmers can ensure efficient water use by synchronizing irrigation plans with projected rainfall patterns. The ACO-ANN model improves prediction accuracy and provides farmers with immediate, data-driven observations, allowing them to formulate better-grounded choices regarding water usage by optimizing the selection of relevant data. This not only cuts expenses and reduces adverse environmental impact, but it also helps to ensure the overall sustainability of agricultural activities by conserving water.

Farmers can enhance crop selection and yields by utilizing precise rainfall predictions, such as those provided by the ACO-ANN model. Farmers can select the most suited crops for the upcoming season by calculating the expected rainfall and amount of precipitation. This decision is based on their water needs. For example, in areas where frequent rain is expected, it is possible to grow crops that require large amounts of water, such as rice. Farmers in drought-prone regions, on the other hand, might choose drought-resistant crops [147] like sorghum or millet that requires less water to thrive. Furthermore, farmers can sow at the optimal moment, ensuring that crops develop in favorable conditions and mature effectively due to early rainfall predictions. This proactive approach to crop management not only increases yields but also lowers the likelihood of crop loss caused by unforeseen variations in weather. The most important thing is that accurate weather predictions help farmers make wise choices that increase crop yields, which is suitable for food security and sustainable agriculture.

2.4 Background of the Study

The Singh et al.[148] propose the problem of rainfall forecasting in India, which is vital for the country's farming industry. The study emphasizes the application of deep learning and machine learning techniques for forecasting rainfall. By examining synoptic patterns in the atmosphere, the researchers Javier Diez-Sierra and Manuel Del Jesus[149] investigated the ability of eight machine learning and statistical approaches to accurately forecast long-lived day-to-day rainfall in a semi-arid region. Specifically, they were interested in how well these algorithms could do this. The rebuilt series are compared to the data that have already been gathered in order to determine which hyper-parameters produce the best results for each family of models. Pham et al. [150] aims to develop and evaluate cutting-edge AI models for predicting received daily precipitation in the mountainous province of Hoa Binh in Vietnam. The models employed meteorological variables as inputs and daily precipitation as output and were validated using various quality assessment criteria. Yazid et al. [151] explored the relationship between rainfall and runoff by using advanced ML models. To get better performance, they used MLP [152] and LSSVM [153] integrated with HHO [151] algorithm. The research seeks to enhance the accuracy of data-driven methods for forecasting runoff variations. In order to better estimate rainfall in Kenyir Lake, the federal state of Terengganu, Malaysia, Wanie et al. [154] created and evaluated multiple ML models. In their work, different scenarios and time frames are analyzed, and two different rainfall prediction strategies are examined and evaluated.

The limitations of standalone models are overcome through the use of hybrid machine learning frameworks, which capitalize on the complementary strengths of various algorithms. Metaheuristic optimizers are frequently employed in conjunction with fundamental predictive algorithms, such as ANN or SVM, to refine and adjust model parameters. Goyal et al. [155] illustrated a composite rainfall prediction model that integrated the FA with SVR. The model's superior performance in minimizing prediction error when compared to classical SVR or ARIMA models was primarily attributable to the improved parameter search capability of FA.

Ahmed et al. [156] constructed a rainfall forecasting framework by combining the HGSO algorithm with MLP neural networks in another noteworthy study. This model was validated across various climatic zones in South Asia and exhibited a significant increase in RMSE and MAE when compared with baseline ML models. The authors attributed this enhancement to the dynamic search process of HGSO, which is capable of identifying the global optimum in complex solution spaces by adapting over time.

Joshi et al. [157] introduced a new ensemble framework called "HydroStack," which utilizes stacking to combine multiple base ML learners, a form of meta-ensemble learning. Their model utilized a meta-regressor that was trained on the outcomes of DT, KNN, and GBM. The model achieved significant accuracy gains in predicting annual rainfall across Indian agro-climatic regions by capitalizing on the individual strengths of each learner through the layering mechanism.

Metaheuristic algorithms are beneficial in the optimization of ML model hyper-parameters when the search space is both extensive and intricate. Abdollahzadeh et al. [158] employed a Guided WOA in conjunction with an AD-PO to implement a dual-optimization approach. Ensemble ML models were trained on rainfall classification tasks using the hybrid optimization framework. The study showed a marked enhancement in classification precision and a faster convergence rate, particularly in datasets with high temporal variability.

From the literature survey, it has been found that many existing works have proposed hybrid models for various classification tasks. Classification in ML is a supervised learning method designed to determine category labels from the input variables. It involves allocating data samples into specific predefined groups or classes, such as distinguishing between spam and non-spam emails or classifying images of animals. However, several research gaps exist. Traditional methods for rainfall forecasting depend on static models and historical data, failing to capture the complexity of severe climatic occurrences, including flooding and prolonged dry spells. Additionally, these models lack accuracy for localized weather conditions and struggle to provide real-time predictions for specific agricultural contexts. Despite the rise of hybrid models, none of the studies have combined metaheuristic with ML models for regression tasks, specifically for predicting daily precipitation. This is critical because accurate, localized, real-time forecasting is necessary to address the increasing unpredictability of rainfall patterns due to climate change. The central objective of this study is to uncover underlying relationships between daily rainfall (output) and multiple climatic factors (inputs) and thus develop a prediction model for daily rainfall. The main focus of this chapter is to address the identified gaps by using a combined ACO-ANN hybrid methodology for more accurate, dynamic, and real-time rainfall forecasting. The key contributions of this chapter are outlined below:

1. The work introduces a hybrid methodology for forecasting daily rainfall.
2. ACO algorithm (Section 2.6.2) and various types of ANN like NNN (Section 2.6.1.1), WNN (Section 2.6.1.3), MNN (Section 2.6.1.2), BNN (Section 2.6.1.4), and TNN (Section 2.6.1.5) are combined to make the hybrid models.
3. Hybrid models were implemented in a MATLAB and Python Development environment and comparisons of various evaluation matrices with state-of-the-art work are done.

2.5 Case Study: Alipurduar, West Bengal

Alipurduar, a small region of West Bengal, is the main area of study. The region is situated in the northeastern part of West Bengal. The area covered approximately 3,136 square kilometers of land. In terms of its geographic location [159], the region lies between 26.4°N and 26.83°N and 89°E and 89.9°E. In terms of altitude, the district varies between 93 and 1,220 meters in elevation. The mean annual rainfall of Alipurduar is 350cm. The district experienced a deficit of precipitation in 2003 and 2013 [160], with only yearly average rainfall recorded being 150cm and 200cm, respectively. The district also experienced extreme rain in 2004 and 2017[161], where average annual rainfall lies between 700 cm to 800cm. Alipurduar has a lot of variations in the amount of precipitation. For this reason, the region is an excellent place to test the effectiveness of hybrid models for predicting daily rainfall.

This study spans from 3rd January 1981 to 20th March 2023. During this span daily values of 33 variables has been recorded. All these values got from the NASA LaRC POWER project[162], which is funded by the respected NASA Earth Science/Applied Science Program.

2.6 Methodology

The main objective of the developed algorithm is to predict how much rain will fall each day in Alipurduar, West Bengal. The proposed model has two parts: selecting features and making predictions. Choosing the features from the dataset represents the most essential stage because it has the most impact on how well the model works. ACO is used in this model to choose the best features. After selecting the features, different ANN are used to predict the daily rain. The experimentation block diagram is shown in Figure 2.1.

2.6.1 Artificial Neural Network (ANN)

McCulloch and Pitts created the ANN in 1943 [163]. A neuron is a simplified version of a biological neuron, whereas ANNs are composed of a significant number of neurons. The amount of neurons used can vary according to the specific issue that must be addressed. Neurons can be connected in various ways to construct an ANN model. According to researchers (e.g., [164],[165]), FF-BP is the most frequently employed type of ANN in a diverse array of fields,

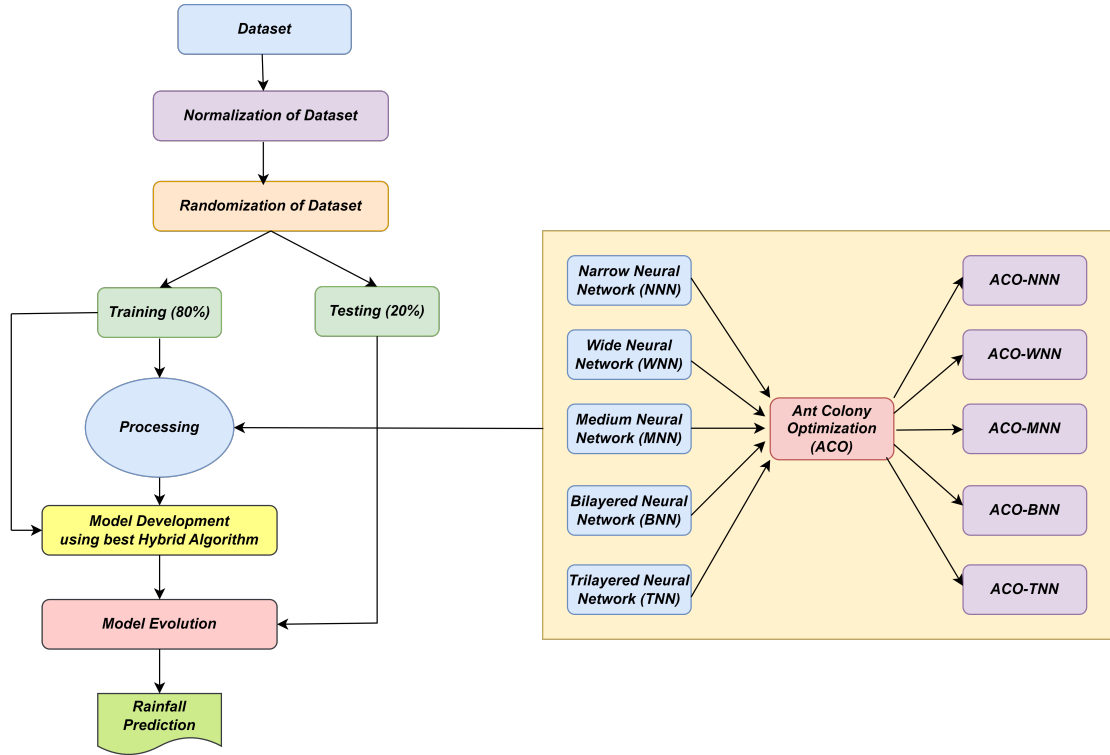


FIGURE 2.1: Experimentation block diagram.

including science and engineering. One of the most well-known FF-ANN is the MLP [164]. MLP is composed of numerous nodes or neurons that are connected by weights across three distinct layers—input, hidden, and output. Complex engineering problems can be resolved by an ANN model that contains a maximum of two concealed layers [166]. Typically, the count of hidden layers and their neurons is identified using an iterative experimentation approach [167]. Initially, each input parameter is designated an intended weight value falling between 0 and 1 in ANN. Subsequently, the weight is multiplied by each input, and the hidden layer(s) receive the accumulated total of these values. The bias (neuron weight) is typically equal to one, and the computed value is then incorporated into it. The attained value is assigned an additional weight in the subsequent round, and the cycle continues to the output layer. the concluding phase of ANN modeling involves aggregating the outputs from all layers. Subsequently, the training phase is completed, after which the model is readied for validation (or verification). In a FF-BP algorithm, information flows from the input layer toward the output layer, which is referred to as the "forward pass." The network then compares the achieved value(s) with the actual value and calculates the error. The error that is acquired is transmitted backward throughout the network to adjust the respective weights, a process known as a "backward pass." The errors of the errors associated with the training and validation sets are minimized using the previously described method. the error is minimized to a threshold specified by the cost function, such as the RMSE, by repeating the FF-BP procedure [168]. Nevertheless, to develop an appropriate ANN model, a database containing an adequate number of datasets is required. The objective of dividing ANN architectures into categories like NNN, MNN, WNN, BNN and TNN is to use the network's complexity, capacity, and computational requirements to the requirements of a specific task. To quantify these computational requirements and structural capacities, the total

number of trainable network parameters (P_{total}) and the structural capacity indicator (C_{struct}) are computed based on the specific architectural configurations. For a network with L hidden layers, where N_0 represents the input dimensionality, N_l denotes the number of neurons in hidden layer l , and N_{out} is the output dimension, the total number of weights and biases to compute is formally expressed as:

$$P_{\text{total}} = \sum_{l=1}^L (N_{l-1} \cdot N_l + N_l) + (N_L \cdot N_{\text{out}} + N_{\text{out}}) \quad (2.1)$$

The architectural capacity and comparative training footprint scale proportionally according to this parametric density, as systematically categorized in Table 2.1. For instance, a Narrow Neural Network (NNN) with a single layer minimizes memory footprints ($P_{\text{total}} = 111$), while a Trilayered Neural Network (TNN) intentionally expands feature mapping capability ($P_{\text{total}} = 1,151$) to capture deeper hidden patterns.

2.6.1.1 Narrow Neural Network(NNN)

NNN has fewer count of neurons present within every layer. The architectural design generally restricts its capacity to recognize intricate patterns or associations in the dataset. Narrow networks are usually faster than other types of ANN. However, they often struggle with high-dimensional data or intricate tasks that demand more neurons for effective feature extraction. Consequently, NNN are used for simpler problems or when computational resources are limited.

2.6.1.2 Medium Neural Network(MNN)

MNN has an intermediate quantity of neurons distributed across the layers. The architectural design enables the network to grasp complicated data patterns without the need for high-cost computational resources. It is applied extensively in practical scenarios, where the model must balance computational performance and efficiency. Medium networks usually have the ability to handle moderately complex tasks while solving common problems such as over-fitting or under-fitting. This makes them an automatic choice for many tasks, such as classification or regression.

2.6.1.3 Wide Neural Network(WNN)

WNN has many neurons in each layer. This architectural design helps the model in extracting a broad spectrum of attributes from the input data. WNN are generally more powerful and capable of managing deeply intricate patterns present within the dataset, which makes them well-suited for tasks with high-dimensional datasets. More neurons mean higher computational costs because the training process requires more resources. Wide networks might also be more prone to over-fitting if they're not managed properly.

2.6.1.4 Bilayered Neural Network(BNN)

BNN is composed of two levels an input level and an output level, with no hidden layers. BNN is simple and can only establish linear relationships between inputs and outputs. Even with the limited capabilities, it can still handle basic tasks like linear regression or binary classification. BNN are useful only for linearly separable problems.

2.6.1.5 Trilayered Neural Network(TNN)

TNN has three layers: an input layer, a hidden layer, and an output layer. The hidden layer allows the network to manage nonlinear relationships between inputs and outputs. TNN can solve complex problems, although it is still simpler compared to deeper network architectures. It is a versatile model that balances between computational efficiency and the ability to learn nonlinear patterns.

2.6.2 Ant Colony Optimization

Dorigo [169] developed ACO as a collaborative multi-agent approach for solving optimization tasks, including problems like the travelling salesman (TSM) scenario. This algorithm is derived from research on the behavior of actual ants in search of a potential source of nourishment. In this optimization algorithm, multiple artificial ants work together to determine the best possible solutions through an indirect sharing of information mediated by a synthetic pheromone. The concept of ACO involves the parallel exploration of multiple computational threads based on problem-specific local data. The exploration process depends on the effectiveness of the solutions found progressively, as well as a flexible memory framework representing the attributes of the problem [170]. In ACO, pheromone is generally associated with the solution elements that artificial ants rely on to generate new outcomes, which subsequently affects their decision-making. This synthetic pheromone builds up during execution through a learning scheme that reinforces successful solutions. A path containing higher pheromone levels has a greater probability of being selected. The pheromones will be associated with an evaporation rule, which will mitigate the possibility of substandard solutions. The ACO algorithm's superiority over other algorithms is primarily due to the appropriateness of the solutions produced by colonies.

2.6.2.1 State Transition Rule(STR)

This rule decreases the classification error in the predicted output. In the proposed approach, the role of the ant is to generate a subset of the given features. Each subset of features represents a specific state. The transition rule is shown as follows:

$$r = \begin{cases} \frac{[\xi_{ij}(t)]^\gamma [\zeta_{ij}]^\delta}{\sum_{m \in \text{allowed}_m} [\xi_{im}(t)]^\gamma [\zeta_{im}]^\delta} & \text{if } j \in \text{allowed}_m \\ 0 & \text{otherwise} \end{cases} \quad (2.2)$$

For an ant, q, ζ is the inversely proportional of computational expense, and parameter δ tells how important heuristics are compared to pheromones. $\xi(g)$ is the measurement of feature (g) in the subset (s).

2.6.2.2 Global Updating Rule (GUR)

This guideline produces feature subsets that exhibit a notably low classification error. It is beneficial for identifying the groups of selected features that produces the lowest error rate during the present iteration. As per this guideline, both the pheromone intensity associated with each feature and the most optimal feature group are incremented. As a result, the ant identifying the best solution is permitted to place pheromone on the chosen feature set. To add dynamism in the search procedure in a more dynamic way, the selection process may be further enhanced by an additional comparative rule. The ant explores states to the nearest current state identified by previous iterations. The rule is applied once each ant has identified its respective solution. The following Eq. 2.3 below demonstrates how the pheromone trail level is updated based on the global update rule:

$$\xi_{ij}(t) = (1 - \varrho) \cdot \xi_{ij}(t) + \varrho \cdot \xi_0 \quad (2.3)$$

$\xi_{ij}(t)$ represents the pheromone quantity at time (t) on edge (i, j), while ϱ denotes the pheromone evaporation rate such that $0 < \varrho < 1$ ξ_0 refers to the initial pheromone value.

2.6.2.3 Local Updating Rule (LUR)

This rule lowers the significance of non-essential features, but it also enables the ant to explore previously unexplored features. It either minimizes the pheromone level of these features or maintains the existing pheromone level, depending on the feature selected. When the LUR rule is applied, the pheromone value of every feature present in the superior subset from earlier iterations. However, the pheromone levels of features not selected in the best subset remain unchanged. Therefore, the pheromone level will not decrease below its initial value. The local pheromone update formula is presented in Eq. 2.4, where m denotes the pheromone evaporation parameter.

$$\xi_{ij}(t) = (1 - m) \cdot \xi_{ij}(t) + m \cdot \xi_0 \quad (2.4)$$

Both global and local updating rules slightly decrease the pheromone levels. As noted previously, the pheromone amount rises for features that belong to the high-quality subset in the current iteration. In contrast, the value for features not part of the good subset remains unchanged. This mechanism prevents the ants from converging on a single, generalized path, encouraging more diverse exploration.

Algorithm 1 Feature Subset Selection using ACO

- 1: Initialize the number of ants.
- 2: Initially the dataset's pheromone trail amount for every feature may initially be identical.. The ant chooses its traits based on probability and state distribution rules.
- 3: Using Eq. 2.2, every ant q prepared in step 1 creates a subset from the original dataset.
- 4: At this stage, the global update rule (Eq. 2.3) is applied to the feature subset generated by the ants in the earlier step. The global update rule is applied to the feature group exhibiting the minimum regression error. To improve the popularity of features, awards would be given to those who made the fewest mistakes.
- 5: In this step, the remaining ants are updated using the local updating rule (Eq. 2.4). That rule may help keep the pheromone quantity of ants the same or change it.

2.6.3 ACO-ANN Hybrid Model

With ACO and ANN algorithms, the hybrid algorithm ACO-ANN is generated (shown in Fig. 2.2). ACO-ANN hybrid algorithm is employed to choose the relevant attributes from the daily rainfall dataset. In the proposed work, ACO is used for feature selection as mentioned in steps 4 and 5 of the Algorithm 1, and ANN is used for rainfall prediction. The original dataset S comprises the complete set consisting of ' N ' distinct attributes. From the original dataset S , subsets of features such as s_1 , s_2 , and s_3 will be generated, and already trained ANN will take those subsets as input for the prediction.

At first, a number of ants are prepared depending on the total amount of features. In the first phase, the prepared ants choose a subgroup of factors from the actual dataset, which contains ' N ' features.

The ACO-ANN hybrid model is outlined in Algorithm 1. The five steps of Algorithm 1 will iterate until the best subset with the least amount of regression error is found for the given start and end numbers.

2.6.4 Data Pre-processing

The data preprocessing phase was carried out comprehensively in Python Development Environment. Missing entries were addressed using an effective imputation approach known as the KNN Imputer. Using the Min-Max Scaler method dataset was scaled between 0 and 1 to ignore the scaling issues; after that the dataset was partitioned into an 80:20 ratio for training and testing purposes. To check the forecasting skill of the model, k-fold cross-validation was applied. The dataset is roughly segmented into five (k) equal folds, and the model is iteratively trained on the remaining $4(k-1)$ folds.

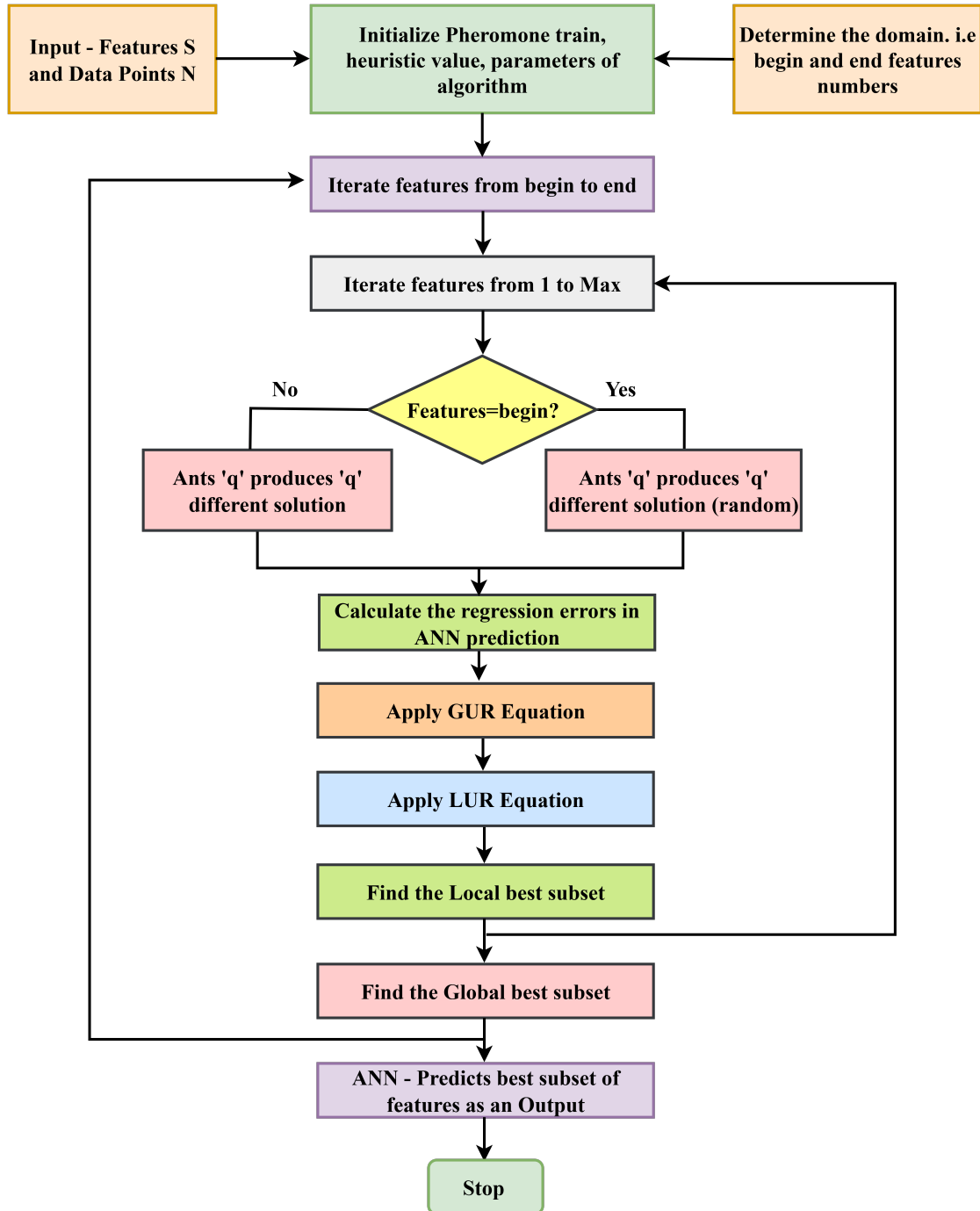


FIGURE 2.2: ACO-ANN hybrid model flowchart.

2.6.4.1 Feature Selection Using ACO

The original meteorological dataset consisted of 33 climatic variables. These variables included temperature, humidity, atmospheric pressure, wind speed, sunshine duration, evaporation, cloud cover, and other weather-related attributes. To identify the most relevant predictors for rainfall forecasting, a wrapper-based feature selection approach using Ant Colony Optimization (ACO) was employed.

In the proposed approach, each ant represented a candidate feature subset. The quality of a feature subset was evaluated using the prediction performance of the ANN model. Feature subsets producing higher prediction accuracy received larger pheromone updates. Through iterative pheromone updating and probabilistic search, ACO gradually converged toward the most informative subset of features.

After convergence, nine features were identified as the most significant predictors and were subsequently used as inputs to the proposed ACO-ANN rainfall prediction model.

2.6.5 Experimentation

TABLE 2.1: Best Architecture Configuration

Neural Network	Input Layer	Hidden Layer 1	Hidden Layer 2	Hidden Layer 3	Output Layer
NNN	9	10	-	-	1
WNN	9	30	-	-	1
MNN	9	20	-	-	1
BNN	9	15	30	-	1
TNN	9	15	30	20	1

Initially, several benchmark regression models, including Linear Regression (LR), Support Vector Regression (SVR), Gaussian Process Regression (GPR), Ensemble Regression (ER), and different ANN architectures such as Narrow Neural Network (NNN), Medium Neural Network (MNN), Wide Neural Network (WNN), Bilayered Neural Network (BNN), and Trilayered Neural Network (TNN), were developed and evaluated. The comparative performance results presented in Table 3 indicate that ANN-based models consistently achieved higher prediction accuracy and lower error values than conventional regression models. Among the ANN variants, TNN demonstrated the best performance and was therefore selected for subsequent hybridization with ACO. The ANN were combined with a metaheuristic algorithm . As already discussed, the proposed model has two parts: feature selection and prediction of daily rainfall. A total of 9 selected features as shown in Fig. 2.3, were obtained from the first phase of the hybrid model. In the second phase, several ANN configurations were tested by changing the count of neurons in each layer. This was done to find the best architecture configuration. Details about the suggested hybrid model's final architecture are shown in Table 2.1. During training and testing for all five-folds, the structure of the model was kept the same. After the training was done successfully, the testing input set was put into the model to see how well it worked.

2.7 Result Analysis

At first, all kinds of regression models, namely LR, DT, SVM, GPR, Ensembles of Trees, and ANN were tested on the dataset. It was observed that ANN worked superior to other types of Regression models, observed in Table 2.2. Next, Hybrid Models have been implemented, and selected features like Cloud Amount, Soil Moisture Profile, Soil Wetness in the Root Zone, Surface Soil Moisture Index, Atmospheric pressure near the surface, Specific moisture content measured at a height of two meters, Relative humidity recorded at the two-meter level, Air temperature taken at a height of two meters and Wind Direction at two meters have been used as input to improve results even more, Shown in Figure 2.3. Result comparisons of various hybrid models are shown in Table 2.3. The table clearly indicates that ACO-NNN works better than other types of hybrid models, shown in Fig. 2.4. In each area, the most impressive results are written in bold. The results are also shown in pictures to give a clearer idea.

The key takeaway is that the simpler model outperformed the more complex ones. While the researchers tested wider and deeper networks, the narrow model combined with the ACO method delivered the most accurate results. This study demonstrates that focusing on the right factors, such as moisture, temperature, and wind, can lead to a better performance. For farmers, this is encouraging that a carefully selected set of inputs and a relatively simple model can provide reliable rainfall forecasts, making the approach both practical and realistic.

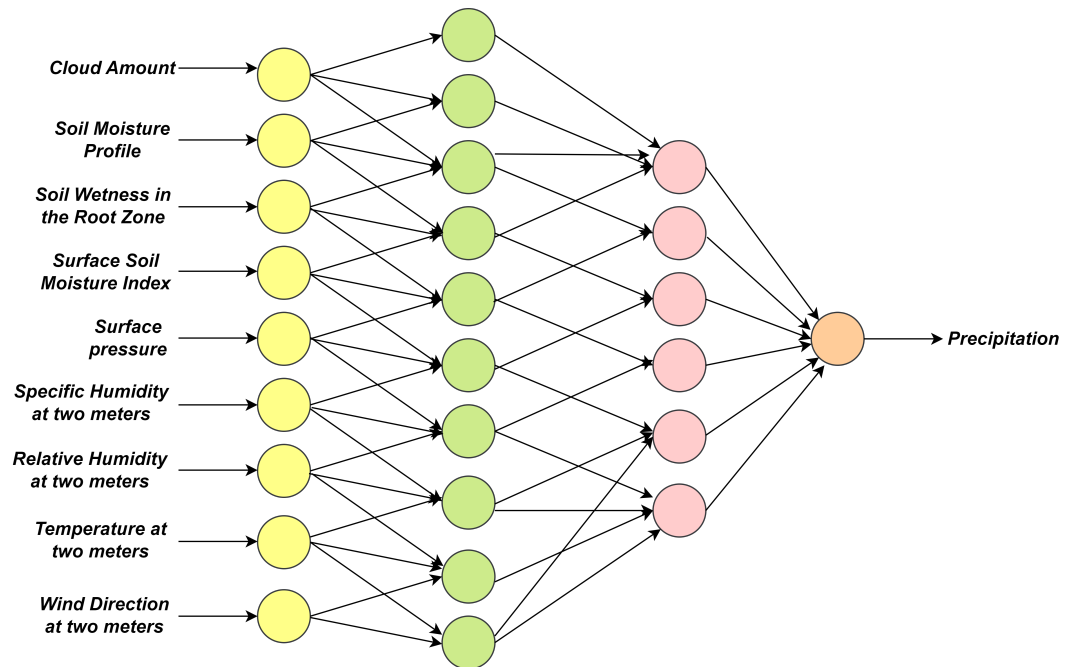


FIGURE 2.3: Abstract ANN model with Selected Features and output

Figure 2.3 illustrates the final architecture of the ANN model after ACO-based feature selection. Although the original dataset contained 33 meteorological variables, ACO identified nine highly informative features that contributed most significantly to rainfall prediction performance. These selected features were subsequently used as inputs to the ANN model.

TABLE 2.2: Results comparisons of various Regression Models.

Model/Approach	RMSE	R^2	MSE	MAE
Neural Networks				
NNN	4.0037	1.9324	10.0222	2.4831
MNN	6.2803	1.7286	28.3881	2.6649
WNN	3.9183	1.9312	9.5163	2.4855
BNN	4.0037	1.9332	10.0222	2.4831
TNN	4.8733	1.8537	16.1003	2.5437

TABLE 2.3: Results of comparison of various hybrid models.

Neural Network	RMSE	R^2	MSE	MAE
ACO-NNN	2.8402	0.92	8.0665	1.3978
ACO-WNN	2.9248	0.91	8.5546	1.4289
ACO-MNN	3.1334	0.9	9.8183	1.612
ACO-BNN	2.9133	0.91	8.4874	1.3632
ACO-TNN	2.8862	0.93	8.3303	1.3608

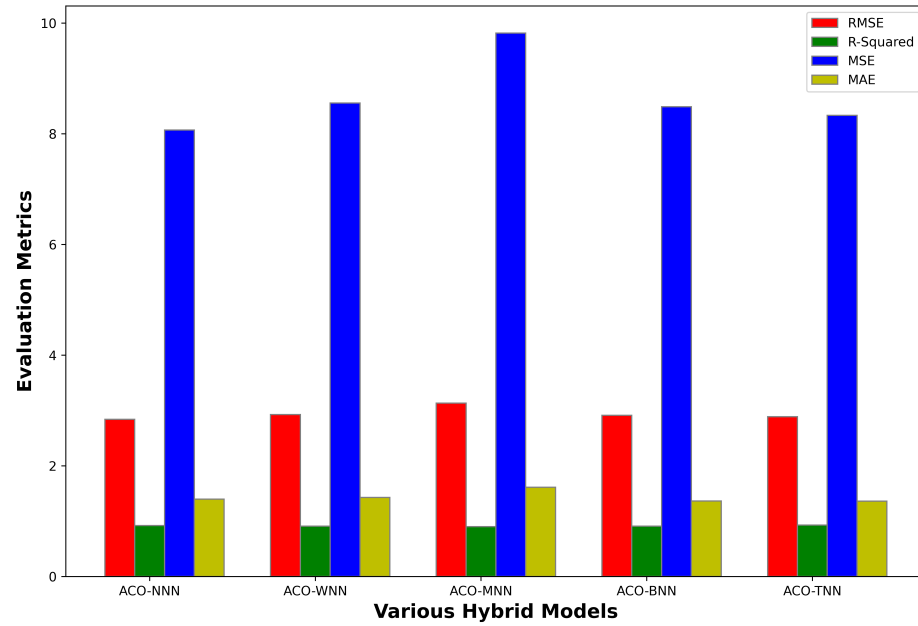


FIGURE 2.4: Result Comparison of various Hybrid Models.

To further validate the hybrid model performance, a comparative evaluation of various wrapper

based feature selection algorithms was conducted, as presented in Table 2.4. The analysis highlights that metaheuristic-based methods generally outperform than the conventional sequential techniques with respect to improving predictive precision and reducing errors. Among these, the ACO-based optimization approach demonstrated the most superior performance, achieving the lowest RMSE (2.84–3.13) and MAE (1.36–1.61) values, alongside the highest R^2 (0.90–0.93). These results indicate that ACO effectively identifies the most significant and informative input variables, thereby enhancing the model’s ability to generate accurate predictions. In comparison, traditional sequential techniques like Forward Selection and Backward Elimination showed higher error metrics, emphasizing the advantage of adaptive, population-based optimization approaches. Overall, integrating ACO with the neural network not only improved model precision but also established an optimal balance between model simplicity and predictive performance, rendering it the strongest feature selection approach in this study.

TABLE 2.4: Results comparisons of various Wrapper Based Feature Selection Methods

Model/Approach	RMSE	R^2	MSE	MAE
Sequential Methods				
Sequential Forward Selection	3.45 – 3.72	0.84 – 0.87	11.90 – 13.80	1.78 – 1.96
Sequential Backward Elimination	3.36 – 3.60	0.85 – 0.88	11.30 – 12.90	1.72 – 1.89
Bidirectional Search	3.22 – 3.48	0.86 – 0.89	10.40 – 12.10	1.65 – 1.82
Recursive Feature Elimination	3.08 – 3.31	0.88 – 0.90	9.50 – 10.90	1.55 – 1.70
Metaheuristic Methods				
Genetic Algorithms	3.02 – 3.25	0.88 – 0.91	9.10 – 10.60	1.49 – 1.68
Simulated Annealing	2.97 – 3.21	0.89 – 0.92	8.80 – 10.30	1.44 – 1.63
Particle Swarm Optimization	2.90 – 3.18	0.89 – 0.92	8.40 – 9.90	1.39 – 1.60
Ant Colony Optimization	2.84 – 3.13	0.90 – 0.93	8.06 – 9.81	1.36 – 1.61

2.8 Chapter Summary

This chapter highlights the importance of understanding the environment in precision agriculture, with specific attention to rainfall forecasting. The impact of climate variability on conventional rainfall behavior [141], resulting in heightened uncertainty and more severe weather events is observed. The constraints of traditional rainfall forecasting techniques, which rely on human observations and static models, were highlighted, underscoring the need for more sophisticated, data-driven models to enhance agricultural efficiency.

As weather patterns become increasingly complex, AI and ML are becoming essential tools for enhancing rainfall forecasts. These cutting-edge technologies help farmers to make informed decisions on irrigation, agricultural cultivation, and disaster resilience. To improve resource management and reduce uncertainty, farmers can more accurately predict weather changes by utilizing data-driven models.

Based on this environmental understanding, the next chapter focuses on optimizing water consumption in agriculture. The theme of the next chapter titled "From Knowledge to Action – Smart Irrigation for a Sustainable Future" will examine the advancement of intelligent irrigation systems using AI and IoT technologies. This chapter will demonstrate how ML models, combined with genetic algorithms, accurately estimate the water needs of crops like coriander. The chapter will examine the issues posed by inefficient irrigation systems and how the integration of AI-driven forecasts with IoT sensors might provide a more intelligent and sustainable approach to water management. This integration aims to reduce water wastage, lower farming expenses, and enhance the long-term sustainability of farming activities.

Chapter 3

From Knowledge to Action – Smart Irrigation for a Sustainable Future

Modern farming is not just about understanding the environment; it is about taking informed actions based on data, preferable latest. The previous chapter discussed how the proposed model ACO-ANN can predict rain; however, simply predicting rain is not enough. To support sustainable farming, predictions need to be used effectively. This is especially important in areas impacted by climate change. Timely decisions based on these predictions can make a big difference. This chapter expands on that idea by demonstrating how AI and the IoT can transform weather forecasts into intelligent methods for watering plants.

Climate change has made it increasingly difficult to predict the timing of the rainfall. Some areas experience drought like conditions, while others are getting too much rain. Farmers face challenges in determining when and how much to water their crops [171]. In places like West Bengal, traditional methods of watering plants are no longer effective. Farmers still need a system that can turn weather forecasts into particular actions on the field, activities that take into account both the weather and the needs of each crop.

This is where intelligent irrigation is beneficial. Innovative systems don't follow a set watering schedule. Instead, they figure out how much water each plant needs at different phases of growth. For instance, coriander (*Coriandrum sativum*) [172] needs varying quantities of water when it is growing, bolting, and flowering. A system that takes into account both the plant's biology and the surroundings can better and more efficiently regulate irrigation.

The primary obstacle in agriculture is water scarcity. Inefficient agricultural methods waste 60% of irrigation water per year [173][174]. A few challenges faced by agriculture globally are low productivity, urbanization, and land degradation. Therefore, a sustainable solution is required to feed the estimated billions of people inhabiting the planet by 2050 [175]. Precision irrigation is a vital solution based on the Evapotranspiration (EV_{T0}) rate. The primary foundation of

water management in agriculture is EV_{T0} (Section ??), determined by total plant transpiration and soil evaporation [176] [177]. Farmers can maximize irrigation efficiency and minimize water consumption by analyzing EV_{T0} . The accurate prediction of EV_{T0} is crucial for implementing precision irrigation.

Some well-known traditional methods of calculating EV_{T0} are complex and require forecast data such as temperature, humidity, and solar radiation. In this work [178], a precise method is proposed to calculate the EV_{T0} rate using a minimal set of environmental variables collected directly from the crop field via Internet of Things(IoT) sensors, avoiding the requirement of forecast data.

This chapter discusses an intelligent irrigation system that leverages both AI and IoT to address this issue. Sensors in the field gather real-time information, including temperature, humidity and soil moisture. This data is real-time and particular to the area than regular weather data. The data obtained is then processed through a hybrid ML model, specifically a GA-ANN [179], which predicts a value known as reference EV_{T0} . This indicates the amount of water the crop is losing and the amount it needs.

Then, based on the plant's growth stage, the estimated water needs are used to determine the amount of irrigation required. There is also an Android app and an IoT-controlled water pump that automatically gives the proper amount of water to the plants. The system can be remotely monitored and managed by farmers, which reduces labor and eliminates the need for guesswork.

ML can effectively estimate EV_{T0} using a small number of essential environmental parameters. Previous studies show that ML effectively resolves the challenges associated with the forecast data for calculating EV_{T0} [180]. Currently, ML solutions mainly depend on weather station data, which may not accurately portray unique field conditions. In the proposed system, IoT sensors gather real-time data on temperature, humidity, and other crucial factors from the coriander field, and the ML model estimates the precise amount of water for irrigation based on these collected data. This system is designed to meet specific water requirements according to the growth patterns of coriander plants through adjustments in crop parameters. While the system is developed for coriander plants, its flexible architecture and IoT-driven design can potentially adapt to various crops, climates, and geographical regions. Thus, the proposed solution makes it possible to create excellent watering systems that meet the crops' needs and field conditions.

3.1 Challenges of Traditional Irrigation Systems

Flood irrigation, sprinkler systems, and surface irrigation are all examples of traditional irrigation methods [181] that typically operate on set schedules or with predetermined amounts of water. These systems are designed to deliver a set amount of water to crops, based on general assumptions about the water needs of different types of plants. However, this strategy is ineffective because it overlooks variations in weather, soil moisture, and the various stages of crop growth. As climate change causes major variability in weather, the timing and strength of rain have become less predictable. If it rains a lot or not at all in some places, using traditional

irrigation systems can waste a lot of water or stress crops[182]. For instance, if it rains heavily immediately after watering, the extra water cannot be absorbed effectively, which can lead to soil erosion, water logging, and crop damage. If there is a drought or insufficient rainfall, however, traditional systems may not provide the crops with enough water, which can cause drought stress, lower yields, and even crop loss[182]. Due to this, conventional irrigation systems are significantly less effective, particularly when the weather is unpredictable.

Over-irrigation and under-irrigation [183] are the two most common issues with traditional irrigation systems. Both the problems are bad for the environment and the health of crops. Over irrigation happens when too much water is used on crops. This excess water can wash away vital nutrients from the soil, leading to nutrient imbalances. It also helps harmful organisms, such as bacteria and fungi to flourish which can harm plants and reduce food production. Under-irrigation, on the other hand, arises when crops don't get enough water. This may occur due to incorrect watering schedules or climatic variables, such as high temperatures or dry conditions. If crops don't get enough water, they experience water stress, which can slow down their growth and result in lower yields, poorer quality, or even complete crop failure. This not only affects the crops but also increases water usage, reduces productivity, and raises overall costs for farmers. Another problem with traditional irrigation methods is that they require a significant monitoring and managing [182]. Farmers generally use their knowledge to determine how much water they need, which can be a time-consuming process and may not be always effective. In certain circumstances, additional workers are required to monitor the water flow, make adjustments to the systems, and address any issues that arise with irrigation.

There is an obvious need for precision irrigation, an advanced technology that can deliver water in real time, based on the specific needs of crops and environmental conditions. Precision irrigation considers several factors, including temperature, humidity, soil moisture, and the stage of crop growth. It ensures that crops receive the proper amount of water at the right time, preventing them from receiving too much or too little. Precision irrigation can maximize water use and enhance crop health[184] by leveraging technology such as AI-based algorithms and IoT sensors. IoT sensors can monitor soil moisture levels, and AI algorithms can automatically adjust watering amounts based on real-time data. This results in increased crop production, reduced labor costs, and more efficient water use.

3.2 Background of the Study

An extensive literature review on smart irrigation and EV_{T0} rate determination is presented below with reputable bibliographic indices. Multiple studies have explored ML methods for predicting crop and reference (EV_{T0}). Chen et al. [193] investigated TCN, LSTM, and DNN to estimate EV_{T0} with restricted environmental data. This research evaluated the efficacy of the models by comparing them to well-known scientific approaches. Scientific approaches like Hargreaves-Samani and Ritchie [194] were used. They emphasized the effectiveness of temperature and humidity-based models and showed outstanding accuracy with an R^2 value of 0.948 and RMSE of 0.096 [195]. Singh et al. [196] assessed the performance of RF, ANN, XGBoost, and

TABLE 3.1: Feature comparisons of proposed work with existing works

Work(Ref.)	Column ¹	Column ²	Column ³	Column ⁴	Column ⁵	Column ⁶	Column ⁷
Bashir et al.[185]	×	×	×	×	×	×	×
Mostafa et al.[186]	×	×	×	×	×	×	×
Yildirim et al.[187]	×	×	×	×	×	×	×
Tr et al.[188]	×	×	×	×	×	×	×
Kocian et al.[189]	×	×	✓	✓	×	×	×
Dong et al.[190]	×	×	×	×	×	×	×
Bidabadi et al.[191]	×	✓	×	×	×	×	×
Rajput et al.[192]	×	×	×	×	×	×	×
Paria et al.(Proposed work) [172]	✓	✓	✓	✓	✓	✓	✓

^aUsed IoT Sensor for Real-time data collection.^bNet solar radiation calculated from collected data using BH1750 sensors.^cUsed Hybridization with Metaheuristic Algorithm for prediction.^d EV_{T0} considered for the specific crop.^e EV_{T0} predicted for each Life cycle of considered Crop.^fReal-time irrigation depends on predicted EV_{T0} rate.^gAndroid application for real-time irrigation monitoring.

CNN in estimating monthly EV_{T0} using hourly data. Their findings demonstrated that CNN models exhibited greater accuracy. Compared to other alternative approaches, the CNN model yields better result in terms of RMSE. Patil and Deka suggested calculating weekly EV_{T0} based on several environmental factors. They applied their solution in India and contrasted it with the Hargreaves (H) method. The RMSE for the suggested method is 0.43 in Jodhpur and 0.33 in Patiala [197]. Keshtegar and Kisi [198] developed a hybrid response surface method (HRSM) by incorporating exponential approximation and second-order polynomial approximation into the response surface method (RSM). They used their technique to calculate Turkey's EV_{T0} .

Previous studies on intelligent irrigation methods used wireless sensors and the Internet of Things(IoT) to oversee crop fields and suggested watering schedules according to environmental factors. Research gaps exists as these systems overlook crop-specific EV_{T0} rates, essential for accurate water usage. Some ML methods use limited meteorological data to predict EV_{T0} but do not incorporate real-time field measurements. Many studies overlooked or calculated net solar radiation using available temperature and humidity data, which may not accurately represent the actual solar radiation in a crop field. Therefore, it is necessary to estimate an efficient solar radiation parameter for calculating EV_{T0} .

Beyond these research gaps the present work presents unique features that are absent in most recent research works. Table 3.1 presents a detailed comparison of these novel aspects, solidifying the innovative nature of the proposed approach.

3.3 Contributions of the present work

- A real-time intelligent irrigation system designed for the different growth stages of coriander plants using various IoT sensors and hybrid ML.
- A novel method is proposed to estimate net solar radiation through sunshine duration collected using BH1750 sensor.
- For the prediction of EV_{T0} , nine different hybrid ANN models based on mutation and crossover operations of the GA are developed.
- An Android application is developed and successfully implemented for remote monitoring and management of intelligent irrigation system in a real coriander field.

3.4 The Estimation of EV_{T0} using Climate-Based Methodologies

Among the various methods available for calculating EV_{T0} , the PM model was reported to be the most effective one for EV_{T0} estimation [199].

3.4.1 PM Model

The PM model uses temperature, humidity, wind speed, sun radiation, and atmospheric pressure of the location to determine EV_{T0} . Equation 3.1 shows the PM model.

$$EV_{T0} = \frac{(0.4079\delta(S_n - H)) + \Gamma \frac{900}{T+273} v_2 (f_s - f_a)}{\delta + \Gamma (1 + 0.3399v_2)} \quad (3.1)$$

where EV_{T0} : refers to the reference evapotranspiration, which is considered in $mmday^{-1}$ unit, Γ : refers to the psychrometric constant, which is regarded as in KPa^0C^{-1} unit, δ : refers to the slope vapour pressure, which is also considered in KPa^0C^{-1} unit, v_2 : refers to the wind speed, which is regarded as in ms^{-1} unit, f_s : refers to the saturated vapour pressure, which is considered in KPa unit, f_a : refers to the actual vapour pressure, which is regarded as in KPa unit, H : refers to the soil heat flux density, which is considered in $MJm^{-2}day^{-1}$ unit, S_n : refers to the net radiation in the crop field, which is considered in $MJm^{-2}day^{-1}$ unit.

3.4.2 Proposed Method to Estimate Solar Radiation

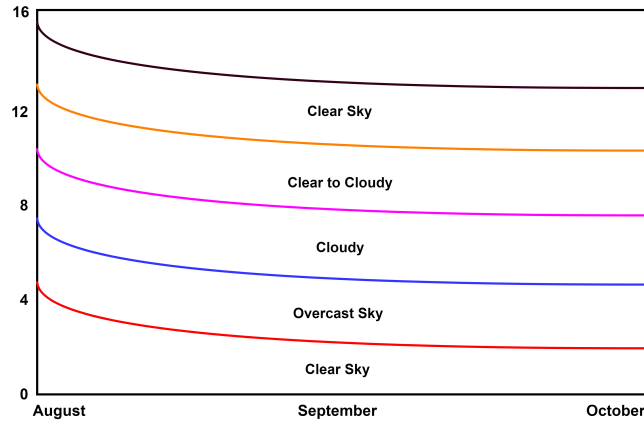


FIGURE 3.1: Sunshine duration with respect to the five cloudiness classifications.

Instead of depending on costly equipment, the proposed method uses BH1750 sensor, a cost-efficient and easily accessible alternative. This sensor accurately classifies sunshine duration into five categories of cloudiness: clear sky, clear to cloudy, cloudy, overcast, and rainy (Fig.3.1). By substituting these data into the selected equations, the net solar radiation can be calculated without requiring advanced instruments.

The estimation of global or short-wave solar radiation is derived by applying the Angstrom equation of sunshine duration [199].

$$S_n = \left(b_n + c_n \frac{o}{O} \right) S_a \quad (3.2)$$

Where S_n : shortwave radiation ($MJm^{-2} day^{-1}$) unit, S_a : extraterrestrial radiation ($MJmm^{-2}day^{-1}$) unit, o : actual duration of sunshine (h), O : maximum sunshine duration (h),

b_n and c_n are parameters expressing the fraction of S_a outreach on overcast days (when $o = 0$), $b_n + c_n$ is the extraterrestrial radiation outreach of the soil surface on clear days ($o = O$).

The Duffie and Beckman equation is used to calculate the daily extraterrestrial radiation [200].

$$S_a = \frac{24(60)}{\pi} H_c e_{rd} [\Omega_n \sin(\Phi) \sin(\Delta) + \cos(\Phi) \cos(\Delta) \sin(\Omega_n)] \quad (3.3)$$

Where S_a : extraterrestrial radiation ($\text{MJm}^{-2} \text{ day}^{-1}$); H_c : solar constant ($0.0819 \text{ MJm}^{-2} \text{ min}^{-1}$); E_{rd} : inverse relative distance Earth-Sun; Ω_n : sunset hour angle; and Φ : latitude of the location (rad) and Δ : solar declination.

Net radiation is calculated by subtracting the net shortwave radiation from the net longwave radiation.

$$S_n = S_{ns} - S_{nl} \quad (3.4)$$

$$S_{ns} = (1 - \beta) S_n \quad (3.5)$$

where S_n and S_{ns} are the global and net shortwave radiation ($\text{MJm}^{-2} \text{ day}^{-1}$) and β is the albedo [201] of the reference crop ($\beta = 0.26$), and $S_{nl} = \varsigma \left[\frac{Temp_{\max, \text{Kel}}^4 + Temp_{\min, \text{Kel}}^4}{2} \right] (0.339 - 0.139\sqrt{f_a})$

$$\times \left(1.349 \frac{S_n}{S_{sr}} - 0.349 \right) \quad (3.6)$$

where S_{nl} : total long wave radiation ($\text{MJm}^{-2} \text{ day}^{-1}$) unit, S_{sr} : radiation on a clear-sky ($\text{MJm}^{-2} \text{ day}^{-1}$), $Temp_{\max, \text{Kel}}$ and $Temp_{\min, \text{Kel}}$ are the maximum temperature and minimum temperature of air (Kel), f_a : actual vapour pressure (kPa) and ς : Stefan-Boltzmann constant [202] for a day ($4.903 \times 10^{-9} \text{ MJKel}^{-4} \text{ n}^{-2} \text{ day}^{-1}$).

$$S_{so} = (0.75 + 210^{-5}y) S_a \quad (3.7)$$

where y is the elevation of the crop field, S_a is given by equation 3.3, and 0.75 is the value for $b_n + c_n$ when $b_n = 0.25$ and $c_n = 0.50$. By substituting the parameter 0.75 with 0.25 in this equation to reflect overcast conditions, the value for the solar radiation on a wholly overcast day, denoted as S_c is obtained. The duration of daylight is determined by the sunset hour angle Ω_n :

$$O = \frac{24}{\pi} \Omega_n \quad (3.8)$$

To estimate the solar radiation, it is assumed that the o (predicted sunshine duration) in equation 3.2 for each cloudiness condition are evenly distributed between 0 and O , as shown in Fig. S1[†]. The equation below is used to calculate the value of o that will be used in equation 3.2 for the five predicted cloudiness conditions:

$$o = co \times O \quad (3.9)$$

Let o represent the predicted sunshine duration in hours and co represent the adjustment coefficient. The adjustment coefficient has sequential values of 0.10, 0.8, 0.6, 0.4, and 0.2 for the

five cloudiness conditions: clear sky, clear to partially cloudy, cloudy, overcast, and rainy. O represents the daylight duration in hours, as shown in the equation. 3.8.

3.5 Crop Water Requirement Estimation

EV_{T0} is used to estimate the water requirements of crops at different phases of the plant life cycle. The EV_{T0c} rate determines the water requirement of the reference crop in a given climate. The crop coefficient (C_c)(which considers the crop's distinct water-use characteristics at various growth stages) is used to modify the value of a particular crop. Equation 3.10 illustrates this computation by calculating the EV_{T0c} rate using the relevant C_c and the life cycle of the reference crop.

$$EV_{T0c} = C_c \times EV_{T0} \quad (3.10)$$

This work examined the irrigation requirements of coriander (*Coriandrum sativum*) at every growth stage. Three main stages characterize the coriander life cycle. In this work, the C_c values for different growth stages of coriander: vegetative growth stage ($C_c=0.21$), bolting stage ($C_c=0.72$), and flowering and seed development stage ($C_c=0.79$) are used from previous literature [203].

3.6 Artificial neural networks (ANNs)

ANN is inspired by the human brain, are powerful tools for processing information. They consist of interconnected processing units called neurons arranged in layers. Information flows from the input layer, through hidden layers (if any), to the output layer. Each neuron applies a simple mathematical function to its inputs, determining its output signal. The strength of the connections between neurons is represented by weights, which are adjusted during training.

Training an ANN involves defining its structure (number of layers and neurons) and adjusting the weights. The BP algorithm is a standard training method that compares the network's output to the desired output and then iteratively adjusts the weights to minimize the error. While BP is widely used, it can sometimes take time to learn and get stuck in sub-optimal solutions.

3.7 Genetic Algorithm (GA)

Inspired by Darwin's theory of evolution, GA are a powerful tool for solving complex optimization problems. They were developed in the 1960s; GA work by mimicking natural selection. A population of potential solutions (chromosomes) is randomly generated, and their fitness is evaluated based on the problem. Similar to natural selection, better solutions are more likely to be selected for reproduction (crossover) to create new solutions (children).(Figure 3.2) Additionally, random mutations are introduced to maintain diversity. This process iterates until an optimal solution is found or a stopping criterion is met. GA offer several advantages over traditional

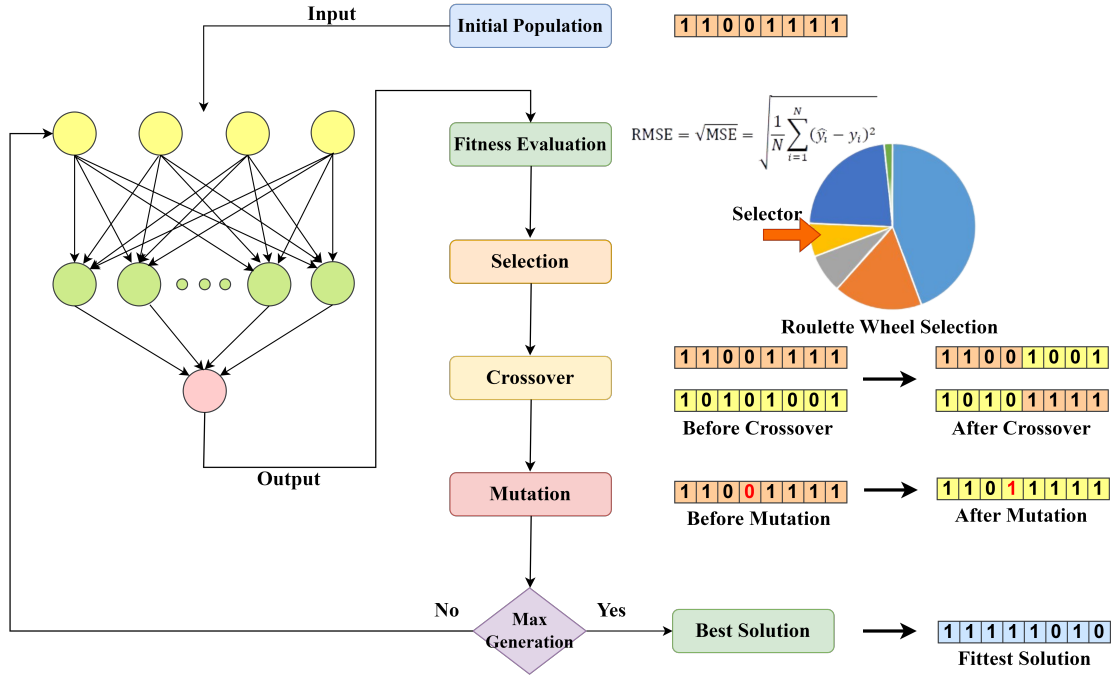


FIGURE 3.2: GA Workflow: Key Steps and Operations.

methods, including handling non-linear problems and working in parallel. However, carefully designing the fitness function and selection method is crucial for successful application.

3.8 ML Hybridization using GA and ANN

While ANN is a powerful tool, it sometimes needs help with slow learning and getting trapped in sub-optimal solutions. Researchers reported a hybrid approach to address these limitations by combining GA with ANN [204]. This method leverages the strengths of both techniques. The GA helps optimize the ANN performance by finding the best configuration (number of hidden layers, initial weights) and adjusting weights and biases to minimize the error function. This hybrid approach has proven to be effective in solving various engineering problems. Fig. 3.3 illustrates the workflow of the GA-ANN algorithm.

3.9 Proposed Framework for Intelligent Irrigation

The proposed framework is used to assist farmers in efficiently managing agricultural farms. This system integrates the IoT and ML functionalities to create an advanced intelligent irrigation system. The work is divided into four sections: data acquisition module, data gathering module, real-time monitoring using an Android application module, and an IoT-based motor controller module.

Unlike the ACO-ANN framework developed in Chapter 2 for rainfall prediction, this chapter utilizes a GA-ANN hybrid model tailored for irrigation management. The integration process

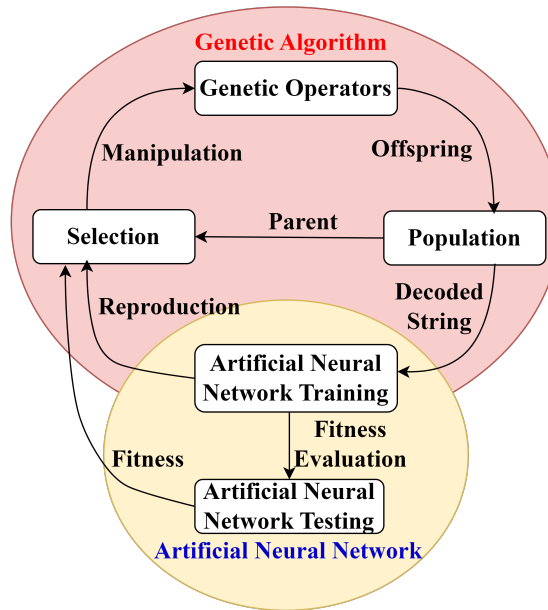


FIGURE 3.3: Integration of GA and ANN.

differs because GA is employed to optimize ANN parameters for evapotranspiration prediction. Therefore, the architecture of the proposed GA-ANN model is illustrated separately in Figure 3.4.

3.9.1 Data Acquisition Module

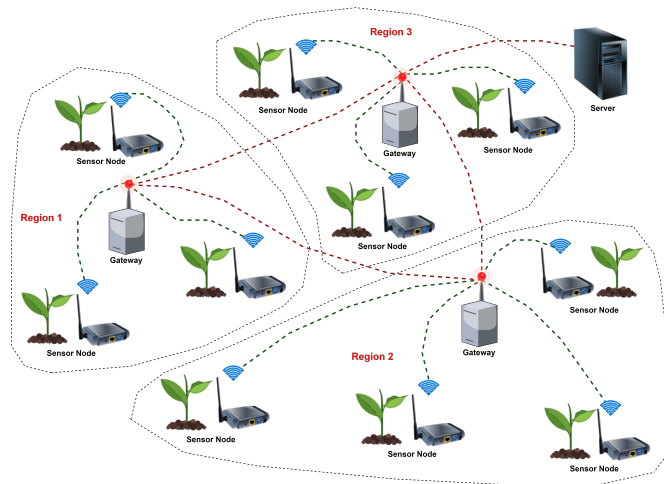


FIGURE 3.4: The Schematic Arrangement of Sensor Nodes in the Crop field.

Initially, agricultural area is divided into several sections. In each zone, sensor nodes are randomly placed to ensure complete coverage of the entire field. A sensor node comprises four sensors and a gateway node. Fig.3.4 illustrates a schematic arrangement of sensor nodes in the crop field. The field data consists of measurements obtained from four low-cost sensors and gateways, as shown in Table 3.2.

TABLE 3.2: Summary of various Sensors and Gateway Used

Sensor	Model Name	Purpose
Temperature	DHT22	Sensing Temperature.
Humidity	DHT22	Sensing Relative Humidity.
Pressure	BMP180	Sensing atmospheric pressure.
Sunshine Duration	BH1750	Measuring Sunshine Duration.
Raspberry Pi	Raspberry Pi V.4	Using as Gateway.

The sensors are linked to an ESP32 micro-controller. The ESP32 nodes transmit data to the Raspberry Pi 3 gateway node using lightweight communication protocol Message Queuing Telemetry Transport(MQTT). A gateway node corresponds to a specific region.

TABLE 3.3: Acronyms used in the Algorithm

Variable	Description
T_{mean}	Average Temperature
$S_{intensity}$	Radiation Intensity
$A_{pressure}$	Atmospheric pressure
T_{max}	Maximum temperature
T_{min}	Minimum temperature
L	Number of hidden layers
n_l	Number of neurons
f_l	Activation function
P	Population size
S	Selection operator
p_c	Crossover rate
p_m	Mutation rate
T_{max}	Maximum generations
W_{best}	Optimal weights
b_{best}	Optimal biases
N	Trained model
T	Tuple for data collected
C_{Humid}	Current Humidity
Th_{min_Humid}	Minimum Threshold
Th_{max_Humid}	Maximum Threshold
f_{target}	Fitness target

The ESP32 node serves as the publisher and transmits the data collected from the field. The Raspberry Pi node serves as the subscriber and receives all sensor data published by the ESP32 node in the corresponding region. For data interchange between the ESP32 node and gateway node, the Mosquitto broker service is utilized. The data gathered by gateway nodes in various regions are consolidated and transmitted to the AWS cloud server using MQTT protocol services. Algorithm 2 provides a comprehensive approach to the data acquisition model. Table 3.3 contains a list of all the acronyms used in the algorithm.

Algorithm 2 Data Acquisition Algorithm

```

1:  $Hrs \leftarrow \{1, 2, \dots, 23, 24\}$ 
2:  $SD \leftarrow \{time : \{Temp_{mean}, S_{intensity}, A_{pressure}, Temp_{max}, Temp_{min}\}$ 
3:  $| time \in Hrs\}$ 
4: Set  $v_2 = 1$ 
5: Set  $H = 1$ 
6: Set  $\epsilon = 0.63$ 
7: for  $time \in Hrs$  do
8:    $SD[time] \leftarrow \text{Collect\_Sensor\_Data}(time)$ 
9: end for
10:  $SD \leftarrow \text{Collect\_Sensor\_Data}()$ 
11:  $T \leftarrow \{\}$ 
12: for  $time \in Hrs$  do
13:   Calculate Psychometric Constant:
14:    $\Gamma = \frac{D_g A_{pressure}}{\epsilon \Lambda}$ 
15:   Calculate Slope Vapour Pressure:
16:    $\delta = \frac{4098 \left[ 0.610799 \exp \left( \frac{17.2699 \times Temp_{mean}}{Temp_{mean} + 237.299} \right) \right]}{(Temp_{mean} + 237.299)^2}$ 
17:   Calculate Mean Saturated Vapour Pressure:
18:    $f_s mean = \frac{f_s(Temp_{max}) + f_s(Temp_{min})}{2}$ 
19:   Calculate Actual Vapour Pressure:
20:    $f_a = f_s(Temp_{min}) \frac{Mh_{max}}{100} = 6.1099 \exp \left[ \frac{17.2699 \cdot Temp_{min}}{Temp_{min} + 237.299} \right]$ 
21:   Calculate Vapour Pressure Deficit (VPD):
22:    $VPD = f_s - f_a$ 
23:   Calculate Net Solar Radiation:
24:    $S_n = S_{ns} - S_{nl}$ 
25:    $S_{ns} = (1 - \beta) S_n$ 
26:
27:   Calculate Reference  $EV_{T0}$ :
28:    $EV_{T0} = \frac{(0.4079\delta(S_n - H)) + \Gamma \frac{900}{T + 273} v_2 (f_s - f_a)}{\delta + \Gamma(1 + 0.3399v_2)}$ 
29:
30:    $T[time] \leftarrow (SD[time][Temp_{mean}], SD[time][S_{intensity}],$ 
31:    $SD[time][A_{pressure}], SD[time][Temp_{max}], SD[time][Temp_{min}], EV_{T0})$ 
32: end for
33: Mosquitto_Gateway( $T$ )

```

3.9.2 EV_{T0} Prediction Module

Data organization is the initial stage of data pre-processing. The temperature and humidity of the experimental field are measured in real time to determine the highest and lowest temperatures and humidity of the day. Missing values are filled in with the KNN imputation of the corresponding months. There is a small amount of bias in replacing missing data with the imputation value. Then, the dataset is divided into training and test datasets. The training-to-test data set ratio is 80:20. To mitigate the presence of outliers, feature scaling process is carried out through standardization by utilizing the pre-processing technique of the MATLAB module.

The ANN utilizes a feed-forward back-propagation architecture to calculate the EV_{T0} rate based on maximum temperature ($Temp_{max}$), minimum temperature ($Temp_{min}$), maximum

Algorithm 3 Hybrid ANN-GA Algorithm For EV_{T0} Prediction

Require: $X, y, L, n_l, f_l, P, S, p_c, p_m, Temp_{max}, f_{target}$

- 1: $W_l \leftarrow \text{initialize_weights}(L, n_l)$
- 2: $b_l \leftarrow \text{initialize_biases}(L, n_l)$
- 3: $P \leftarrow \text{create_initial_population}(P, W_l, b_l)$
- 4: $t \leftarrow 0$
- 5: $f_{best} \leftarrow -\infty$
- 6: **while** $t < Temp_{max}$ and $f_{best} < f_{target}$ **do**
- 7: **for** $i = 1$ to P **do**
- 8: $W_i, b_i \leftarrow \text{decode_chromosome}(P[i])$
- 9: $a_l, L_i \leftarrow \text{Feed_Forward_Back_Propagation_ANN}(X, y, L, n_l, f_l, W_i, b_i)$
- 10: $f_i \leftarrow \text{RMSE}(y, a_l[-1])$
- 11: $P[i].\text{fitness} \leftarrow f_i$
- 12: **end for**
- 13: $P_s \leftarrow \text{Roulette_Wheel_Selection}(P)$
- 14: $O \leftarrow \text{Two_Point_Crossover}(P_s, p_c)$
- 15: $O' \leftarrow \text{Bit_Flip_Mutation}(O, p_m)$
- 16: $P \leftarrow O'$
- 17: $f_{best} \leftarrow \max_{P_i \in P} P_i.\text{fitness}$
- 18: $W_{best}, b_{best} \leftarrow \text{decode_chromosome}(\text{argmax}_{P_i \in P} P_i.\text{fitness})$
- 19: $t \leftarrow t + 1$
- 20: **end while**
- 21: **return** W_{best}, b_{best}

Algorithm 4 EV_{T0} Prediction and Irrigation Pump Control

- 1: $N \leftarrow \text{Train_Hybrid_ANN_GA_model}(T)$
- 2: **Set** $Th_{min_Humid} \leftarrow \text{Minimum Threshold Humidity}$
- 3: **Set** $Th_{max_Humid} \leftarrow \text{Maximum Threshold Humidity}$
- 4: **for** $k \in \text{time}$ **do**
- 5: $C_{Humid} \leftarrow N(T_k)$
- 6: **if** $C_{Humid} \leq Th_{min_soil_moist}$ **then**
- 7: **while** $(Th_{max_Humid} - Th_{min_Humid})/2 > C_{Humid}$ **do**
- 8: $\text{Start_Irrigation_Pump}()$
- 9: **end while**
- 10: **else**
- 11: **while** $Th_{max_Humid} \geq C_{Humid}$ **do**
- 12: $\text{Start_Irrigation_Pump}()$
- 13: **end while**
- 14: **end if**
- 15: $\text{Stop_Irrigation_Pump}()$
- 16: **end for**

humidity ($Humid_{max}$), minimum humidity ($Humid_{min}$), atmospheric pressure(Q) and solar radiation(S_n). The number of neurons in an ANN is determined by the number of input variables, with one additional neuron in the output layer. To avoid convergence problems, including slow convergence and getting trapped at the local minimum, a hybrid approach combining GA and ANN is proposed here to enhance the performance of ANNs and achieve the global minimum (shown in Algorithm 3). This method involves determining and adjusting critical parameters of

an ANN, such as the optimal number of hidden layers and the initial weights, by considering a suitable fitness function. The primary objective of employing the hybrid GA–ANN model is to optimize a set of weights and biases to minimize the RMSE. The adaptability of the proposed model is not limited to coriander plants. The hybrid ANN-GA model can be retrained with region-specific data, and by modifying the crop coefficient (C_c), the proposed model can address the water requirements of various crops and environmental conditions.

3.9.3 Real-Time Monitoring using Android Application Module

A real-time monitoring application based on the Android operating system is created using Flutter. The Android application displays sensor data obtained from the experimental field. The application can autonomously initiate or terminate irrigation operations to obtain the specified moisture level necessary for plant development.

3.9.4 IoT-based Irrigation Pump Controller Module

By utilizing a Wi-Fi-enabled node and the response from an Android application, it is possible to activate or deactivate the irrigation pump by manipulating the relay switch. Furthermore, the server can autonomously access the irrigation pump. Algorithm 4 outlines the process through which the water motor is controlled.

The proposed irrigation controller consists of sensor nodes, communication infrastructure, cloud services, gateway devices, and an actuator subsystem. Environmental parameters are collected using field-deployed sensors and transmitted through ESP32-based sensing nodes. The collected data are forwarded via Wi-Fi to the cloud server using the MQTT protocol. The ANN-GA prediction model estimates evapotranspiration and crop water requirements. Based on the predicted irrigation demand, control commands are transmitted to a Raspberry Pi gateway, which activates or deactivates the irrigation pump through a relay module. Farmers can also monitor system status and manually override pump operation through the Android application.

3.10 Data Collection

TABLE 3.4: Dataset of Vegetative Growth Cycle[August, 2023]

Date	f_s	f_a	δ	Γ	S_n	v_2	Temp	EV_{T0}
01/08/23	63.97	48.77	0.204	0.066	1.170	0.293	26.61	3.20
02/08/23	64.10	48.86	0.204	0.066	1.170	0.331	26.26	3.72
03/08/23	64.12	48.86	0.204	0.066	1.170	0.293	26.26	3.46
.	.	.	.	.	.	.	.	.
29/08/23	64.25	48.98	0.204	0.066	1.170	0.352	25.85	4.12
30/08/23	64.49	49.16	0.204	0.066	1.170	0.355	25.23	3.99

TABLE 3.5: Dataset of Bolting Cycle[September,2023]

Date	f_s	f_a	δ	Γ	S_n	v_2	Temp	EV_{T0}
01/09/23	63.91	48.72	0.204	0.067	1.253	0.39	26.77	4.26
02/09/23	63.95	48.75	0.204	0.068	1.253	0.42	26.66	4.78
03/09/23	64.03	48.81	0.204	0.067	1.254	0.38	26.44	4.36
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
29/09/23	64.20	48.94	0.204	0.067	0.253	0.44	25.99	5.01
30/09/23	64.39	49.08	0.204	0.067	0.255	0.42	25.47	4.77

TABLE 3.6: Dataset of Flowering and Seed Development Cycle[October, 2023]

Date	f_s	f_a	δ	Γ	S_n	v_2	Temp	EV_{T0}
01/10/23	64.56	48.36	0.204	0.067	1.154	0.23	25.69	4.23
02/10/23	64.25	49.56	0.204	0.066	1.154	0.25	26.79	3.26
03/10/23	65.35	50.32	0.204	0.066	1.158	0.36	22.26	3.59
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
29/10/23	65.82	50.25	0.204	0.068	1.155	0.32	20.32	2.33
30/10/23	65.83	50.65	0.204	0.068	1.163	0.26	20.22	3.21

This work examines crop irrigation needs using empirical data collected from a private home garden situated in Brajanathchak, Haldia, West Bengal, India (latitude 22.037400°N, longitude 88.053528°E). The garden has three sensor nodes that collect data every 10 minutes from August 1st to October 30, 2023. This work employed a three-month-long dataset consisting of snapshots captured during the vegetative growth cycle (Table 3.4), bolting cycle (Table 3.5), and flowering and seed development cycle (Table 3.6).

3.10.1 Experimental Testbed and System Configuration

The experimental testbed was deployed in a coriander cultivation field at Haldia, West Bengal, India. The system consisted of three distributed sensor nodes equipped with environmental sensors, an ESP32 microcontroller for data acquisition, a Raspberry Pi gateway for local processing, and a relay-controlled irrigation pump. Wireless communication was established through Wi-Fi using the MQTT protocol. The irrigation decision engine was implemented using the proposed GA-ANN model running on a cloud-connected server. The Android application provided real-time monitoring and remote pump control. The complete configuration enabled end-to-end automated irrigation management under real field conditions. Table 3.7 summarizes the hardware and software configuration employed for the implementation of the proposed IoT-based smart irrigation system.

TABLE 3.7: Hardware and Software Configuration of the Smart Irrigation Testbed

Component	Specification	Purpose
ESP32	Dual-core MCU	Sensor acquisition
BH1750	Light sensor	Solar radiation estimation
DHT22	Temperature & Humidity	Weather monitoring
Raspberry Pi 4	Gateway	Local processing
Relay Module	5V relay	Pump switching
Water Pump	AC Motor Pump	Irrigation
MQTT Broker	Communication	Data transfer
Android App	Monitoring Interface	User control

3.11 Experimentation

Depending upon the selection, mutation, and crossover operators of GA, nine HML configurations are evaluated by testing different combinations of: HML1 (utilizing roulette wheel selection, single-point crossover, and bit flip mutation), HML2 (employing roulette wheel selection, two-point crossover, and bit flip mutation), HML3 (using roulette wheel selection, uniform crossover, and bit flip mutation), HML4 (utilizing ranking selection, single point crossover, and bit flip mutation), HML5 (employing ranking selection, two point crossover, and bit flip mutation), HML6 (using ranking selection, uniform crossover, and bit flip mutation), HML7 (utilizing tournament selection, single point crossover, and bit flip mutation), HML8 (employing tournament selection, two point crossover, and bit flip mutation) and HML9 (using tournament selection, uniform crossover, and bit flip mutation). The parameters for the GA are determined through iterative testing and are presented in Table 3.8. The population size, selection rate, mutation rate, and number of rounds are fine-tuned through rigorous testing to obtain optimal results.

TABLE 3.8: GA Variables and their Values

Variables	Values
Number of Iterations(U)	200
Population Size(m)	100
Selection Rate(d)	0.7
Crossover Rate(r)	0.8
Mutation Rate(n)	0.02
Solution Space(E)	5

Choosing the optimal design in feed-forward back-propagation ANN is crucial to ensure accurate and efficient system operation. To achieve an optimal design, a solution space is defined in four dimensions. Each individual in the GA population corresponds to a unique ANN setup. The gene structure determines the parameters of the network design and training process. Optimal gene structure is a key factor for providing better results. Equations 3.11 and 3.12 provide the population size and constraint function for the number of hidden layers (NHL) and neurons per layer (NPL) to ensure diversity and practicality in the generated structures.

$$Q_{lNHL_j}(y) = \begin{cases} 1, & y < 1 \\ y, & 1 \leq y \leq 30 \\ 30, & y > 30 \end{cases} \quad (3.11)$$

$$Q_{lNPL_j}(z) = \begin{cases} 1, & z < 1 \\ z, & 1 \leq z \leq 20 \\ 20, & z > 20 \end{cases} \quad (3.12)$$

The constraint functions, $Q_{lNHL_j}(y)$ and $Q_{lNPL_j}(z)$, as specified by equations 3.11 and 3.12, are utilized in the GA population to control the allowable values for NHL and NPL . The functions ensure that NHL ranges from 1 to 30 and NPL ranges from 1 to 20 for each iteration and gene. An ANN randomly selects the activation function (ACF)(Table 3.11) and training function (Tfc)(Table 3.11) for its setup and training.

$$F(y_j) = RMSE(ANN_j) \quad (3.13)$$

The fitness of each ANN configuration (gene) in the GA is evaluated using Equation 3.13. In this context, $F(y_j)$ denotes the fitness score of gene j , and $RMSE$ of its associated ANN (ANN_j). The lower $RMSE$ value indicates the superior performance of ANN (ANN_j).

According to earlier study [205], a learning rate of 0.2 has been found effective, therefore the same has been adopted in this work. The $RMSE$ is employed to evaluate the efficiency of networks produced by each gene. The $RMSE$ values are also used to evaluate fitness in the GA for selecting parents in the following generation.

The constraint functions derived from equations 3.11 and 3.12 guarantee that mutations are confined to the specified range for the number of hidden layers and neurons per layer. This prevents the formation of network architectures that are invalid or ineffective.

3.12 Performance Evaluation

The efficiency of predictions is a vital aspect of ML for measuring the efficacy of a model. To accomplish this, the MSE, RMSE, MAE, and R^2 are used as evaluation metrics [206].

Table 3.11 demonstrates that the HML4 model is the most dominant due to its lowest RMSE value. Every gene listed in Table 3.11 corresponds to an artificially created ANN that has been trained to forecast the water requirements of the plant. The water consumption quantities projected by the most influential genes in each model serve as the basis for developing a customized irrigation timetable that caters to various stages of life cycle of the plant. Fig. 3.5 presents visually appealing representations of EV_{T0} prediction for each life cycle. HML4 is the most accurate model for determining the amount of water for coriander, covering all growth stages, such as vegetative growth, bolting, blooming, and seed development. The HML4 predictions

TABLE 3.9: Results for Vegetative Growth, Bolting, Flowering and Seed Development Cycle

Vegetative Growth Cycle									
Indicator	HML1	HML2	HML3	HML4	HML5	HML6	HML7	HML8	HML9
<i>RMSE</i>	0.172	0.064	0.048	0.030	0.087	0.052	0.094	0.122	0.032
<i>MSE</i>	0.029	0.004	0.002	0.001	0.007	0.002	0.008	0.015	0.001
<i>R²</i>	0.764	0.967	0.981	0.992	0.939	0.978	0.929	0.880	0.991
<i>MAE</i>	0.171	0.024	0.023	0.020	0.019	0.036	0.049	0.064	0.026
Bolting Cycle									
<i>RMSE</i>	0.037	0.146	0.184	0.034	0.075	0.038	0.051	0.162	0.086
<i>MSE</i>	0.001	0.021	0.033	0.001	0.005	0.001	0.002	0.026	0.007
<i>R²</i>	0.987	0.814	0.707	0.989	0.950	0.987	0.977	0.771	0.935
<i>MAE</i>	0.020	0.089	0.025	0.020	0.016	0.029	0.022	0.161	0.042
Flowering and Seed Development Cycle									
<i>RMSE</i>	0.127	0.199	0.166	0.028	0.044	0.149	0.041	0.087	0.104
<i>MSE</i>	0.016	0.039	0.027	0.001	0.002	0.022	0.001	0.007	0.010
<i>R²</i>	0.920	0.805	0.864	0.996	0.990	0.890	0.991	0.962	0.946
<i>MAE</i>	0.123	0.072	0.102	0.020	0.017	0.080	0.029	0.058	0.054

had average errors of 0.092, 0.085, and 0.042 for the three growth stages Table 3.9, while the average errors for the other models varied from 0.12-0.29. HML4 is a more dependable model for forecasting the water requirements of coriander plants.

Fig. 3.6(a), (b), (d), and (e) illustrate the correlation between the forecasted EV_{T0} (output from the network) and observed EV_{T0} (measured) for the training, validation, test, and overall dataset. All four regression graphs have a highly significant correlation coefficient (r) that surpasses 0.75. This suggests that the ANN model fully understands the data. Moreover, the dataset's overall gradient of 0.77 indicates that for each incremental rise in the measured EV_{T0} , the projected EV_{T0} increases by an average of 0.78 units. Furthermore, the fixed value of 0.22 signifies the estimated EV_{T0} when the observed value is zero. For instance, if the measured EV_{T0} is 0.4, it corresponds to a predicted EV_{T0} of 0.324 based on the regression equation.

Fig. 3.6(c) provides a more detailed analysis of the performance of the ANN model by examining the distribution of errors. The error histogram indicates that most predictions (≈ 1600) exhibit a high degree of accuracy, with near zero errors. Nevertheless, the histogram also suggests the presence of potential outliers in the data. Although most errors are typically within a range of -0.4051 to 1.041, there are instances where errors exceed these limits. These data points suggest outliers that pose challenges to the model.

Fig. 3.6(f) illustrates the progressive learning of the ANN model. MSE measures the errors between the predicted and actual values for the training, validation, and testing data. The crucial aspect is the validation error, which indicates the model's ability to generalize unfamiliar data. The value of 0.16 is achieved at time 11, indicating successful learning. There is a small rise in inaccuracy over the next 11 epochs. To mitigate over-fitting, the training process is terminated

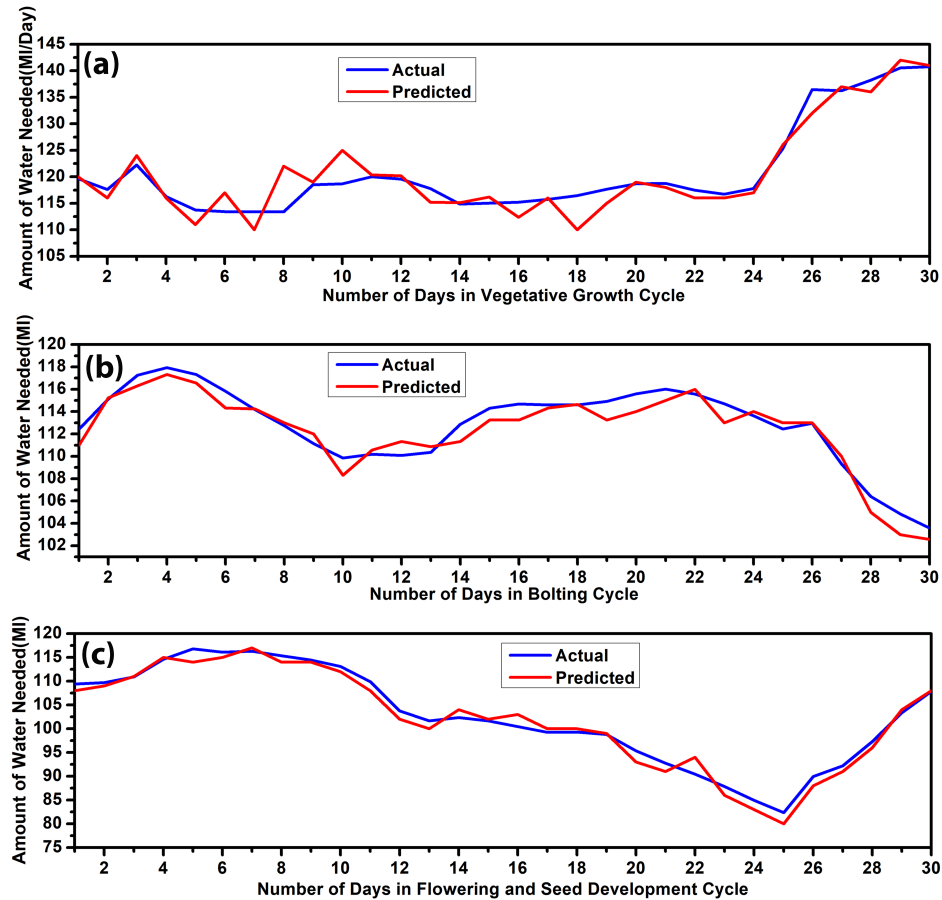


FIGURE 3.5: (a) Actual vs HML4 Predicted Water Requirement for Vegetative Growth Cycle, (b) Actual vs HML4 Predicted Water Requirement for Bolting Cycle, (c) Actual vs HML4 Predicted Water Requirement for Flowering and Seed development Cycle

at epoch 31. Overall, the final MSE for all the data points is relatively low, indicating that the model successfully captures the correlation between the input and output data.

3.13 Results Comparison

TABLE 3.10: Results Comparisons with Existing Work

Work(Ref.)	MSE	RMSE	MAE	R ²
Bashir et al.[185]	-	0.86	0.75	0.94
Mostafa et al.[186]	-	0.39 - 0.74	0.38 - 0.58	0.91 - 0.95
Yildirim et. al.[187]	-	0.327	0.232	0.961
Tr et al.[188]	-	0.016	0.011	-
Kocian et al.[189]	-	0.108	-	0.84 - 0.94
Dong et al.[190]	-	0.11 - 0.28	0.15 - 0.26	0.64 - 0.82
Bidabadi et al.[191]	0.22 - 0.29	0.47 - 0.54	-	0.83 - 0.96
Rajput et al.[192]	0.053	0.230	0.126	0.982
Paria et al.(Proposed) [172]	0.02	0.16 - 0.32	0.020	0.98 - 0.99

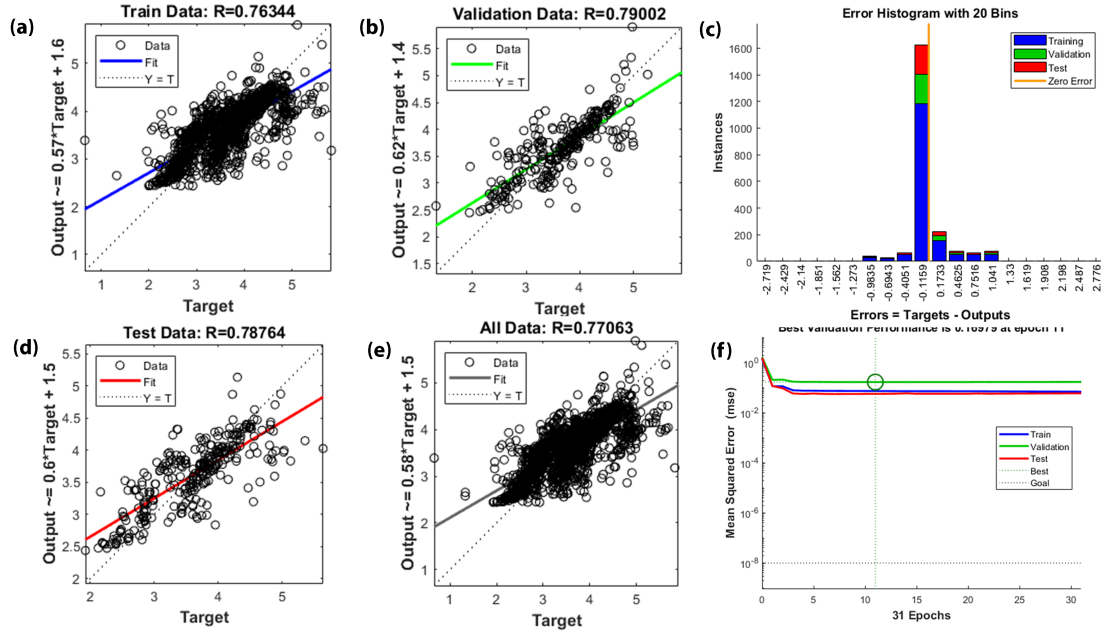


FIGURE 3.6: (a) Correlation between the forecasted EV_{T0} and the observed EV_{T0} for training, (b) Correlation between the forecasted EV_{T0} and the observed EV_{T0} for validation, (c) Error Histogram, (d) Correlation between the forecasted EV_{T0} and the observed EV_{T0} for testing, (e) Correlation between the forecasted EV_{T0} and the observed EV_{T0} for all data, (f) MSE between the predicted and actual values for the training, validation, and testing data

The proposed approach achieves superior performance compared to existing methods for EV_{T0} estimation (shown in Table 3.10). Notably, this model boasts the highest R^2 values (0.98-0.99), which indicates a very strong correlation between predicted and observed EV_{T0} values. Although Tr et al. [188] achieved a lower RMSE (0.016), their results may not be directly comparable due to unavailability of dataset. Compared with other studies the proposed approach demonstrates significantly lower and more consistent RMSE and MAE values across a wider range, which suggests greater generalizability and accuracy. This highlights the effectiveness of this method in precisely estimating EV_{T0} for various agricultural applications.

3.14 Real-time Smart Irrigation System

Fig. 3.7 illustrates the comprehensive structure of a smart irrigation system, developed to monitor and notify farmers regarding the water requirements of plants using up-to-the-minute crop field data. Sensors collect vital data such as temperature, humidity, atmospheric pressure, and duration of sunshine.

The sensor data are transmitted to an attached ESP32 micro-controller and then aggregated from several nodes through a Raspberry Pi gateway before reaching the AWS cloud server. The most successful model, HML4 on the cloud server, is utilized to process the input data and estimate the water requirements of the plant. This process initiates the regulation of the attached irrigation pump to supply the required water for irrigation purposes.

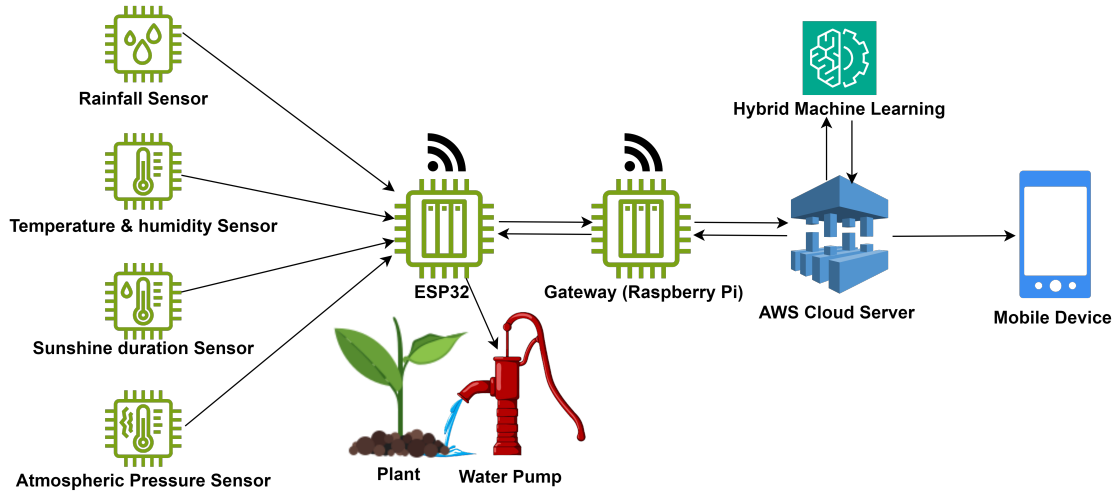


FIGURE 3.7: Experimentation block diagram

Fig. 3.8 presents the real-time practical application of the intelligent irrigation system. This ESP32 micro-controller is immediately linked to the sensors positioned near the plant's roots and gets the actual data values from the crop field. The AWS Cloud Server utilizes the hybrid ML model HML4, which is established previously as illustrated in Fig. 3.7. The HML4 at the AWS Cloud server activates the relay linked to the irrigation pump to enable the required amount of water via the irrigation pump (which has a pumping capacity of 19.2 ml per second [207]). Each movement is meticulously calculated and fine-tuned to guarantee that the plant receives the optimal moisture.

However, the system does not end its functionality here. An exclusive Android application provides real-time updates at each stage of the process. Fig.3.9 displays the app interface, which is easy to handle for the farmers. This allows farmers to track sensor readings and the overall operation of the system. The innovative irrigation system continuously provides up-to-date information about crop fields, helping farmers to make informed decisions regarding the maintenance of plants. Fig. 3.9 displays the user-friendly Android application created for the smart watering system. The interface is divided into several Display Panels [Fig. 3.9(a) - (f)]. This interface provides farmers with an overview of the plant's condition. The sensor data, including temperature, humidity, Atmospheric Pressure and net radiation, are shown in Fig. 3.9(b) - (f). The regularly updated data, with a refresh rate of 10 minutes, enables farmers to remain well informed about the requirements of their plants. The data screens, shown in Fig. 3.9, offer historical insight. Farmers can retrieve daily water usage data, enabling them to monitor plant water consumption over time. In addition, the system logs the cumulative water supply from the previous irrigation cycle, as well as the average temperature and solar radiation, providing a comprehensive overview of the plant's surroundings and water requirements.

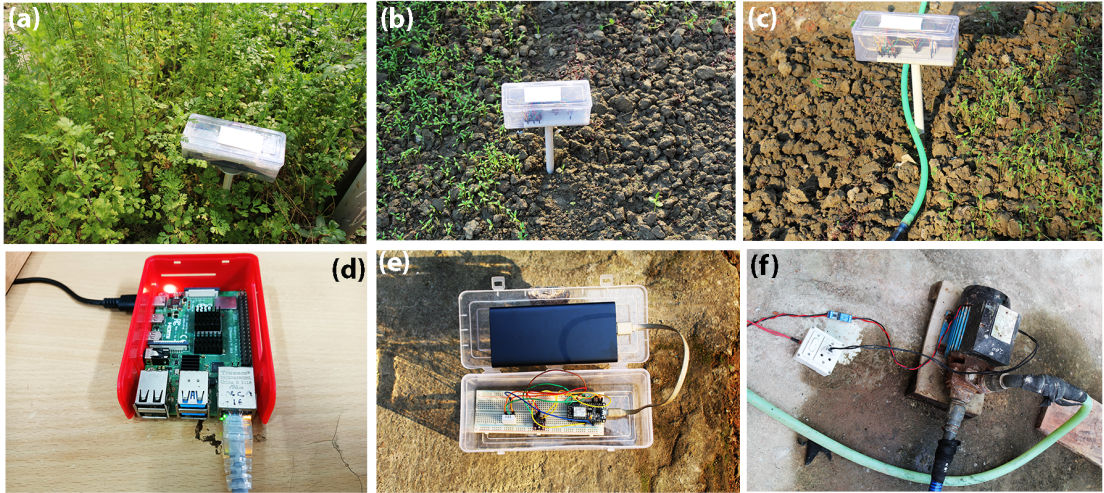


FIGURE 3.8: (a) Sensor node at Region 1, (b) Sensor node at Region 2, (c) Sensor node at Region 3, (d) Gateway(Raspberry Pi), (e) In-Depth view of each Sensor Node, (f) Irrigation Pump with power supply and relay module

3.15 Chapter Summary

As global warming worsens and groundwater decreases, controlling groundwater for agriculture is essential. Intelligent watering systems are used to get better results with less water wastage. The primary goal of this work is to reduce over-irrigation of coriander plants. A novel approach is developed to estimate net solar radiation using sunshine duration data from the BH1750 sensor. Nine HML models are developed. Among them, HML4 showed the best performance for prediction of EV_{T0} . The exact water requirements for each growth stage are determined by integrating the predicted EV_{T0} with C_c values. Additionally, a custom Android application is successfully developed for remote monitoring and management, demonstrating the practical applicability of the intelligent irrigation system in coriander fields. This work shows significant advancements in precision agriculture, offering a promising solution for efficient water management in crop production.

This intelligent watering system extends beyond its original use in coriander. Its versatility and adaptability make it appropriate for a wide range of crop plants globally in different geographical and climatic situations. This presents tremendous opportunities for advancing environmental sustainability, water conservation, and efficient resource utilization in agriculture. The system provided vital knowledge for optimizing irrigation infrastructure. Integrating various sensor data with groundwater flow analysis into real-time smart irrigation systems could provide considerable outcomes for future developments and might improve irrigation techniques and optimize water control in agriculture. There are many more possibilities of future research in the field for smart irrigation, which are listed below:

- Future research will aim to enhance the accuracy and efficiency of the proposed hybrid models using advanced metaheuristic techniques such as Particle Swarm Optimization (PSO) or Differential Evolution to refine the model's performance.

- Additional training with more extensive and diverse datasets will be conducted to improve generalizability of the models to different crops and climatic conditions.
- The IoT system can be expanded to support a larger agricultural field by implementing robust multi-node sensor networks and integrating advanced communication protocols such as LoRaWAN for efficient long-range data transmission [208].
- Incorporation of explainable AI (XAI) techniques in the proposed work can enhance the interpretability of model outcomes. This may help farmers and stakeholders to better understand the key factors influencing irrigation recommendations.
- Upgradation of Android applications with predictive analytics and automated decision-making features may further improve the efficiency of irrigation management.

After establishing efficient water management systems, the next important challenge in sustainable agriculture is boosting soil fertility. The upcoming chapter deals with the complexities of soil management, specifically focusing on how AI-driven models classify soil fertility. By using tools like ACO-XgBoost and XAI, farmers will gain a deeper understanding of the nutrient composition of their soil. The AI-driven models will not only consider the macro and micro-nutrient levels within the soil but also provide a clearer understanding of its overall health. With this knowledge, farmers can select the most suitable crops for their specific soil conditions. They will also be able to adopt more effective fertilization practices, leading to improved crop yields and sustainable agricultural outcomes. The next chapter will further explore the role of AI in transforming traditional soil management practices and making them more accessible, accurate, and actionable.



FIGURE 3.9: (a) Home Screen, (b) Graph for Temperature, (c) Graph for Atmospheric Pressure, (d) Graph for Specific Humidity, (e) Graph for Relative Humidity, (f) Graph for Net Radiation, (g) Screen for Pump on, (h) Screen for Pump Off

TABLE 3.11: Models & their Best Genes

Models	Column ¹	Column ²	Column ³	Column ⁴	Column ⁵	Column ⁶	Column ⁷	Column ⁸
HML1	4	5	7 – 12 – 7 – 6 – 4	Logsig, hardlim, tansig, hardlims, purelin	1	purelin	traincgp	0.194
HML2	4	4	6 – 13 – 6 – 3	hardlim, radbas, tansig, purelin	1	purelin	traincgp	0.632
HML3	4	4	7 – 11 – 9 – 5	Logsig, hardlim, hardlim, tansig	1	tansig	traincgp	0.812
HML4	4	5	10 – 14 – 8 – 6 – 5	tansig, radbas, tansig, logsig , purelin	1	purelin	trainbfg	0.165
HML5	4	3	6 – 9 – 3	radbas , radbas, tansig	1	tansig	trainbfg	0.923
HML6	4	5	9 – 12 – 8 – 6 – 4	Logsig, logsig, hardlim, tansig, purelin	1	purelin	traincgp	0.215
HML7	4	4	7 – 10 – 8 – 2	Logsig, hardlim, tansig, tansig	1	tansig	trainbfg	0.546
HML8	4	5	12 – 15 – 11 – 9 – 6	Logsig, radbas, tansig, hardlim, purelin	1	purelin	trainbfg	0.248
HML9	4	3	8 – 11 – 5	Logsig, radbas, purelin	1	purelin	traincgp	0.844

^aInput Count^bHidden Layer Count.^cNeuron Count in Hidden Layers.^dActivation Functions Utilized in Hidden Layers.^eQuantity of Neurons in Output Layer^fActivation Function utilized in Output Layer^gTraining Function^hRMSE

Chapter 4

The Soil Beneath – Classifying Soil Fertility for Better Crop Choices

In India, agriculture is the main means of earning a living; about 70% of rural families rely on farming as their major income source. India is the largest producer and consumer of agricultural products, accounting for 25% of global production and 27% of global consumption. Understanding the soil's chemical, physical, and biological properties is vital for the productivity of agriculture. It is crucial to increase soil fertility according to its chemical properties in order to increase crop yield. However, soil deterioration through erosion, compaction, nutrient depletion, and contamination, causes a reduction in soil fertility. Crops grown in nutrient-deficient or nutrient-excess soils can experience diseases that lead to poor productivity. Therefore, accurately assessing soil parameters is vital for improving yields and ensuring sustainable farming practices.

To maintain soil quality necessary nutrients for the proper growth of the crops is required. This will ensure fertility of the soil. Desired plant growth is obtained by delivering adequate nutrients, such as NPK to the plants. Soil fertility directly influences both the amount and the quality of agricultural produce. Healthy soil helps to produce nutritious crops with higher yields. On the other hand, low-fertility soil leads to low yields and poor crop quality. Additionally, excessive use of fertilizers can harm the environment.

Soil fertility is influenced by several factors, including acidity level, organic matter, and cation exchange capacity (CEC), nutrient proportions, as well as the soil's textural composition. Soils with low pH have poor nutrient retention, leading to reduced fertility. The essential soil chemical parameters for plant growth are categorized into non-mineral and mineral nutrients. Non-mineral nutrients, such as carbon, oxygen and hydrogen are derived from both air and water. Mineral nutrients are divided into macronutrients (including N, K, P, Ca, S, and Mg) and micronutrients (such as B, Mn, Cl, Fe, Cu, Mo, Zn, and Ni). They play critical roles in various growth

stages of crops, such as seed germination, root and leaf development, photosynthesis, and disease resistance. Imbalances of soil chemical parameters can adversely affect the crop growth and yield.

In the previous chapter, the focus was on water management and its importance for sustainable farming. While water is essential for crop growth, well-organized soil management is crucial to enhance agricultural productivity. But proper soil management is a key challenge in the present day. Farmers must gain an understanding of soil health so they can take well-informed decisions about nutrient application and overall soil management. This chapter focuses on soil fertility, presenting it as the next major challenge in sustainable agriculture.

To address this challenge, AI plays an increasingly important role. AI, particularly ML models, has revolutionized the classification of soil fertility. However, individual models may not be sufficient to capture all the necessary parameters or features required for accurate predictions. While ML techniques have shown effectiveness in various domains, their ability to handle the complexity of soil fertility and considering factors such as micro and macro-nutrients, soil texture, and environmental conditions, can be limited when used in a single model. To overcome these limitations, hybrid ML models have been proposed. Such hybrid approaches integrate the advantages of several algorithms—like pairing ACO with XGBoost to enhance predictive accuracy by boosting the efficiency of each individual model.

Moreover, hybrid models enhance the ability to identify the varied and nonlinear patterns naturally present in soil-related data. This integration is crucial for achieving the level of precision and efficiency required for classifying soil fertility. While traditional soil testing methods are time-consuming, expensive, labor-intensive, and require lab work, AI can automate this process, delivering faster and more efficient outcomes. Explainable AI further enhances the model's utility by helping farmers understand how these models make decisions, providing transparency, and fostering trust in the system. The ability to clarify how decisions are derived makes AI solutions more actionable and accessible for farmers, ultimately improving soil management practices and helping farmers make better-informed choices related to soil management and nutrient application.

The literature review highlights various advancements in soil fertility classification using ML and deep learning techniques. However common limitation across these studies is that proper explanation regarding the results obtained was not satisfactory. Thus, absence of XAI does not always make these proposals suitable. While many of the works propose methods that successfully classify soil fertility based on chemical and physical soil parameters, none of them tries to identify and explain the key factors influencing the classification process, which is crucial for making the results more interpretable and actionable for farmers. Furthermore, a variety of algorithms, including CNN, RF and NB, have been used to enhance classification accuracy; but, there is no use of hybrid models combining these techniques with XAI approaches to not only improve accuracy but also offer transparency. A promising approach seems to be the development of a new hybrid algorithm that incorporates XAI, which can provide both high performance in soil fertility classification and a clear understanding of the critical factors, ultimately supporting an efficient decision-making in precision agriculture.

The main contributions of this work are as follows:

1. A hybrid machine learning model is introduced where the hyperparameters of the XGBoost algorithm are fine-tuned using the ACO technique to achieve better classification performance.
2. The hybrid model ACO-XgBoost gives a high accuracy (98%) compared to the traditional ML models.
3. To provide better understanding, the best-performing model (ACO-XgBoost) is investigated using XAI techniques to highlight the most important features contributing to soil fertility classification.

4.1 Reasons for Soil Fertility Classification

Soil fertility classification is important for precision agriculture. It helps in optimizing resources, increasing crop yields, and supports sustainable farming. By assessing soil nutrient levels, farmers can apply fertilizers as per requirement, leading to both cost savings and the elimination of wasteful excess. The approach improves productivity and protects the environment by preventing over-fertilization. In the long run, it helps maintain soil health and ensures successful farming.

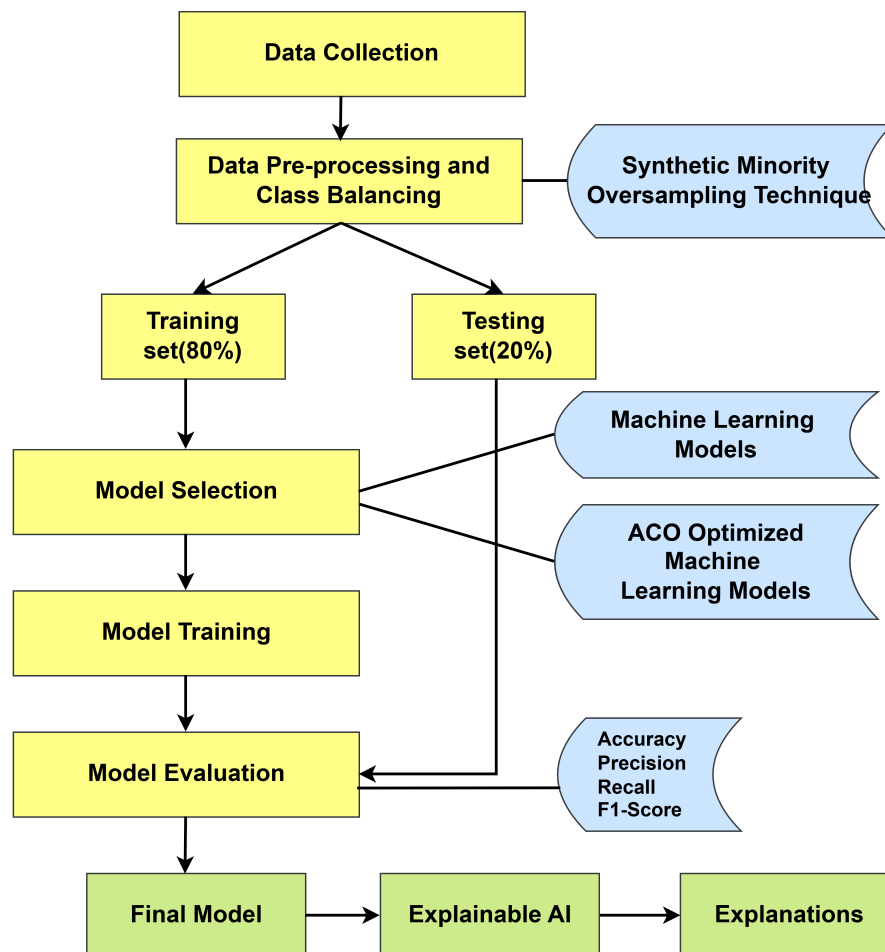


FIGURE 4.1: Workflow for Soil Fertility Classification

4.2 Methodology

Fig. 4.1 represents the workflow for the system used for categorizing soil fertility levels. The procedure starts by gathering the dataset from the Kaggle data source [214]. The dataset divides the soil into three fertility classes. Class 0 (Low Fertility) soil has fewer vital nutrients, meaning it will need a lot of fertilizer to grow healthy plants. Class 1 (Medium Fertility), which includes soils with moderate nutrient levels; these can support good crops with some careful fertilization and management. Finally, Class 2 (High Fertility) represents nutrient-rich soils that are ideal for farming, as they may lead to optimal crop growth with little need for added fertilizers. Once the data is collected, it undergoes several data preprocessing steps, such as cleaning the dataset by removing irrelevant columns, handling missing values, and applying normalization to normalize the feature values so they fall within a uniform scale. This step also balances the dataset classes using SMOTE, making equal distribution for the classes. After balancing, each class contains exactly 440 data points. The training and testing split for each class are as follows: for class 0, the ratio is 356:84; for class 1, it is 347:93; and for class 2, it is 353:87. Model training is then performed using traditional and ACO optimized ML algorithms. After training, each model is tested using evaluation indicators including accuracy, precision, recall, and the F1-measure. These evaluation measures help assess how effectively the model performs the classification task. Once the final model is selected, an explanation of how the model makes decisions is provided, which helps in understanding the predictions made by the model and ensures trust and reliability in its use.

4.2.1 Dataset Description

The soil fertility dataset includes multiple soil attributes—such as N, P, K, pH, EC, OC, S, Zn, Fe, Cu, Mn, and B along with a target label that indicates the fertility category (class 0, class 1, or class 2). [214]

4.2.2 Machine Learning Algorithm

In this work, various A range of machine learning techniques LR, DT, RF, SVM, NB, KNN, Gradient Descent, and XGBoost—are employed to train the classification model. Each of these models are evaluated based on their performance. To further enhance its performance, the The machine learning model’s hyperparameters are fine-tuned using the ACO optimization technique. This hybrid approach, combining the strengths of both ML and ACO (proposed hybrid model shown in Algorithm 5), led to improved classification accuracy and efficiency, achieving better results compared to other models used in this work.

4.2.2.1 SHapley Additive exPlanations (SHAP)

The core concept of SHAP [215] revolves around the Shapley Value, that represents the the mean incremental influence a feature has on the model’s prediction when evaluated across every

Algorithm 5 Proposed Hybrid XgBoost with ACO for Soil Fertility Classification

```

1: Input: Soil properties  $x_i \in \mathbb{R}^{12}$  (N, P, K, pH, EC, OC, S, Zn, Fe, Cu, Mn, B) to
   predict soil fertility class  $y \in \{0, 1, 2\}$ .
2: Output: Optimized hyper-parameters  $\mathbf{h}^*$  and trained XgBoost model.
3: Initialize pheromone levels  $\tau_{ij}$ , number of ants  $m$ , iterations  $N_{\text{iter}}$ , evaporation rate
    $\rho$ , and influences  $\alpha, \beta$ .
4: for each iteration  $t = 1 \dots N_{\text{iter}}$  do
5:   for each ant  $A_k, k = 1 \dots m$  do
6:     Step 1: Construct Solution
7:     for each hyper-parameter  $h_i \in \mathbf{h}$  do
8:       Select  $h_j$  based on  $P(h_j|h_i) = \frac{\tau_{ij}^\alpha \cdot \eta_{ij}^\beta}{\sum_{j'} \tau_{ij'}^\alpha \cdot \eta_{ij'}^\beta}$ 
9:     end for
10:    Step 2: Evaluate Solution
11:    Train XgBoost with  $\mathbf{h}_k$ , compute validation error  $\mathcal{L}_k$ 
12:    Update  $\mathbf{h}_{\text{best}}$  if  $\mathcal{L}_k < \mathcal{L}_{\text{best}}$ 
13:  end for
14:  Step 3: Update Pheromone
15:  for each hyperparameter pair  $(i, j)$  do
16:    Update  $\tau_{ij} \leftarrow (1 - \rho) \cdot \tau_{ij} + \frac{1}{\mathcal{L}_k}$ 
17:  end for
18: end for
19: Step 4: Return Best Hyper-parameters  $\mathbf{h}^*$ 
20: Step 5: Retrain Final Model with  $\mathbf{h}^*$ 
21: Step 6: Classify Fertility  $\hat{y}_i = \text{XgBoost Classifier}(x_i|\mathbf{h}^*)$ 

```

possible subset of features. SHAP values are stored as tensors that quantify how much each feature contributes to the model's output.

For total contribution in feature distribution i , its Shapley value is computed as the mean of its marginal contributions across every feasible subset of features $S \subseteq N \setminus \{i\}$. Eq. 4.1 defines how feature i contributes marginally to a specific subset S .

$$\Delta_i(S) = \Phi(S \cup \{i\}) - \Phi(S). \quad (4.1)$$

Since there are multiple subsets S and possible feature orders, the marginal contribution $\Delta_i(S)$ must be weighted based on the length of the subset $|S|$ and the total count of permutations of n features.

The number of permutations where S appears before i and the remaining features appear after i are shown in Eq. 4.2

$$|S|! \cdot (n - |S| - 1)! \quad (4.2)$$

Thus, the weighting factor is normalized by the total number of permutations $n!$, described in Eq. 4.3:

$$\text{Weight}(S) = \frac{|S|! \cdot (n - |S| - 1)!}{n!}. \quad (4.3)$$

The Shapley value $\phi(i)$ aggregates the weighted marginal contributions for all subsets $S \subseteq N \setminus \{i\}$, described in Eq. 4.4

$$\phi(i) = \sum_{S \subseteq N \setminus \{i\}} \text{Weight}(S) \cdot \Delta_i(S). \quad (4.4)$$

Substituting $\Delta_i(S)$ into the Eq. 4.4 we got Eq. 4.5

$$\phi(i) = \sum_{S \subseteq N \setminus \{i\}} \text{Weight}(S) \cdot [\Phi(S \cup \{i\}) - \Phi(S)]. \quad (4.5)$$

Where,

- $\phi(i)$: The Shapley value of feature i indicates how much that feature influences the model's prediction for an individual sample.
- $\Phi(S)$: The prediction of the regression model when using only the features in subset S . Missing features ($N \setminus S$) are set to baseline or imputed values (e.g., their mean).
- $\Phi(S \cup \{i\}) - \Phi(S)$: The incremental effect of including feature i in subset S on the resulting prediction.
- $\frac{|S|!(n-|S|-1)!}{n!}$: A weighting term that guarantees fairness by considering every possible arrangement of feature combinations.
- $S \subseteq N \setminus \{i\}$: S includes every feature except i , implying that the Shapley value evaluates feature i 's influence across all possible subsets; here, N denotes the entire set of features utilized in the model.
- n : The total number of features.

For handling dimensionality issues in regression tasks, the input features x are typically represented as vectors. When computing $\Phi(S)$ for a subset S :

- Features not in S are replaced by baseline values (e.g., mean or median of the training data).
- For models requiring fixed input dimensions (e.g., neural networks), missing features can be masked or replaced with baseline embedding.

For instance, when $N = \{x_1, x_2, x_3\}$ and $S = \{x_1, x_3\}$, the resulting input vector can be represented as:

$$x = [x_1, \text{baseline}, x_3]. \quad (4.6)$$

If the regression model involves sequential data (e.g., time series or embedding):

- Subsets S must follow the sequential order.
- Masking or zero-padding is used for features outside S , ensuring that S remains interpretable.

For regression models, predictions $\Phi(S)$ are real-valued. Unlike classification, where outputs are probabilities, the contributions in regression are directly to the predicted value.

TABLE 4.1: Evaluation of ML models' performance for different classes

	Accuracy	Precision	Recall	F1-Score	Class
LR	0.73	0.73	0.73	0.73	0
	0.74	0.73	0.73	0.74	1
	0.73	0.72	0.73	0.73	2
DT	0.87	0.87	0.87	0.86	0
	0.88	0.87	0.87	0.87	1
	0.87	0.87	0.88	0.87	2
RF	0.93	0.93	0.93	0.93	0
	0.93	0.92	0.93	0.93	1
	0.93	0.93	0.93	0.92	2
SVM	0.87	0.87	0.87	0.87	0
	0.86	0.87	0.87	0.86	1
	0.87	0.87	0.86	0.86	2
NB	0.59	0.50	0.59	0.48	0
	0.60	0.58	0.58	0.52	1
	0.59	0.58	0.58	0.48	2
KNN	0.79	0.80	0.79	0.78	0
	0.80	0.80	0.80	0.79	1
	0.80	0.80	0.79	0.78	2
Gradient Descent	0.92	0.92	0.92	0.92	0
	0.92	0.92	0.91	0.92	1
	0.92	0.92	0.92	0.92	2
XgBoost	0.93	0.93	0.93	0.93	0
	0.94	0.93	0.93	0.93	1
	0.93	0.93	0.94	0.94	2

4.3 Results and Discussion

Table 4.1 showcases XGBoost's performance, demonstrating that it consistently outperforms the other models across all three categories. With an accuracy ranging between 0.93 and 0.94, XGBoost exceeds the performance of traditional models such as Gradient Descent (0.92), and RF (0.93) etc. XgBoost also dominates precision between 0.93, followed closely by Gradient Descent (0.92). The recall follows the same trend with XgBoost, achieving 0.93 to 0.94. At the same time, models like Naive Bayes struggle, particularly in Class 2, with a recall of just 0.48. Additionally, XgBoost maintains an excellent f1 score of 0.93 - 0.94 across all classes. Gradient Descent perform well but do not match the overall balanced performance of XgBoost. Furthermore, conventional models such as LR, NB, and SVM show weaker results, particularly in recall and F1-score. Overall, A model that consistently proves to be the strongest performer for this classification problem, showing reliable and superior results over all major evaluation measures.

The performance evaluation presented in Table 4.2 demonstrates the significant improvement achieved when traditional machine learning models are optimized using the Ant Colony Optimization (ACO) algorithm. An optimization process that significantly boosts key evaluation scores for every model and class when compared to their unoptimized counterparts. Notably, A hybrid model that delivers remarkably high results, achieving nearly flawless accuracy between 0.97 and 0.98 across every class. This result highlights the superior learning and generalization capability of the ACO-XgBoost model in handling complex, multidimensional data. Similarly, ACO-Gradient Descent also achieved remarkable stability, yielding accuracy and precision around 0.94–0.95, thereby confirming the robustness of ACO as an effective optimization strategy. Models such as ACO-RF and ACO-DT show moderate yet consistent improvement, maintaining accuracy around 0.89–0.93, indicating that ACO fine-tunes model parameters to achieve better convergence and reduced error rates. In contrast, algorithms like ACO-NB and ACO-KNN, while still showing improvement from their original forms, demonstrate relatively lower performance compared to ensemble and gradient-based models, suggesting that these simpler classifiers have limited adaptability even after optimization. An improvement trend that indicates the algorithm successfully maintains a proper balance between discovering new possibilities and refining known solutions throughout the optimization workflow, enabling models to identify optimal feature subsets and parameter combinations that enhance predictive accuracy. This also demonstrates the potential of hybridizing metaheuristic algorithms like ACO with classical machine learning techniques for real-world data modeling, particularly in complex and nonlinear domains. A conclusion showing that applying ACO-driven optimization leads to substantial gains in the predictive accuracy of machine learning classifiers.

The superior performance of XGBoost can be attributed to the inherent characteristics of the soil fertility dataset, which contains complex nonlinear relationships among macro- and micronutrient attributes. Unlike conventional machine learning models, XGBoost employs gradient boosting with sequential tree construction, enabling it to effectively capture intricate interactions among soil parameters such as nitrogen, phosphorus, potassium, organic carbon, pH, and micronutrients. Furthermore, XGBoost incorporates regularization mechanisms that reduce overfitting while improving generalization performance. The model also handles feature heterogeneity and varying feature importance more effectively than individual decision trees and conventional ensemble approaches. As a result, XGBoost achieves higher classification accuracy and more robust predictions. The integration of Ant Colony Optimization (ACO) further enhances performance by selecting the most informative subset of soil attributes, reducing redundancy and enabling the classifier to focus on the most influential fertility indicators.

Random Forest also demonstrated competitive performance because of its ensemble learning capability; however, it constructs trees independently and aggregates their predictions through voting. In contrast, XGBoost builds trees sequentially, where each new tree attempts to correct the errors generated by previous trees. This iterative error-minimization strategy allows XGBoost to learn subtle patterns present in the soil fertility data more effectively. Consequently, XGBoost achieved the highest overall classification performance among the evaluated models.

Table 4.3 presents a comparative analysis between the proposed ACO-XgBoost model and previously reported machine learning approaches from the literature. The comparison clearly shows

TABLE 4.2: Results of ACO optimized ML models' performance for different classes

	Accuracy	Precision	Recall	F1-Score	Class
ACO-LR	0.73	0.74	0.74	0.74	0
	0.74	0.73	0.73	0.73	1
	0.75	0.75	0.75	0.75	2
ACO-DT	0.89	0.89	0.89	0.89	0
	0.88	0.88	0.88	0.88	1
	0.89	0.89	0.89	0.89	2
ACO-RF	0.93	0.93	0.93	0.93	0
	0.93	0.92	0.93	0.93	1
	0.93	0.93	0.93	0.92	2
ACO-SVM	0.89	0.89	0.89	0.89	0
	0.89	0.88	0.88	0.88	1
	0.87	0.88	0.88	0.88	2
ACO-NB	0.52	0.52	0.52	0.50	0
	0.58	0.58	0.58	0.55	1
	0.55	0.55	0.55	0.52	2
ACO-KNN	0.64	0.64	0.64	0.64	0
	0.64	0.63	0.63	0.64	1
	0.64	0.63	0.63	0.63	2
ACO-Gradient Descent	0.95	0.95	0.95	0.95	0
	0.94	0.94	0.95	0.95	1
	0.94	0.94	0.94	0.95	2
ACO-XgBoost	0.98	0.97	0.97	0.98	0
	0.97	0.98	0.98	0.98	1
	0.98	0.98	0.97	0.98	2

that the proposed model delivers stronger results across all major evaluation measures. While earlier studies, such as Jayalakshmi et al. [216] and Chandra et al. [217], employed Random Forest (RF) models achieving accuracies of 0.96 and 0.97 respectively, and Rahman et al. [219] utilized Gaussian SVM attaining an accuracy of 0.94, none matched the overall predictive strength of the ACO-XgBoost model. Similarly, the Gaussian NB model reported by Keerthan et al. [218] yielded only moderate results with an accuracy of 0.72. Conversely, the ACO-XgBoost model produced exceptional outcomes, securing 0.98 accuracy, precision, recall, and F1-score, which indicates a remarkably stable and well-balanced classification performance. This improvement stems from the incorporation of ACO, which efficiently fine-tunes the feature selection process and optimizes the model's hyperparameters. Therefore, the proposed ACO-XgBoost framework not only outperforms traditional and previously implemented models but also establishes a new benchmark for accuracy and robustness in predictive modeling within this domain.

4.3.1 Explanation of Correct and Incorrect Classifications Using XAI

4.3.1.1 Correct Classification:

The model's prediction aligns with the proper label in Fig. 4.2 (a), where Actual: 2 and Predicted: 2. The SHAP values for each feature contribute to the final prediction, which is correctly classified as Class 2. The total model output represented as $f(x) = -3.488$, is reached through

TABLE 4.3: Results of comparison with other implemented work.

References	Model Used	Accuracy	Precision	Recall	F1-Score
Jayalakshmi et al.[216]	Boosting(C5.0)	0.96	0.96	0.97	0.96
Chandra et al.[217]	XAI based RF	0.97	0.96	0.97	0.97
Keerthan et al.[218]	RF	0.72	0.72	0.72	0.72
Rahman et al.[219]	Gaussian SVM	0.94	0.94	0.94	0.94
Proposed	ACO-XgBoost	0.98	0.98	0.98	0.98

the sum of individual feature contributions. For instance, N contributes with a value of +0.755, indicating a strong influence in the correct classification toward Class 2. Other features, such as Mn with a contribution of +0.242 and pH with -0.751, also play roles, although their contributions are smaller. The color gradient indicates each feature's SHAP value, with positive contributions in blue (pushing the output toward Class 2) and negative contributions in red (pulling it away). The cumulative sum of these contributions brings the final prediction within the correct class range, confirming the model's accurate prediction.

This section shows how well the model pinpoints the key features that most strongly influence soil-fertility classification. A balanced consideration of both positive and negative feature impacts helps the model reach accurate decisions, even when certain attributes like pH introduce contradictory information. Applying SHAP clarifies the model's reasoning, providing insight into the steps and feature influences that shape each prediction. Such openness greatly strengthens confidence in decisions generated by AI systems.

4.3.1.2 Incorrect Classification:

In Fig. 4.2 (b), Actual: 1, Predicted: 2, the model incorrectly predicts Class 2 instead of Class 1. The total model output, $f(x) = -5.011$, is influenced by individual features. Mn, with a strong contribution of +5.194, and N, with +0.907, push the prediction toward Class 2, even though the actual label is Class 1. Negative contributions from Cu (-1.626) and B (-0.469) work to move the prediction away from Class 2 but aren't enough to overcome the positive contributions. The SHAP values show that despite the negative contributions, the combination of features still leads to an incorrect prediction, with the final output closer to Class 2 rather than Class 1.

This section highlights a potential limitation of the model, where it places excessive importance on certain features like manganese and nitrogen, causing the model to make a biased prediction. Even though the negative contributions from other features should have steered the model toward Class 1, the overwhelming influence of manganese and nitrogen led to the wrong classification. This suggests that the model may need to be adjusted to ensure a more balanced consideration of all features, rather than over-relying on specific ones that could lead to errors.

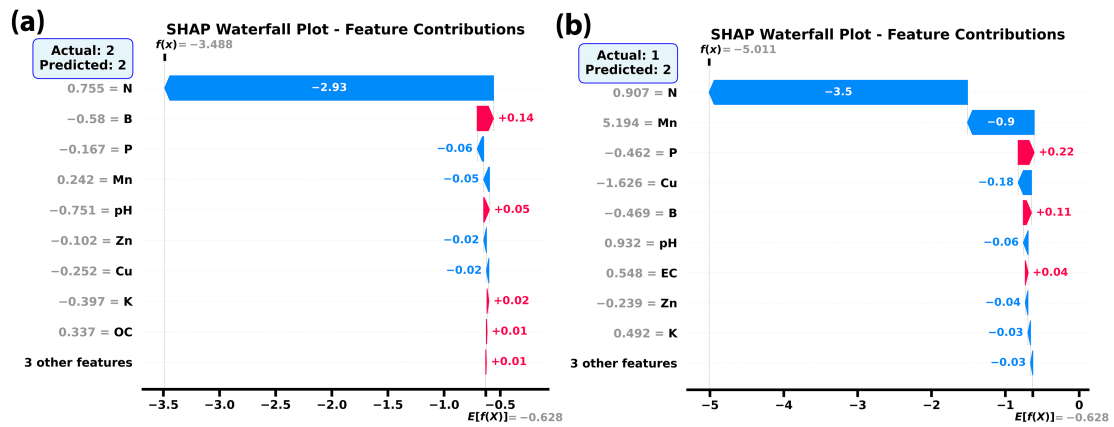


FIGURE 4.2: SHAP Waterfall Plots for Feature Contributions: (a) Correctly Classified Sample, (b) Incorrectly Classified Sample

4.4 Chapter Summary

This work developed a robust soil classification model by using micro and macro soil nutrients to classify soil fertility accurately. The proposed hybrid ML approach, integrating ACO with XgBoost, demonstrated superior performance over traditional algorithms, achieving an outstanding accuracy of 98%. The hyper-parameter optimization of XgBoost with ACO significantly enhances the model's performance. Moreover, XAI techniques provide a valuable understanding of the key soil properties driving the classification process, which ensures greater interpretability and transparency in the decision-making process. This work advances soil fertility classification and contributes to developing more efficient and interpretable machine-learning models in agricultural science, offering practical implications for precision farming and soil management.

The next chapter highlights the sustainable crop and fertilizer recommendations. With optimal irrigation and healthy soil conditions, farmers now require guidance on which crops to plant and how to fertilize them effectively. This vital advice will be delivered by a recommendation system integrating eco-friendly hybrid ML with XAI. The system's recommendations will be tailored to the soil's specific conditions, ensuring proper selection of crops for the particular environment. This optimization minimizes fertilizer waste and significantly reduces environmental impact to vigorously promote sustainable farming practices. This chapter sets the foundation for a future where agriculture is both productive and environmentally sustainable. By following these detailed recommendations, can be taken a significant step towards true sustainability in crop management.

Chapter 5

Building for the Future – Sustainable Crop and Fertilizer Recommendations

The previous chapter dealt with soil fertility and classification. So, now equipped with water requirement and idea about soil fertility, the farmers are empowered to make informed decision about crop selection and fertilizer application. Optimizing Crop selection and nutrient management guided by soil characteristics and nutrient analysis maximize benefits while minimizing the environmental impact. This chapter proposes a novel system that recommends crops and fertilizers using advanced ML techniques. Unlike conventional methods that depend on previous experiences, the proposed system recommends the most capable crop and fertilizer for a certain soil type using real-time data. The proposed system utilizes CatBoost and XgBoost to forecast the capable crops and fertilizers depends on soil nutrients. The eco-friendly design of the proposed system helps to reduce both energy consumption and CO_2 emissions. To simplify farming operations, a web application is being further created for farmers. The web application utilizes live data from soil sensors to provide instant recommendations for crops and fertilizers. This chapter demonstrates how the proposed system empowers farmers with a sustainable farming strategy.

Crop recommendations is a vital aspect in precision agriculture [220]. The rapidly growing population in West Bengal, India, demands adoption of smart agriculture to secure long-term food availability. In this context, there is an urgent need for a more focused approach to forecast capable crops depends on specific soil nutrients and land types within a limited area. Proper guidance on appropriate land management is also crucial to boost crop productivity. Effective guidance in agriculture encompasses several critical aspects, including predicting appropriate crops, recommending suitable fertilizers, estimating farming costs, and ensuring proper soil preparation. Traditionally, farmers manually predict crop productivity by using past experiences for the specific crop. Manual methods fail to mitigate the ever-changing characteristics of the soil

and environmental variables. Hence, the deployment of AI in crop yield forecasting is vital for achieving a higher accuracy and mitigating the constraint of conventional manual approaches.

An intelligent crop and fertilizer recommendation system is a new approach which uses ML and data analytics to recommend suitable crops and fertilizers for specific environmental and soil parameters. Accurate crop forecast empower farmers to take scientifically guided decisions for crop selection and planting time optimization. Crop recommendation systems help agronomists to make intelligent judgments about crop imbedding, watering schedules [172], and fertilizer application by evaluating soil and environmental parameters.

The primary focus of sustainable precision agriculture [221] optimizes crop yields by efficiently using limited land resources to strengthen a country's economic foundation. To achieve this goal, an advanced data analytics-based recommendation system is crucial for predicting land-specific crop productivity and indorsing the most capable crops for cultivation, along with the appropriate fertilizers to ensure optimal crop growth. Traditional crop endorsement systems can suggest capable crops based on limited soil nutrient data, but they often lack transparency and fail to promote environmental sustainability. In addition, conventional crop recommendation models are not practical for direct use in real-world agricultural environments. This work presents a crop and fertilizer endorsement system intended to effectively address these limitations. As far as our acquaintance, this is the first integrated system that maps crop and fertilizer recommendations into a single system based on soil's micro and macro-nutrients, specifically designed for the local conditions of West Bengal, India. The main aim of this work are as follows:

1. A sustainable crop and fertilizer recommendation system is proposed here.
2. Proposed system uses ML and optimization algorithms for precise crop recommendations.
3. To interpret the best-performing model, XAI algorithms are used to reveal the critical drivers behind each crop selection.
4. Identified an eco-friendly model, as well as optimal model by testing the performance of each model based on CO_2eq emissions and electricity consumption.
5. To develop, a farmer-friendly web app to select the most appropriate crops and the corresponding fertilizer suggestions depends on real-field data captured via IoT sensors, for optimal growth.

5.1 Literature Survey

Mahale et al [222] developed a crop endorsement and predicting system for Mumbai using ML techniques. The system employed RF for crop recommendation, achieving 92% accuracy, and an LSTM model for three-month climate forecasting. Utilizing an expectation-maximization technique for data optimization, this approach enhanced crop recommendations tailored to Maharashtra's agro-climatic zones, significantly aiding farmers' decision-making with a user-friendly interface and a crop calendar. Akkem et al [223] suggested a web-based crop endorsement system enhanced with XAI for smart farming. They utilized techniques like SHAP, LIME to clarify

model decisions. The results showed improved transparency of the system, helping farmers make data-driven choices through interpretable AI models.

Lu et al [224] created an ML-based in-season rice nitrogen level analysis using NDVI and NDRE canopy sensors. Integrating RF and XgBoost with multi-source data, their method enhanced yield and N efficiency over traditional practices by up to 15% and 71%, respectively, optimizing recommendations for on-farm applications. Senapaty et al [225] introduced a decision assistance system for crop recommendation using thirteen ML classifiers, including RF and SGD. The Synthetic Minority Oversampling Technique improved class balance, achieving an accuracy of 1.00 with the SGD method on area-specific crop data for enhanced farmer decision-making. Rahman et al [226] developed an AI and IoT-intended hydroponic approach for crop recommendation and nutrient management using sensors and ML models like Random Forest, achieving 97.5% accuracy. This system improved plant growth through continuous monitoring and real-time adjustments, enabling sustainable, resource-efficient hydroponic farming.

Sravanthi and Moparthi [227] developed an IoT-based system integrating MKGPCNN for crop selection and CGSAAN for plant illness prediction. Utilizing soil and environmental data, their model achieved 99% accuracy in crop prediction and 98% in disease identification, assisting farmers with reliable crop and fertilizer choices. Liben et al [228] applied ML to develop location-specific fertilizer recommendations (SSFR) for wheat in Ethiopia, achieving yield gains up to 25% over traditional methods. Tested on 277 farms, SSFR improved N and water use efficiency, increasing profitability by an average of 580 per hectare. Xu et al [229] generated a multi-task knowledge-based method named Shared-MMoE for crop disease diagnosis and treatment recommendation based on plant electronic medical records. Using cascaded task dependencies, their method got AUC scores of 96.33% for diagnosis and 95.36% for medication, improving accuracy and reliability in intelligent pest and disease management.

TABLE 5.1: Comparative feature analysis of proposed work with previous literature

Work (Ref.)	Deployment of XAI Models	Eco -friendly AI Model	Farmer -friendly web app.	IoT-based real-field validation
Bera <i>et. al.</i> [220]	✗	✗	✗	✗
Bhola <i>et. al.</i> [230]	✗	✗	✗	✗
Shams <i>et. al.</i> [231]	✓	✗	✗	✗
Xu <i>et. al.</i> [229]	✗	✗	✗	✗
Rahaman <i>et. al.</i> [226]	✗	✗	✗	✗
Akkem <i>et. al.</i> [223]	✓	✗	✓	✗
Sravanthi <i>et. al.</i> [227]	✗	✗	✗	✗
Dey <i>et. al.</i> [221]	✗	✗	✗	✗
Proposed	✓	✓	✓	✓

A comparative analysis with recent literature is conducted to identify existing feature gaps (Table 5.1). Although many studies have focused on AI models for crop recommendation since 2021, several limitations and gaps still exist, which are described below:

1. Many ML-based crop recommendation models were less accurate and failed in real deployment due to the use of static data and not considering the dynamic factors, such as weather and climate.
2. Most AI tools in farming do not explain their recommendations, which makes it hard for farmers to trust or understand them. Without explainability, the reasoning behind suggestions remains unclear.
3. Earlier crop recommendation systems focused primarily on yield, often overlooking sustainability.
4. Crop recommendation models often use conventional ML models. They could be more accurate if they adopt ML combined with optimization algorithms.
5. Most AI-based crop recommendation systems utilize broad datasets, often overlooking local needs. Regions like West Bengal, with unique soil and climate, do not have specific solutions. Existing models for other places are not suitable for West Bengal.
6. Most crop recommendation tools can not adapt to changing field conditions because they use static data and do not integrate with real-time IoT data.

5.2 Predictive Analysis

Due to the complicated relationship between crop and soil dietary factors, a suitable classification model is required for crop recommendation. Effective crop recommendation requires careful consideration of factors such as non-linear relationships, feature importance, model scalability, handling of heterogeneous and missing data, data distribution assumptions, and the sophistication of the problem. The specific strengths of different ML methods are utilized to address the particular needs and characteristics of the crop recommendation task. Integrating XAI into the crop recommendation system increases computational costs. Therefore, it is essential to choose ML algorithms that effectively capture the complicated non-linear association among input parameters (soil nutrient factors) and output (crop classes) while ensuring the best prediction accuracy and applicability in real world.

5.3 Hybrid ML Models

The present work uses ML models such as CatBoost[235] and XgBoost[236] for crop recommendation. Both algorithms have their effectiveness with complex data and substantial recommendation accuracy. Still, to enhance performance, hyper-parameter tuning was conducted using three optimization algorithms: SSA [232], PPSO [233], and GWO [234].

Six optimized hybrid models were made by pairing CatBoost[235] and XgBoost[236] with each optimization algorithm. These hybrid models were then evaluated using statistical methods to estimate accuracy, precision, recall, and f1-score to determine the best model for crop recommendation.

5.4 Proposed Crop Recommendation Model

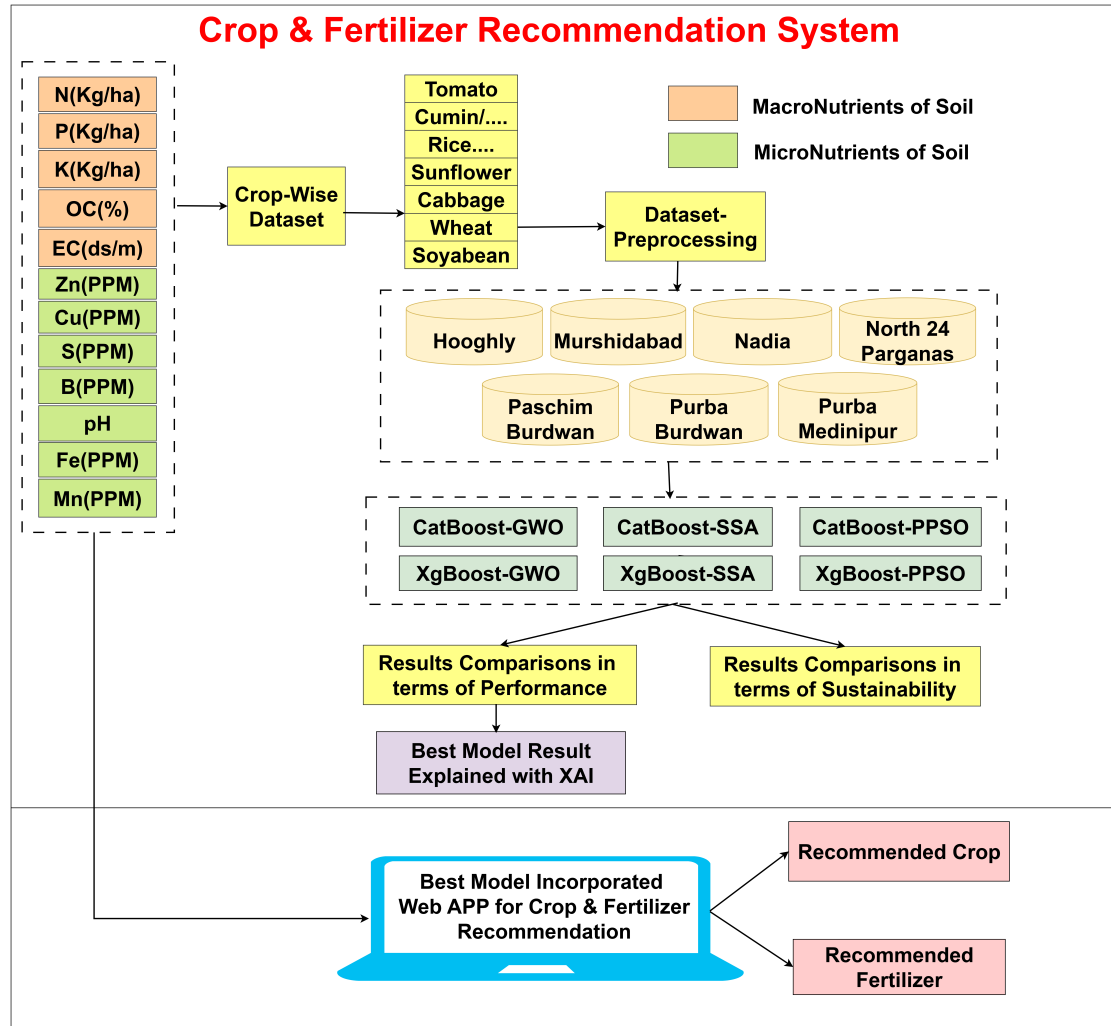


FIGURE 5.1: Workflow of the sustainable crop & fertilizer Recommendation system using hybrid ML, XAI, and eco-friendly AI

5.4.1 Workflow of Proposed Recommendation System

The workflow diagram (Fig. 5.1) represents a systematic process for implementing the proposed crop and fertilizer recommendation system. The steps are elaborated as follows:

1. The first step is the development of a sustainable and efficient crop and fertilizer recommendation system suitable for smart farming applications. The recommendation system is structured in two stages: stage 1 focuses on crop recommendation, while stage 2 provides fertilizer recommendations based on the crops selected in the first stage.
2. In the second step, data is collected and pre-processed. The dataset is collected from the soil health card portal (Ministry of Agriculture and Farmers Welfare, Government of India) [238] for seven different districts (Hooghly, Murshidabad, Nadia, North 24 Parganas,

Paschim Bardhaman, Purba Bardhaman, Purba Medinipur) of West Bengal, India, which includes soil parameters such as Cu, B, S, Fe, Zn, Mn, pH, N, K, P, OC, and EC. Missing values are handled and normalization is applied to ensure the quality of dataset and its suitability for ML models. After applying pre-processing techniques, district-wise dataset's records are labelled with crops using the research document [239] created by ICAR-Indian Institute of Soil Science, Bhopal.

3. ML models CatBoost[235] and XgBoost[236] are utilized in this step for predictive analysis. Optimization techniques, such as SSA[232], PPSO[233], and GWO [234], are applied to fine-tune hyper-parameters to enhance model accuracy and efficiency. State-of-the-art work is also compared in this step with the hybrid models optimized with benchmark optimizers such as Genetic Algorithm(GA) (Section 3.7) and PSO [237].
4. In this step, XAI techniques such as LIME (Section 6.2.5.1) and SHAP (Section 4.2.2.1) are applied to ensure transparency in model predictions of the best performing model.
5. In this step, CO_2eq emissions and electricity consumption are evaluated for each hybrid model to determine the most eco-friendly one.
6. A streamlit-based (python framework) web application is developed to assist farmers in selecting the most suitable crops based on soil nutrient levels and recommending optimal fertilizers for optimum crop growth.
7. Fertilizers suggestions are validated based on real-time field data collected via IoT sensors.

5.4.2 Eco-friendly AI Metrics and CodeCarbon

In eco-friendly AI, it is crucial to assess the weather impact of AI models, especially in terms of their carbon emissions. Eco-friendly AI metrics offer performance indicators for evaluating ecological effects throughout an AI model's life-cycle. CodeCarbon is a Python tool that measures carbon emissions and electricity consumption from computational processes, particularly during the implementation of ML models.

5.5 Crop-Wise Soil Nutrients-Based Fertilizer Recommendation System

The crop-wise soil nutrients-based fertilizer recommendation system was previously created under the AICRP on STCR, providing a scientific and data-driven approach for optimizing fertilizer application across different crops and soil conditions with 20% increased yield compared to traditional fertilizer application methods.[239]

Aligned with the STCR-based fertilizer recommendation methods, Algorithm 6 is developed here to compute fertilizer requirements systematically. It uses soil nutrient levels and target yield to ensure efficient, site-specific nutrient management, promoting optimum crop productivity and preserving sustainable soil health. To accurately calculate the required fertilizer amount, A is

used as the yield coefficients for N , P_2O_5 , and K_2O while B represents the soil contribution coefficients. The values of constant A and B for each crop are shown in Table 5.2.

Algorithm 6 Fertilizer Requirement Calculation Algorithm

```

1: Inputs:
2:    $T \leftarrow$  Targeted yield (q/ha)
3:    $S = (N_0, P_0, K_0) \leftarrow$  Initial soil test values (kg/ha)
4:    $A = (A_N, A_P, A_K) \leftarrow$  Yield coefficients for  $N$ ,  $P_2O_5$ , and  $K_2O$ 
5:    $B = (B_N, B_P, B_K) \leftarrow$  Soil contribution coefficients
6: Initialize:
7:    $F_N \leftarrow 0, F_{P_2O_5} \leftarrow 0, F_{K_2O} \leftarrow 0$ 
8: for each nutrient  $X$  in  $\{N, P_2O_5, K_2O\}$  do
9:   if  $X = N$  then
10:     $F_N \leftarrow \max(0, A_N \times T - B_N \times N_0)$ 
11:   else if  $X = P_2O_5$  then
12:     $F_{P_2O_5} \leftarrow \max(0, A_P \times T - B_P \times P_0)$ 
13:   else if  $X = K_2O$  then
14:     $F_{K_2O} \leftarrow \max(0, A_K \times T - B_K \times K_0)$ 
15:   end if
16: end for
17: Output:
18:   return  $(F_N, F_{P_2O_5}, F_{K_2O})$ 

```

5.6 Experiment Design

This section gives a detailed description of the dataset, along with a detailed analysis and the experimental setup necessary for implementing the crop and fertilizer recommendation system.

5.6.1 Dataset Description & Data Pre-Processing

The dataset used in this work consists of 33,134 samples across twelve input variables related to soil micro and macro-nutrients. The dataset is collected manually from the soil health card portal [238] for seven different districts (Hooghly, Murshidabad, Nadia, North 24 Parganas, Paschim Bardhaman, Purba Bardhaman, Purba Medinipur) of West Bengal, India. Key variables include Cu, B, S, Fe, Zn, Mn, pH, N, K, P, OC and EC. A categorical "Label" variable with 13 unique crop varieties, such as "Tomato", "Cumin / Isabgol / Guar / Mothbean", "Rice-CR 1009 / ADT-31 / ASD-16 / IR-20", "Sunflower", "Cabbage", "Wheat" and "Soybean" furnish a range of crop choices.

After collecting data for each district of West Bengal, India, several pre-processing tasks are performed on the dataset. Missing values are identified across district-wise datasets and imputed using the KNN imputation method. Once the missing values are handled, each district-wise dataset records are labelled with crops using research document published by ICAR, Bhopal [239]. For better clarity and efficient crop management, crops are classified into distinct classes. To enhance the model performance, bias reduction, and generalization improvement, SMOTE [241] is applied to balance the classes for each district.

TABLE 5.2: Nutrient Factors for Fertilizer Recommendation for various crops

Tomato		
Nutrient X	A_X (Yield Factor)	B_X (Soil Factor)
Nitrogen (N)	5.33	0.46
Phosphorus (P_2O_5)	3.88	4.16
Potassium (K_2O)	5.16	0.25
Cumin..		
Nutrient X	A_X (Yield Factor)	B_X (Soil Factor)
Nitrogen (N)	5.38	0.46
Phosphorus (P_2O_5)	5.07	2.46
Potassium (K_2O)	4.86	0.34
Rice-CR 1009..		
Nutrient X	A_X (Yield Factor)	B_X (Soil Factor)
Nitrogen (N)	3.28	0.18
Phosphorus (P_2O_5)	4.80	5.02
Potassium (K_2O)	2.83	0.54
Sunflower		
Nutrient X	A_X (Yield Factor)	B_X (Soil Factor)
Nitrogen (N)	31.4	1.63
Phosphorus (P_2O_5)	8.98	2.91
Potassium (K_2O)	4.70	0.41
Cabbage		
Nutrient X	A_X (Yield Factor)	B_X (Soil Factor)
Nitrogen (N)	1.57	0.62
Phosphorus (P_2O_5)	0.60	0.91
Potassium (K_2O)	0.48	0.09
Wheat		
Nutrient X	A_X (Yield Factor)	B_X (Soil Factor)
Nitrogen (N)	12.88	0.80
Phosphorus (P_2O_5)	2.15	0.23
Potassium (K_2O)	4.65	0.29
Soybean		
Nutrient X	A_X (Yield Factor)	B_X (Soil Factor)
Nitrogen (N)	6.60	0.35
Phosphorus (P_2O_5)	6.05	3.19
Potassium (K_2O)	3.86	0.21

5.6.2 Experimental Environment

The proposed work is implemented in Python environment with Pandas 2.2.3, NumPy 2.0, Matplotlib, Seaborn, Optuna 4.1.0, Hyperopt 0.2.7, SciPy 1.14.1, CodeCarbon 2.7.2, Scikit-Learn, LIME and SHAP. Experiments are conducted on a Linux machine with 12 GB RAM, an AMD Ryzen 5 5600H CPU @ 2.70GHz (12 cores), and Radeon Graphics.

As mentioned earlier, ML algorithms like XgBoost and CatBoost are tuned using three optimization algorithms: SSA, PPSO, and GWO have been used for crop recommendation. The hyper-parameters are set to ideal values for optimal performance: SSA is configured with a population size of 500, a max of 1000 iterations, and a coefficient $C1$ starting at 2 and decreasing linearly to 0 [242]. PPSO utilizes a population size of 500, 1000 iterations, an inertia weight w of 0.7 (dynamically adjusted from 0.9 to 0.4), and coefficients $c1 = 1.5$ and $c2 = 1.5$ [243]. The

GWO setup includes a population size of 500, 1000 iterations and an exploration-exploitation parameter a that decreases linearly from 2 to 0 [244].

5.6.3 Evaluation Metrics

The performance of the proposed crop recommendation models was evaluated using Accuracy, Precision, Recall, and F1-score. These metrics were computed from the confusion matrix and are defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5.1)$$

$$Precision = \frac{TP}{TP + FP} \quad (5.2)$$

$$Recall = \frac{TP}{TP + FN} \quad (5.3)$$

$$F1-Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5.4)$$

Where TP, TN, FP, and FN denote True Positive, True Negative, False Positive, and False Negative predictions, respectively.

5.7 Results

The values of Accuracy, Precision, Recall, and F1-score obtained for all hybrid models are summarized in Tables 5.3 and 5.4. These metrics were used to compare the predictive performance of the competing crop recommendation models and to identify the most effective hybrid framework. The confusion matrices in Fig. 5.2 highlight the performance of the CatBoost method optimized with SSA across all districts compared to other hybrid models (CatBoost-GWO, CatBoost-PPSO, XgBoost-GWO, XgBoost-SSA, XgBoost-PPSO). In Hooghly district (see Fig. 5.2(a)), the model correctly classifies 63 samples, achieving 99.64% accuracy. For Murshidabad (see Fig. 5.2(b)), it attained an accuracy of 99.69%. Similarly, in Nadia (see Fig. 5.2(c)), CatBoost-SSA provides an accuracy of 99.80%. This CatBoost-SSA model also shows exceptionally well performance in North 24 Parganas (see Fig. 5.2(d)) and Paschim Bardhaman (see Fig. 5.2(e)) with no misclassifications. In Purba Bardhaman, the model maintained an excellent accuracy of 99.75%(see Fig. 5.2(f)). For Purba Medinipur (see Fig. 5.2(g)), CatBoost-SSA continues to deliver reliable performance with no misclassifications.

Each hybrid model is executed 30 times for every seed value, ranging from 25 to 55, and the final score are obtained by averaging the results across these runs. Additionally, Table 5.3 presents

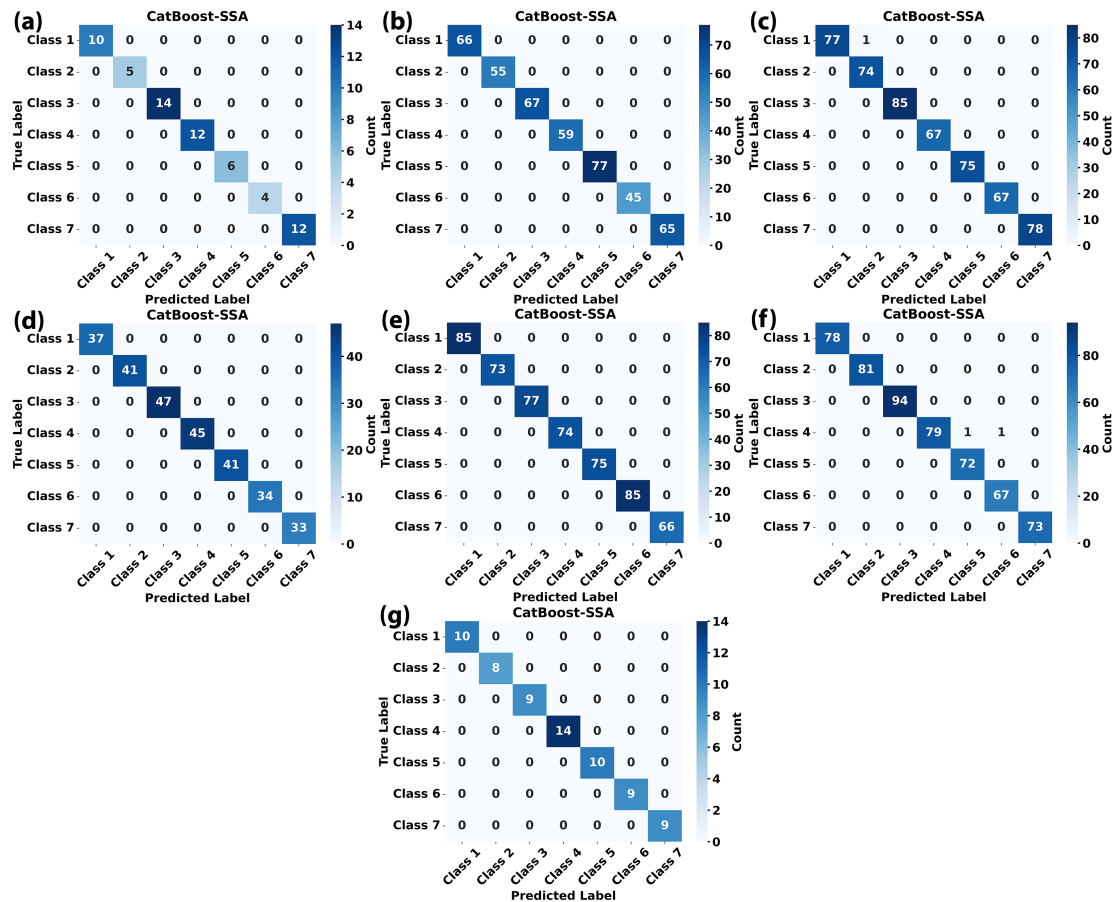


FIGURE 5.2: Confusion matrices for best model in seven districts, illustrating the true vs. predicted class distributions

the best-performing algorithm for each district. Among the CatBoost-based hybrid models, SSA optimization consistently delivers the highest performance across all districts, achieving an accuracy range of 99.64% to 99.87%. Thus, SSA greatly enhances the classification performance of the CatBoost model, highlighting its vital role as an optimization technique in achieving exceptional classification results with CatBoost.

Table 5.4 presents a relative analysis of CatBoost and XgBoost method optimized using 5 different algorithms, including benchmark models such as GA and PSO, along with SSA, GWO, and PPSO. Among the CatBoost-based models, SSA achieves the best performance with an accuracy range of 99.64% - 99.87% and perfect score (0.99) for seven different districts. PPSO also significantly improves CatBoost's performance, yielding near-perfect results with an accuracy range of 92.75% - 99.23%. In contrast, the benchmark models GA and PSO exhibit slightly lower performance. This analysis highlights SSA's superior optimization capability in enhancing classification performance.

TABLE 5.3: The performance of the CatBoost model optimized with SSA are reported, as CatBoost-SSA consistently delivers the best results across all districts

	Hooghly	Murshidabad	Nadia	North 24 Parganas	Paschim Bardhaman	Purba Bardhaman	Purba Medinipur
Accuracy	99.64	99.69	99.80	99.75	99.87	99.75	99.82
Precision	0.99	0.99	0.99	0.99	0.99	0.99	0.99
Recall	0.99	0.99	0.99	0.99	0.99	0.99	0.99
F1-Score	0.99	0.99	0.99	0.99	0.99	0.99	0.99

TABLE 5.4: Performance of CatBoost and XgBoost models with benchmark optimizers GA and PSO

XgBoost Based Model					
	XgB-GA	XgB-PSO	XgB-GWO	XgB-SSA	XgB-PPSO
Accuracy	64.32 - 85.11	68.77 - 87.21	80.18 - 99.43	72.13 - 98.87	88.09 - 97.10
Precision	0.62 - 0.84	0.67 - 0.88	0.80 - 0.98	0.72 - 0.98	0.88 - 0.97
Recall	0.63 - 0.85	0.67 - 0.88	0.80 - 0.98	0.72 - 0.98	0.88 - 0.97
F1-Score	0.63 - 0.84	0.68 - 0.87	0.80 - 0.98	0.72 - 0.98	0.88 - 0.97
CatBoost Based Model					
	CatB-GA	CatB-PSO	CatB-GWO	CatB-SSA	CatB-PPSO
Accuracy	76.42 - 91.42	76.86 - 92.43	79.30 - 99.62	99.64 - 99.87	92.75 - 99.23
Precision	0.76 - 0.91	0.74 - 0.92	0.79 - 0.99	0.99 - 0.99	0.92 - 0.99
Recall	0.76 - 0.91	0.76 - 0.91	0.79 - 0.99	0.99 - 0.99	0.92 - 0.99
F1-Score	0.77 - 0.91	0.76 - 0.92	0.79 - 0.99	0.99 - 0.99	0.92 - 0.99

5.8 Results Comparison with Existing Work

This section compares the performance of the proposed crop recommendation framework with existing crop recommendation systems reported in the literature. The comparison is based on Accuracy, Precision, Recall, and F1-score obtained from crop recommendation datasets. The results presented in Table 5.5 do not correspond to rainfall prediction or fertilizer recommendation tasks; rather, they specifically evaluate the effectiveness of the proposed crop recommendation model against previously published crop recommendation approaches.

It highlights that, despite the differences in soil nature, climate conditions, and diverse soil growth patterns across the districts, the CatBoost model optimized with SSA outperforms other models. This demonstrates the effectiveness and adaptability of the proposed system, which can handle varied weather conditions. Despite these challenges, the proposed method demonstrates its capability for broader applicability in precision agriculture across various regions.

TABLE 5.5: Results comparison with existing work

Work(Ref.)	Accuracy	Precision	Recall	F1-Score
Wankhade <i>et. al.</i> [245]	95.10%	0.95	0.94	0.95
Bera <i>et. al.</i> [220]	64.71%	0.63	0.64	0.64
Sravanthi <i>et. al.</i> [227]	96.93%	0.96	0.95	0.95
Gopi <i>et. al.</i> [246]	93.18%	0.92	0.93	0.93
Bhola <i>et. al.</i> [230]	66.67%	0.65	0.65	0.65
Abdullahi <i>et. al.</i> [247]	78.80%	0.74	0.62	0.65
Proposed	99.64 - 99.87	0.99 - 0.99	0.99 - 0.99	0.99 - 0.99

5.9 Sustainable AI Assessment

In evaluating the sustainability of AI models, key factors such as energy consumption, CO_2eq emissions, and computational efficiency are considered along with predictive accuracy. The sustainability assessment of CatBoost and XgBoost models are optimized using GWO, PPSO, and SSA across seven different districts.

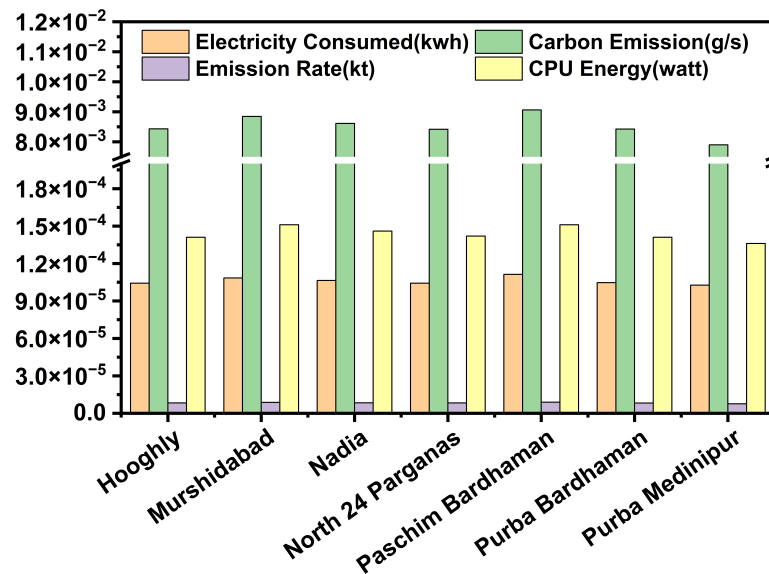


FIGURE 5.3: A comparison of electricity consumption, CO_2eq emissions, emission rate, and CPU energy usage for the best performing model (XgBoost-PPSO) across seven districts

As shown in Fig. 5.3, the XgBoost-PPSO model performs optimally well across all districts in terms of sustainability, balancing electricity consumption, CO_2eq emissions, and CPU energy usage. Models that consume less electricity emit lower CO_2eq emissions and require minimal CPU energy are more desirable. As observed CatBoost-SSA model demonstrates the highest predictive accuracy in Section 5.7, also maintaining relatively balanced sustainability metrics. The electricity consumption of the CatBoost-SSA model across all districts remained in the range of 1.063×10^{-4} kWh to 1.127×10^{-4} kWh, with corresponding CO_2eq emissions ranging from 8.521×10^{-3} g/s to 9.224×10^{-3} g/s. While this is marginally higher than some alternative models, it remains within an acceptable threshold for sustainable AI practices. The emission

rates, varying between 8.400×10^{-6} kt and 9.900×10^{-6} kt, and CPU energy consumption of the CatBoost-SSA model remains in the range of 1.460×10^{-4} watts to 1.640×10^{-4} watts, which is slightly higher than some models and also not significantly different from its counterparts. Notably, XgBoost models, especially those optimized with PPSO and SSA, demonstrate slightly lower CPU energy usage in certain cases, yet they do not outperform CatBoost-SSA in prediction accuracy. CatBoost-SSA is not the most eco-friendly model, but its prediction accuracy made it an automatic choice for applying in real-world scenarios where both performance and environmental efficiency are crucial.

5.10 The Significance of Features Towards Recommendation Through XAI

The main features that influence the recommendation process of each district are shown through the SHAP GSP. SHAP GSP for each district, highlighting the most significant features prompting their recommendation processes, are shown in Fig. 5.4. Fig. 5.4 (a) - (g) shows that for all AI models, 'N', 'P' and 'K' are the most influential features. The importance of various features in effective crop recommendation is shown by the length of the bar along the 'Mean Absolute SHAP Value' axis in Fig. 5.4.

5.10.1 Accurate and Inaccurate Recommendation Explained with XAI

As observed 'N', 'P', and 'K' are the most influential factors for overall crop recommendations across all districts. This section analyzes individual feature contributions to both accurate and inaccurate recommendations for each class across the districts. As an example, the accurate and inaccurate recommendations are discussed here for Hooghly district only.

5.10.1.1 Accurate Recommendation

The LIME local plot in Fig. 5.5(a), generated using LIME, shows that 'N(kg/ha)', 'OC (%)', and 'pH' have the most significant negative contributions, while features like 'EC (dS/m)', 'Cu (ppm)', 'P (kg/ha)', 'S (ppm)', 'B (ppm)', 'K (kg/ha)', and 'Mn (ppm)' have minor positive or negative influences. The SHAP waterfall plot in Fig. 5.5(b), gives feature contributions compared to LIME. In Fig. 5.5(a) - (b), x-axis shows various features, and y-axis shows impacts of each feature on the final recommendation. It highlights 'OC (%)', 'N (kg/ha)', 'pH', 'P (kg/ha)', and 'EC (dS/m)' as the dominant factors with positive SHAP values (+4.05), (+3.16), (+1.12), (+0.61), (+0.47) respectively, while features such as 'K (kg/ha)' (-0.31), 'Cu (ppm)' (-0.24), and 'S (ppm)' (-0.13) have a negative impact on the recommendation. Conversely, 'B (ppm)' (+0.3) and three other features (+0.1) have minor positive effects on the recommendation.

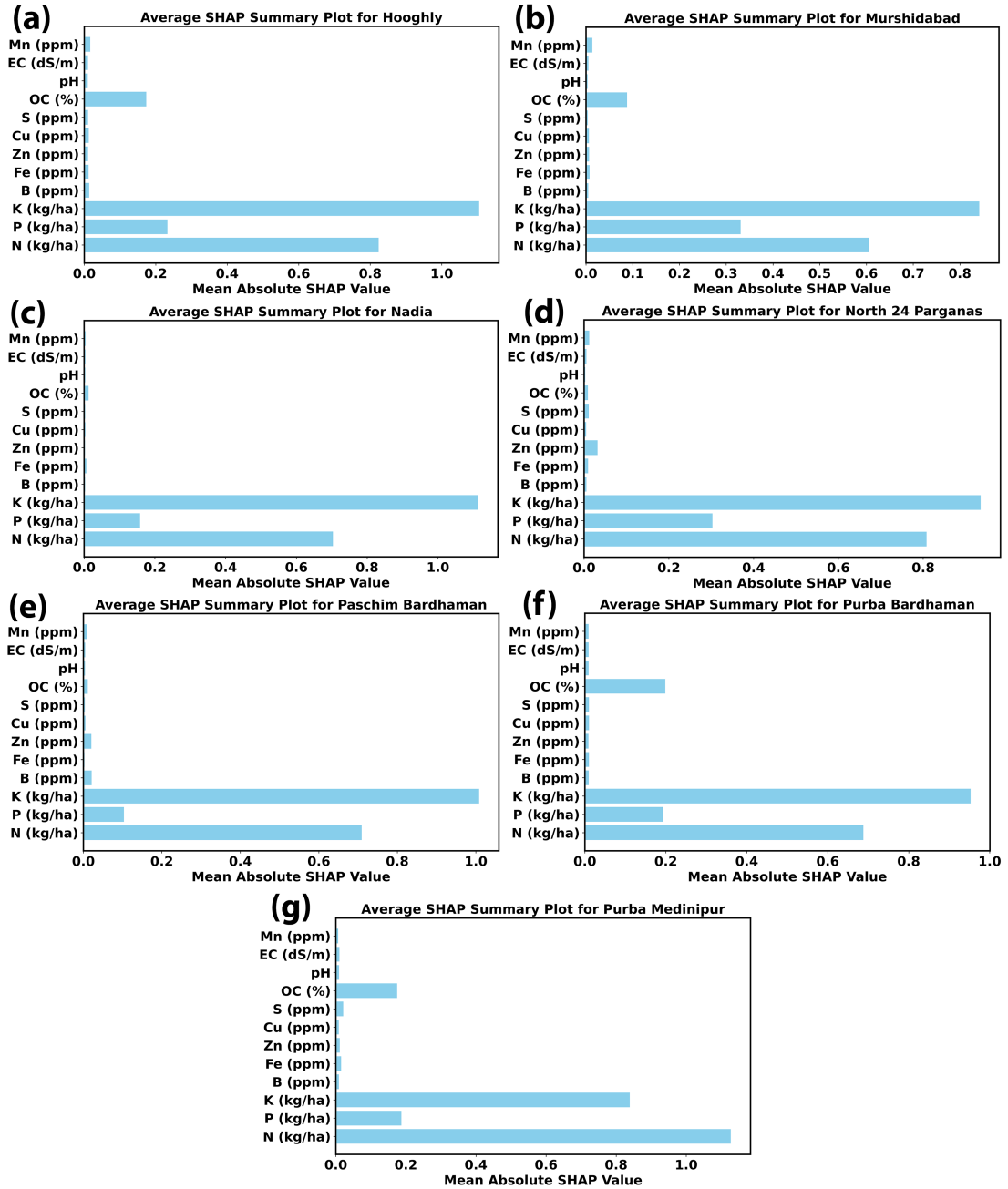


FIGURE 5.4: SHAP summary plots: feature importance analysis for crop recommendations using XAI for each seven districts

5.10.1.2 Inaccurate Recommendation

The LIME local plot in Fig. 5.5(c), generated using LIME, shows that 'N(kg/ha)' and 'pH' contribute most significantly in a positive way, while features like 'EC (dS/m)', 'OC (%)', 'Cu (ppm)', 'P (kg/ha)', 'S(ppm)', 'B(ppm)', 'K(kg/ha)', and 'Mn(ppm)' have minor positive or negative influences. The waterfall plot in Fig. 5.5(d), gives a more detailed feature contributions compared to LIME. In Fig. 5.5(c) - (d), the x-axis shows the various features, and the y-axis shows the impacts of each feature on the local recommendation. It highlights 'OC (%)',

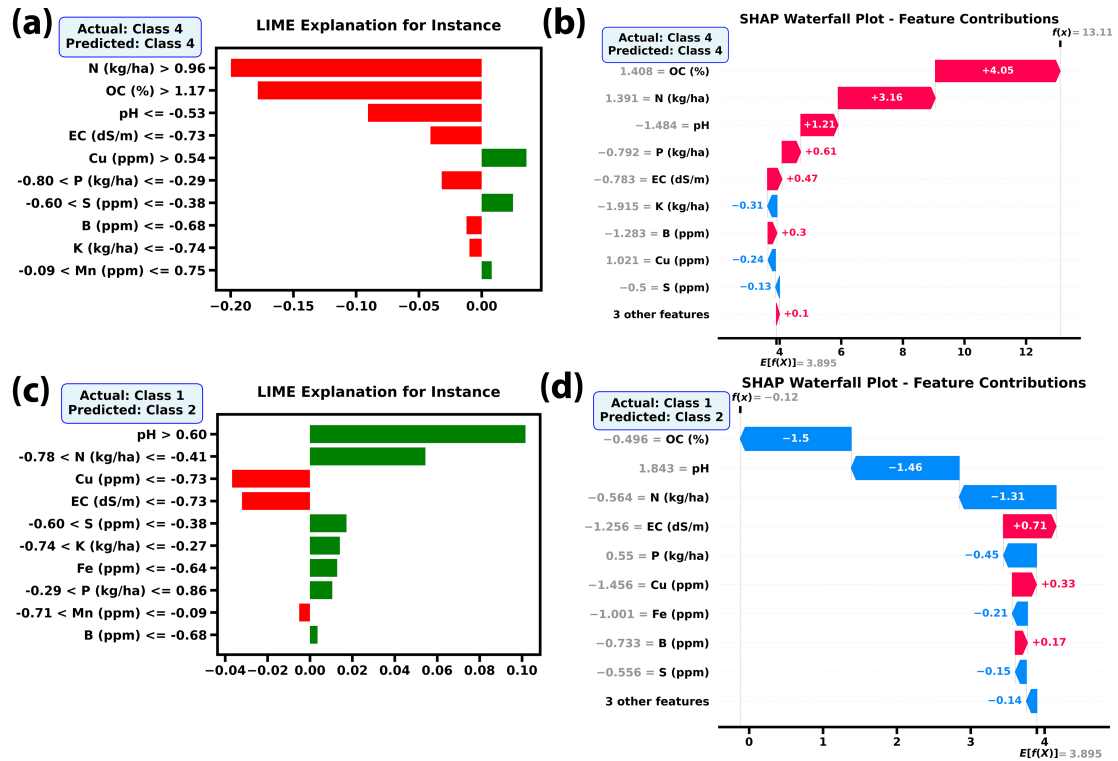


FIGURE 5.5: LIME and SHAP interpretations for model predictions in Hooghly district: (a), (c) LIME feature importance explanations for correctly and incorrectly classified instances, and (b), (d) Waterfall plots showing feature contributions to the model predictions

'N(kg/ha)', and 'pH' as dominant factors with negative SHAP values (-1.5), (-1.46), and (-1.31), respectively, while features such as 'EC (dS/m)' (+0.71), 'Cu(ppm)' (+0.33), and 'B(ppm)' (+0.17) provide a positive impact on the accurate recommendation. Conversely, 'P (kg/ha)' (+0.3) and '3 other features' (+0.1) have minor positive effects on the recommendation.

5.11 A Farmer-Friendly Web Application for Crop and Fertilizer Recommendation

The web app, presented in Fig. 5.7, assists farmers for crop selection and fertilizer recommendations. Using ML models, the web application provides instant recommendations for precision and sustainable agricultural practices.

The web application is developed using Streamlit Python framework with the best-performing CatBoost-SSA model. Given input parameters like macro-nutrient and micro-nutrient levels in the soil, including N, P, K, OC, EC, pH, Cu, B, S, Fe, Zn, and Mn in the web application, the suitable crop will be recommended instantly. As shown in Fig. 5.7(a), after giving the soil nutrient values for the selected garden located at Haldia, the web application recommends 'Sunflower' as the most suitable crop. Following this recommendation, Sunflower seeds are planted on January 30th, 2025. The web application also integrates an IoT-based NPK sensor

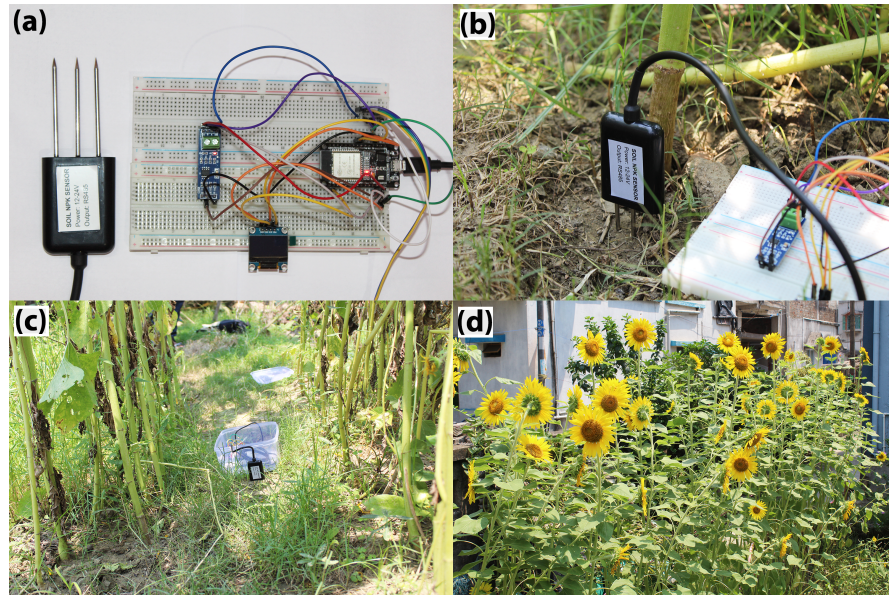


FIGURE 5.6: (a) Breadboard setup with NPK sensor, RS485 module, OLED display and ESP32 micro-controller unit for data acquisition. (b) NPK sensor inserted near sunflower roots for real-time soil nutrient monitoring. (c) Distributed sensor nodes deployed along the sunflower plantation rows. (d) Enhanced sunflower growth observed after applying NPK nutrients based on sensor data feedback.

that continuously monitors NPK levels throughout the Sunflower's growth cycle (see Fig. 5.6). A summary of the collected NPK sensor data is presented in Table 5.6. In addition to crop recommendations, the web application is further used for fertilizer recommendations along with the required quantity of fertilizer to achieve the desired yield. By providing the target yield (kg/ha) and NPK sensor data as input, the application shows the required fertilizer quantities, as shown in Fig. 5.7(b). For the recommended crop, Sunflower (Class 4), the web application suggests fertilizers such as N: 131 kg per hectare, P: 181 kg per hectare, and K: 48 kg per hectare. The observed improved growth of Sunflower (see Fig. 5.6(d)) after applying the suggested fertilizers demonstrates the web application's ability to support sustainable and efficient farming practices.

5.12 Chapter Summary

The recommendation system demonstrates the substantial potential of ML methods in addressing the challenges of recommendation in precision agriculture. By using locally relevant soil nutrient data from West Bengal, India, the system gives crop and fertilizer recommendations to increase agricultural product while decreasing environmental impact. The use of advanced ML algorithms, such as CatBoost optimized with the SSA, shows an outstanding accuracy of nearly $\approx 99\%$, and the level of precision not only highlights the superiority of the model in crop recommendation but also proves its robustness when applied to diverse soil conditions. Integrating XAI in this work increases transparency by revealing the key drivers behind each crop selection. The sustainability aspect of the model remains equally important, as it accounts for CO_2eq emissions and energy consumption during model execution. While the XgBoost-PPSO

TABLE 5.6: Summary of NPK sensor data collected in a sunflower garden from 30th January 2025 to 5th April 2025, indicating levels of NPK in kg/ha.

Date	N(Kg/ha)	P(Kg/ha)	K(Kg/ha)
30-Jan-2025	302.12	28.04	232.41
31-Jan-2025	302.11	27.02	232.57
01-Feb-2025	301.21	23.54	212.69
.	.	.	.
.	.	.	.
28-Feb-2025	275.03	15.39	200.41
01-Mar-2025	275.01	14.79	196.11
02-Mar-2025	274.86	14.17	198.23
.	.	.	.
.	.	.	.
03-Apr-2025	340.31	29.46	296.07
04-Apr-2025	338.85	28.58	296.10
05-Apr-2025	336.19	25.52	295.26

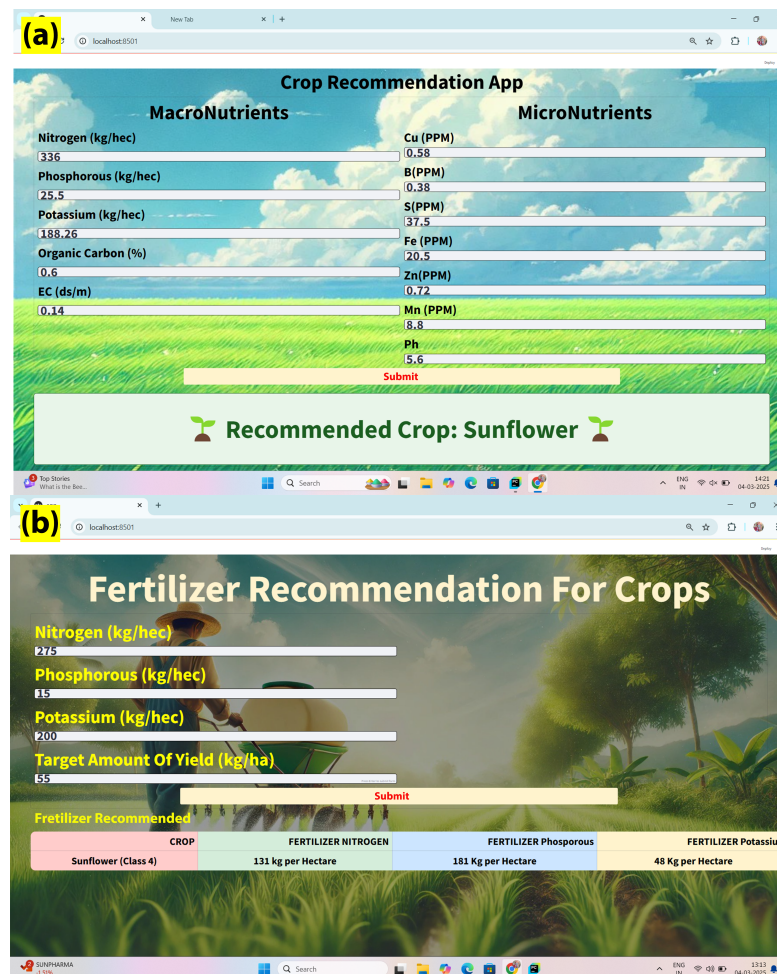


FIGURE 5.7: (a) AI-powered crop recommendation web application: analyzing macro and micro-nutrient levels to suggest optimal crop selection, (b) AI-based fertilizer recommendation system: web application for optimized crop fertilizer management

model records greater environmental efficiency, the CatBoost-SSA model is still chosen due to its superior accuracy. Thus, CatBoost-SSA makes it an automatic choice for real-world deployment, where both predictive power and environmental efficiency are essential. Furthermore, the creation of a farmer-friendly web app integrated with IoT sensors bridges the gap between complicated ML methods and their implementation in real crop fields.

The next chapter will explain how XAI helps to predict crop yields and clarify which factors significantly influence yield prediction. By making these decisions more transparent, farmers will feel more confident in using the system for their farming practices. In the next chapter, farmers gain both precise guidance and a clear understanding of yield prediction.

Chapter 6

The Final Frontier – Enhancing Farmer Decision-Making with Explainable AI

Making informed decisions in farming is crucial. For sound decision-making, farmers must to consider various aspects including factors like soil conditions, climatic patterns, and resource availability. In traditional crop yield prediction techniques, farmers have to make random decisions without adequate support. This chapter offers an Explainable Artificial Intelligence (XAI) (Section 6.2.1) based novel framework to provide crop yield predictions with crucial insights. Moving beyond the sustainable crop and fertilizer recommendations covered in the previous chapter 5, this chapter focuses on empowering farmers with a deeper understanding of crop yield predictions.

A key challenge highlighted in this chapter is the inherent difficulty in understanding traditional ML models. Despite their accuracy, the traditional ML models typically do not offer clear interpretability of their predictive outcomes. Without knowing the reasons behind the specific predictions, it is difficult for the farmers to trust the system. The proposed XAI framework (Section 6.2.1) addresses this problem by offering comprehensive explanations for crop yield predictions. In this framework, farmers can grasp the specific influence of variables like temperature, precipitation, and soil condition etc. The previous chapter highlighted the importance of using eco-friendly ML models (Section 5.4.2) to reduce environmental footprint while improving efficiency. Along with accuracy, this chapter steps further by ensuring transparency and interpretability.

In agriculture [248], AI and ML have become an essential tools nowadays. Regarding accuracy, complex models [249] are often better than simple statistical models [250]. ML models like RF and SVM (Chapter 2) can directly work with unstructured data, while traditional models need specific data formats. Also, models like RF and SVM are proficient at detecting hidden nature and relationships in data.

The utilization of AI and ML in statistical modeling can be good as well as bad at the same time. While AI and ML enhance the accuracy of predictions, interpretation of the predicted results remains a challenge [251]. Statistical models like RF and gradient boosting are all efficient algorithms, but their predicted outcomes are still quite hard to interpret [251].

Making a equilibrium between accuracy and interpretability of AI models rubble a hurdle [252]. However, it is widely acknowledged that ideal models should be precise and understandable. Clear explanations are even more crucial for decision-making than higher accuracy with a black-box model [253]. This problem can be solved using XAI [253]. XAI focuses on creating tools that make complex algorithms understandable without compromising accuracy [254]. XAI makes a balance between understandability and accuracy. XAI has increased significant traction in recent years, spreading its wings across various scientific disciplines like biodiversity research [255], geoscience [256], hydrology [257], etc. In agriculture, the adoption of XAI techniques has begun to blossom since 2020 [253]. Applications in areas like crop yield estimation [258], crop classification [259], and even intelligent irrigation systems [260][172] are observed. However, using XAI methods in agriculture remains limited to a few approaches. Despite this, it is clear that the full potential of XAI and its diverse techniques is yet to be fully explored within the agricultural domain.

To empower farmers with AI-driven decision-making, it is crucial to provide explanations behind crop yield predictions. For explanation-orientated prediction, an XAI framework [261] that uses three techniques: SHAP [262] for global model insights, counterfactual explanations to explore 'what-if' scenarios [263], and LIME [264] for understanding individual predictions is proposed in this chapter. The framework consumes agricultural data from real crop fields and generates explanations with predictions. Equipped with these explanations based on prediction, farmers can make more informed agrarian decisions.

The proposed work provides explanation based on improved crop yield prediction through XAI. Primary contributions of this work are listed below:

- An XAI framework is developed to interpret the rationale underlying the model's outputs of AI models in crop yield prediction.
- The introduced XAI framework offers a comprehensive understanding by offering global, local and counterfactual explanations to explore 'what-if scenarios.
- The effectiveness of the introduced XAI framework is assessed through two real-world agricultural crop-yield datasets.
- A novel XAI-based feature selection approach is introduced here to detect, prioritize, and assess the influence of key variables in predicting crop yield across ML models.
- A comprehensive sensitivity analysis is further introduced to check the effects of individual features and the impact of hyper-parameter tuning across the ML models.

6.1 Literature review

Shams et al. [231] proposed integration of XAI into crop recommendation systems (CRS) to enhance transparency and interpretability. Author utilized LIME to provide clear and accurate recommendations. Proposed method outperformed traditional ML models in metrics like MSE and R^2 . Akkem et al. [223] presented a study that enhances crop recommendation systems using XAI techniques to improve transparency and faith in AI-based innovative farming solutions. The work [223] explored state-of-the-art XAI methods such as LIME, SHAP, IG, and LRP to provide interpretable insights into AI-based crop recommendations. Cartolano et al. [265] analyzed the use of XAI models in smart system to enhance the understandability of AI-enabled crop recommendation models. The study used SHAP and LIME to explain the rationale behind decision-making of ML models for predicting optimal crop selection based on various agricultural parameters. Malashin et al. [266] presented a study on forecasting sustainable crop yields using deep learning and XAI tools to enhance agricultural decision-making. The research employed a DNN optimized through GAs to achieve high predictive performance, with an R^2 score of 0.92, and utilized XAI methods like LIME to improve model understandability. Pei et al. [267] investigate the impact of phenology on improving crop yield forecasting by comparing Ground-observed and satellite-derived phenology data. Their study segmented the growing season into phenological stages to enhance prediction accuracy using ML models, with ground-based data slightly outperforming satellite-based observations. The findings suggested an integrating phenology-derived time window to improve yield forecasting for more precise agricultural decision-making.

6.2 Methodology

6.2.1 Proposed framework

The main objective of this study is to create an explainable AI-based structure to interpret features influencing crop yield forecasting. The components of the proposed pipeline is described in Fig. 6.1. This work utilizes two datasets for crop yield prediction. The data is partitioned into training and testing subsets using an 80:20 split. The next step is to evaluate the outputs generated by opaque AI models. However, since these models function as black boxes, explanations for their decisions are necessary. SHAP is employed to provide global explanations. A SHAP global summary plot is used to represent the mean absolute significance of each feature across all input samples. In contrast, a SHAP beeswarm visualization assists in examining the relationships among key features and crop yield predictions. For local explanations, LIME is applied to clarify the reasons behind specific predictions. These explanations increase the illustrate of the AI models and make them more transparent compared to conventional black-box methods. Beyond model explanations, another approach called 'counterfactual explanations' is used for 'what-if' scenarios to understand a model-specific prediction. Thus, the proposed XAI framework offers a complete understanding of global and local factors influencing crop yield predictions.

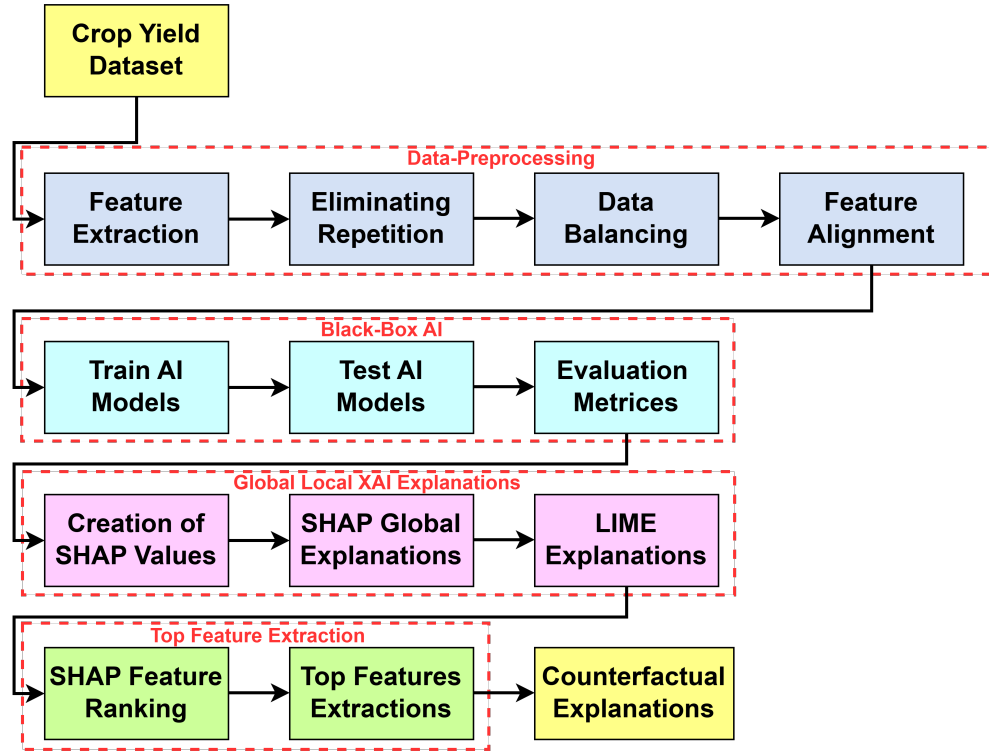


FIGURE 6.1: Proposed XAI framework for crop yield prediction

6.2.2 A novel feature selection method using SHAP

Effective feature selection is an important task for enhancing the performance of AI methods. To address this, a novel XAI-based method for feature extraction, ranking is proposed and their influence on the results of AI models. Proposed SHAP-based feature selection technique is elaborated in Algorithm 7.

Algorithm 7 comprises an adaptive feature attribution mechanism by modifying the traditional SHAP-based feature selection and ranking technique. The key modification is done by including a weight factor based on model accuracy, ensuring that models with higher accuracy exert greater influence on the feature ranking. In the situation of crop yield prediction, domain-specific modifications are done by assigning higher weights to models with better accuracy (using w_j). The proposed algorithm ensures that features are ordered based on their SHAP values and the predictive power of the models in which they are assessed. This modification makes the proposed framework more flexible for feature related challenges in crop yield prediction, where accurate feature attribution is critical for model interpretability and performance. The modified SHAP scores provide a more reliable measure of feature importance, directly contributing to improved prediction models for crop yield.

6.2.3 Dataset description

To check the efficacy of the framework, real crop yield agricultural datasets collected from various authentic sources are used. Descriptions of dataset 1 and 2 are explained in Section 6.2.3.1 and

Algorithm 7 SHAP-based feature selection and ranking algorithm

- 1: Identify the features of the dataset. Let the features are: $F = \{f_1, f_2, \dots, f_m\}$, where m is the total amount of features.
- 2: For each feature f_i , compute SHAP values using n AI models: SHAP_{ij} represents the SHAP value of feature f_i for model M_j .
- 3: Compute the importance score is F_i for each feature f_i :

$$F_i = \sum_{j=1}^n \text{SHAP}_{ij} \cdot \text{Accuracy}_j,$$

where Accuracy_j is the accuracy of the j -th model M_j .

- 4: Rank features as per their importance scores F_i in descendent order.
- 5: For a feature f_i , compute its overall average ranking R_i across all models:

$$R_i = \frac{1}{n} \sum_{j=1}^n \text{Rank}_{ij},$$

where Rank_{ij} is the rank of feature f_i in model M_j based on F_i .

- 6: Assign higher weights to models with better accuracy by modifying F_i :

$$F_i = \sum_{j=1}^n w_j \cdot \text{SHAP}_{ij} \cdot \text{Accuracy}_j,$$

where:

$$w_j = \frac{\text{Accuracy}_j}{\sum_{j=1}^n \text{Accuracy}_j},$$

and w_j is the weight for model M_j .

- 7: Extract the top- k ranked features: $F_k = \{f_{(1)}, f_{(2)}, \dots, f_{(k)}\}$, where k is user-defined or determined through evaluation metrics.
- 8: Train AI models using subsets of the dataset corresponding to F_k for: $+ k \in \{\text{number of features}\}$, and evaluate their performance.

6.2.3.2, respectively.

6.2.3.1 Dataset 1

A publicly available dataset from the FAO [268] and the World Data Bank [269] is used. The dataset comprises a total of 29,689 records. It includes eight features: country names, crop types, planting years (1990–2013), average annual rainfall(mm), pesticide usage (in tons), average temperature(°C), annual crop product (in tons), and crop yield values (in hectograms per hectare).

6.2.3.2 Dataset 2

A publicly available dataset from the Global Program on Enabling Systems Transformation (EST), ICRISAT [270] is used as Dataset 2. The dataset contains 7,740 records from the state of

West Bengal, India. It covers crop types of potatoes, rice, and wheat from the years 1958 to 2022. Dataset contents nine features: actual evapotranspiration(mm) [172], water deficit(mm), maximum temperature($^{\circ}$ C), minimum temperature($^{\circ}$ C), potential evapotranspiration(mm) [172], annual rainfall(mm), total area(ha), total production(tons), yield(kg). During pre-processing, irrelevant columns, such as country names, crop types, and planting years, are not considered in the present work.

6.2.3.3 Dataset Preprocessing

Dataset 1 comprises agricultural and environmental data for various crops cultivated across multiple countries in the world from the year 1990 to 2013. Dataset 2 includes agricultural production and environmental data for crops in India from the year 1958 to 2022. For Dataset 1, duplicate and incomplete rows are removed. Dataset 2 is found without redundancies. The datasets are further balanced using the SMOTE to ensure balanced dataset. Since the datasets include categorical features, one-hot encoding is applied to convert them into numerical values. Subsequently, feature normalization is performed using standard scaling, transforming feature values.

6.2.4 Model selection and prediction

To evaluate the proposed framework, six distinct AI models: Random Forest (RF) [271], Adaptive Boosting (AdaBoost) [272], Categorical Boosting(CatBoost) [272], Light Gradient-Boosting Machine(LGBM) [272], KNN [272], and DT [271] are adopted first. The black-box characteristics of the predictions of the models are further analyzed and explained using the XAI framework.

6.2.5 Model explainability

There are two main categories of model-agnostic interpretation: local explanations with LIME [273] provide insights for individual predictions. Another category is global explanations with SHAP [215] revealing features that influence the behavior of the model. Beyond model explanations, there is another approach called 'counterfactual explanations.' [274] to explore the 'what-if' scenario.

6.2.5.1 Local Interpretable Model Explanation (LIME)

LIME [273] focuses on finding an interpretable model g that is faithful (similar) to the original model f in the local region around the instance x . This is achieved by minimizing a distance function L between the predictions of the two models.

The specific data point x for which an explanation is sought. It lies in the input feature space $\mathbb{R}^d$, where d is the total count of features.

The points x_i sampled in the neighborhood of x . The set of neighboring points is denoted in Eq. 6.1

$$\{x_1, x_2, \dots, x_N\} \quad (6.1)$$

These points are generated by perturbing the features of x , either by adding noise or sampling from the feature distributions.

The distance function $d(x, z)$ measures how similar the perturbed point z is to the original instance x . Common choices for $d(x, z)$ described in Eq. 6.2 and Eq. 6.3

Euclidean Distance:

$$d(x, z) = \sqrt{\sum_{j=1}^d (x_j - z_j)^2} \quad (6.2)$$

Cosine Similarity:

$$d(x, z) = 1 - \frac{\sum_{j=1}^d x_j z_j}{\|x\| \|z\|} \quad (6.3)$$

The weights w_i reflect the proximity of the neighboring point x_i to the original instance x . They are computed using a kernel function $k(x, x_i)$, expressed in Eq. 6.4

$$w_i = \exp\left(-\frac{d(x, x_i)^2}{\sigma^2}\right) \quad (6.4)$$

where: σ is a kernel width parameter controlling the neighborhood size. Points closer to x (smaller $d(x, x_i)$) get higher weights.

The interpretable model g is a function, such as LR or DT, designed to mimic f 's behavior locally. It is parameterized by θ , and its predictions for a point z are denoted as $g(z; \theta)$.

To ensure local fidelity, g must approximate f 's predictions in the neighborhood of x . The discrepancy between f and g is measured using the weighted squared error, shown in Eq. 6.5

$$L(f, g) = \sum_{i=1}^N w_i \cdot [f(x_i) - g(x_i)]^2 \quad (6.5)$$

The weights w_i emphasize points close to x by assigning higher importance based on their distance, shown in Eq. 6.6

$$w_i = \exp\left(-\frac{d(x, x_i)^2}{\sigma^2}\right) \quad (6.6)$$

Substituting w_i into the Eq. 6.5 we got Eq. 6.7

$$L(f, g) = \sum_{i=1}^N \exp\left(-\frac{d(x, x_i)^2}{\sigma^2}\right) \cdot [f(x_i) - g(x_i)]^2 \quad (6.7)$$

To optimize the loss function the parameters θ of the interpretable model g are chosen to minimize $L(f, g)$:

$$\theta^* = \arg \min_{\theta} \sum_{i=1}^N \exp \left(-\frac{d(x, x_i)^2}{\sigma^2} \right) \cdot [f(x_i) - g(x_i; \theta)]^2 \quad (6.8)$$

As an example linear regression, the interpretable model g is represented in Eq. 6.9

$$g(x_i; \theta) = \theta_0 + \sum_{j=1}^d \theta_j x_{i,j} \quad (6.9)$$

where: - θ_0 : Intercept term. - θ_j : Coefficients for feature j .

Substituting $g(x_i; \theta)$ into the Eq. 6.5 we got Eq. 6.10

$$L(f, g) = \sum_{i=1}^N \exp \left(-\frac{d(x, x_i)^2}{\sigma^2} \right) \cdot \left[f(x_i) - \left(\theta_0 + \sum_{j=1}^d \theta_j x_{i,j} \right) \right]^2 \quad (6.10)$$

To minimize $L(f, g)$, compute the gradients with respect to θ_0 and θ_j shown in Eq. 6.11 and Eq. 6.12

Gradient w.r.t. θ_0 :

$$\frac{\partial L}{\partial \theta_0} = -2 \sum_{i=1}^N w_i \cdot [f(x_i) - g(x_i; \theta)] \quad (6.11)$$

Gradient w.r.t. θ_j :

$$\frac{\partial L}{\partial \theta_j} = -2 \sum_{i=1}^N w_i \cdot [f(x_i) - g(x_i; \theta)] \cdot x_{i,j} \quad (6.12)$$

The core idea behind counterfactual [274] explanations is understanding the causal effect of changing specific features in an instance. Counterfactual explanations are represented as a sequence of tensors that demonstrate the minimum changes required in input parameters to achieve a desired outcome. The counterfactual probability can be expressed in Eq. 6.13:

$$P(Y(x') \mid \text{do}(X = x')) \quad (6.13)$$

where $Y(x')$ is treated as the predicted output for a counterfactual instance x' . To make this clearer, it should be explicitly stated that:

- **$Y(x')$:** Represents the mapping from the counterfactual instance x' to its predicted outcome.
- **Input:** Features x' of the instance after an intervention.
- **Output:** Predicted outcome $Y(x')$.

In causal inference, $\text{do}(X = x')$ implies modifying the underlying SCM. An SCM contains a Eq. 6.14:

$$X = f_X(P_X, U_X), \quad Y = f_Y(X, U_Y), \quad (6.14)$$

where, f_X and f_Y : Functional relationships, P_X are the parents of X , and U_X are exogenous factors influencing X . Y is dependent on X , and X acts as the parent of Y in the causal structure or Bayesian Network.

An intervention $\text{do}(X = x')$ removes the natural causal mechanism for X and replaces it with a fixed value x' . The updated model is shown in Eq. 6.15

$$X = x', \quad Y = f_Y(x', U_Y). \quad (6.15)$$

The probability of observing $Y(x')$ given $\text{do}(X = x')$ is computed in this altered model.

To understand $P(Y(x') \mid \text{do}(X = x'))$, the do-calculus rules and Bayesian expansion are used.

Rule of Intervention by definition is shown in Eq. 6.16

$$P(Y(x') \mid \text{do}(X = x')) = P(Y \mid X = x'), \quad (6.16)$$

if there are no unobserved con-founders between X and Y .

Expand the conditional probability using joint distribution, in Eq. 6.17

$$P(Y(x') \mid X = x') = \frac{P(Y(x'), X = x')}{P(X = x')} \quad (6.17)$$

where, $P(Y(x'), X = x')$: Joint probability of $Y(x')$ and $X = x'$, $P(X = x')$: Marginal probability of $X = x'$.

The joint probability $P(Y(x'), X = x')$ can be extracted from the SCM as given:

Decomposition of $P(Y(x'), X = x')$ using the chain rule in Eq. 6.18

$$P(Y(x'), X = x') = P(Y(x') \mid X = x')P(X = x'). \quad (6.18)$$

Replacing $P(Y(x') \mid X = x')$ with the causal model's prediction is described in Eq. 6.19

$$P(Y(x') \mid X = x') = P(Y = f_Y(x', U_Y)) \quad (6.19)$$

This probability is computed by integrating over U_Y as shown in Eq. 6.20

$$P(Y(x') \mid X = x') = \int P(U_Y) \delta(Y - f_Y(x', U_Y)) dU_Y, \quad (6.20)$$

where δ is the Dirac delta function ensuring $Y = f_Y(x', U_Y)$.

The marginal probability $P(X = x')$ is obtained from the prior distribution of X shown in Eq. 6.21:

$$P(X = x') = \int P(U_X) \delta(X - f_X(P_X, U_X)) dU_X. \quad (6.21)$$

Substituting the joint and marginal probabilities back into Eq. 6.17 results in Eq. 6.22:

$$P(Y(x') | X = x') = \frac{\int P(U_Y) \delta(Y - f_Y(x', U_Y)) dU_Y}{\int P(U_X) \delta(X - f_X(P_X, U_X)) dU_X}. \quad (6.22)$$

With no con-founders and X and Y being independent conditional on the intervention, resulting in Eq. 6.23

$$P(Y(x') | \text{do}(X = x')) = \int P(U_Y) \delta(Y - f_Y(x', U_Y)) dU_Y. \quad (6.23)$$

In practice, counterfactual explanations are computed using observed data and predictive models. For ML models:

- Train a model $\hat{Y} = g(X)$.
- Replace X with x' (counterfactual instance) in the Eq. 6.24:

$$Y(x') = g(x'). \quad (6.24)$$

Then the conditional probability becomes in Eq. 6.25

$$P(Y(x') | \text{do}(X = x')) = P(Y = g(x') | X = x'). \quad (6.25)$$

Structure of dependencies and SCM learning The dependencies among features using causal inference techniques, focusing on identifying and quantifying the causal relationships between the variables that influence crop yields is established. This is done by employing a SCM approach, where each feature (e.g., rainfall, fertilizer, temperature) is considered a variable, and the causal influence between them is modeled in a functional relationship. The dependencies among the features are learned using the Peter-Clark algorithm [275]. This algorithm aims to learn the causal structure by identifying conditional dependencies between features. Peter-Clark algorithm tests for conditional independence between variables by performing statistical tests for independence given different sets of variables. It starts with an assumption of no causal relationships between features and iteratively refines the model by adding or removing relationships based on statistical evidence. The Peter-Clark algorithm builds a functional model that reflects the causal dependencies. Once the causal structure is established, the learned SCM is used to generate counterfactual predictions. A counterfactual prediction is made by manipulating one or more features (e.g., changing the amount of rainfall or fertilizer applied) while holding other features constant. Thus it can be observed how these changes impact the predicted crop yield.

6.2.6 Evaluation metrics

MSE [206], RMSE [276], MAE [206], and R^2 [206] are used as evaluation metrics in this work.

6.3 Performance evaluation

This section comprehensively evaluates results on Dataset 1 and Dataset 2. In Section 6.3.1, the results of various black box AI techniques in predicting crop yield is analyzed. In Section 6.3.3, the effect of feature selection techniques on the efficiency of crop yield prediction is examined. In Section 6.3.4, the influence of diverse hyper-parameters on the performance of the AI techniques is evaluated. In Section 6.3.5, the selection technique is compared with different standard feature selection methods.

TABLE 6.1: Result of black-box AI techniques using all features from Dataset 1

Models	Training				Testing			
	MSE	RMSE	MAE	R^2	MSE	RMSE	MAE	R^2
RF	3588.03	59.90	3.08	0.99	19771.1	98.84	6.49	0.98
AdaBoost	40292.13	200.72	19.57	0.947	63250.6	251.49	23.08	0.92
CatBoost	526.38	22.94	3.35	0.99	19401.12	139.28	10.82	0.97
LGBM	2707.86	52.03	5.67	0.99	36210.52	190.29	15.56	0.95
KNN	42721.70	206.69	10.58	0.94	108593.07	329.53	18.67	0.86
DT	9.390	3.06	4.22	1.0	18419.88	135.71	7.88	0.97

bold values indicate best results

6.3.1 Performance of various blackBox AI models

The results of each black-box AI model for crop yield prediction for both Dataset 1 and Dataset 2 is presented here. Tables 6.1 and 6.2 review the key performance metrics of each model on Dataset 1 and 2, respectively. As shown in Table 6.1, DT model demonstrates best performance during training, while the RF model shows best performance during testing with Dataset 1. For Dataset 2, Table 6.2 highlights that CatBoost achieves the highest performance in both the train and test phases.

TABLE 6.2: Result of black-box AI techniques using all features from Dataset 2

Models	Training				Testing			
	MSE	RMSE	MAE	R^2	MSE	RMSE	MAE	R^2
RF	3292.64	57.38	39.01	0.99	26787.02	163.66	114.6	0.94
AdaBoost	72039.97	268.4	216.73	0.86	107026.84	327.14	264.55	0.79
CatBoost	363.76	19.07	14.98	0.999	19337.78	139.06	100.15	0.96
LGBM	477.25	21.84	7.26	0.99	29969.25	21.84	113.33	0.94
KNN	9328.94	96.58	65.98	0.98	31020.39	176.12	170.13	0.94
DT	224.77	16.07	188.71	0.98	33411.14	182.78	126.98	0.93

bold values indicate best results

6.3.2 Performance comparisons with existing XAI frameworks

To assess the effectiveness of the framework, the proposed framework is validated against previously reported XAI frameworks using both datasets. A detailed comparative analysis with existing approaches is described in Table 6.3. The results specify that the present framework outperforms previously reported XAI frameworks.

The quantitative results presented in Table 6.3 demonstrate the robustness of the proposed XAI framework. Compared with existing explainability frameworks, the proposed method achieves superior predictive performance while simultaneously improving model transparency. The SHAP-based feature selection approach reduces the dimensionality of the input space without compromising prediction accuracy. Across both datasets, the framework consistently produces higher R^2 values and lower MAE, RMSE, and MSE values than competing approaches. Furthermore, the integration of global explanations (SHAP summary plots), local explanations (LIME and SHAP waterfall plots), and counterfactual explanations provides a comprehensive interpretability mechanism that is absent in most existing frameworks. These results quantitatively validate the effectiveness, robustness, and practical applicability of the proposed XAI framework.

TABLE 6.3: Performance comparisons of proposed XAI-framework with existing frameworks

Dataset 1				
	Mohan <i>et. al</i> [252]	Shams <i>et. al.</i> [231]	Martin <i>et. al.</i> [277]	Proposed
MSE	36210.52	71100.02	63250.6	19771.1
RMSE	190.29	266.65	140.28	98.84
MAE	15.56	201.23	10.81	6.49
R^2	0.95	0.87	0.96	0.98
Dataset 2				
	Mohan <i>et. al</i> [252]	Shams <i>et. al.</i> [231]	Martin <i>et. al.</i> [277]	Proposed
MSE	29969.25	31449.02	33411.14	19337.78
RMSE	21.84	161.9	182.78	139.06
MAE	113.33	115.19	126.98	100.151
R^2	0.94	0.72	0.93	0.96

bold values indicate best results

6.3.3 Effects on various combinations of features

Each black-box AI model is trained using different sets of top-k ranked features from Dataset 1 and Dataset 2 to analyze the impact of individual features. Parameters are carefully chosen using the proposed SHAP-based feature selection technique. Here, K represents the count of top-ranked parameters used as input for the regression models. Various regressors are then tested with these various values of K , and their performance are recorded to measure the effects of feature selection on model result.

Table 6.4 demonstrates that most of the models achieve optimal performance while utilizing all five features of Dataset 1. Notably, RF, AdaBoost, CatBoost, and LGBM require the top five

TABLE 6.4: Performance of AI models on Dataset 1 using top selected features through proposed SHAP-based feature selection and ranking method

		SHAP- RF	SHAP- AdaBoost	SHAP- CatBoost	SHAP- LGBM	SHAP- KNN	SHAP- DT
K=1	MSE	10882.42	80443.12	123315.02	73109.02	104095.02	50385.12
	RMSE	104.31	283.62	351.16	270.39	322.63	224.47
	MAE	9.48	17.66	263.44	204.97	242.9	153.34
	R^2	0.86	0.90	0.77	0.86	0.81	0.91
K=2	MSE	10426.81	105055.62	157335.23	78202.02	131235.02	71570.02
	RMSE	102.11	324.11	396.65	279.65	362.25	267.53
	MAE	10.42	242.2	308.38	214.69	285.9	192.14
	R^2	0.89	0.81	0.71	0.86	0.76	0.87
K=3	MSE	10246.11	120945.6	71100.02	106045.02	50926.02	36542.61
	RMSE	101.22	347.76	266.65	325.64	225.67	191.16
	MAE	9.18	267.61	201.23	251.55	163.21	131.1
	R^2	0.91	0.78	0.87	0.80	0.91	0.93
K=4	MSE	9922.18	141505.02	70667.12	109525.02	51421.02	12426.02
	RMSE	99.61	376.16	265.83	330.94	226.76	111.47
	MAE	6.84	264.58	199.53	247.63	165.78	76.92
	R^2	0.94	0.74	0.87	0.80	0.90	0.98
K=5	MSE	19771.1	63250.6	19401.12	36210.52	108593.07	18419.88
	RMSE	98.84	251.49	139.28	190.29	329.53	135.71
	MAE	6.49	23.08	10.82	15.65	18.67	7.88
	R^2	0.98	0.92	0.97	0.95	0.86	0.97

bold values indicate the best results

features ($K = 5$) to get the best result. KNN and DT models provide the best prediction with top three and four features, respectively. This reveals the potential of top features in enhancing prediction performance. Additionally, the full feature ranking derived from the novel SHAP-based ranking method for Dataset 1 is presented in Table 6.5. This ranking is calculated by averaging the SHAP values of each feature across all AI models.

TABLE 6.5: Feature ranking for Dataset 1

Feature	RF	AdaBoost	CatBoost	LGBM	KNN	DT	Avg.	Rank
Area	3	2	2	2	3	3	2.5	2
Production	1	1	1	1	1	1	1	1
Ann. Rain	5	5	5	5	4	5	4.83	5
Fertilizer	4	4	3	4	2	4	3.5	4
Pesticide	2	3	4	3	5	2	3.16	3

Table 6.7 demonstrates that RF, KNN and DT models achieve optimal performance by utilizing all eight features ($K = 8$) on Dataset 2. On the other hand, AdaBoost, CatBoost, and LGBM models provide the best performance with the top three, seven and three features, respectively. Additionally, the full feature ranking derived from the novel SHAP-based ranking method for Dataset 2 is presented in Table 6.6. This ranking is calculated by averaging the SHAP values of each parameter across all AI models.

TABLE 6.6: Feature ranking for Dataset 2

Feature	RF	AdaBoost	CatBoost	LGBM	KNN	DT	Avg.	Rank
Evapo_act	7	7	8	6	6	7	6.8	8
Water_def	8	8	7	7	4	6	6.6	7
Temp_Max	5	4	3	5	7	5	4.8	4
Temp_Min	4	6	5	4	8	3	5	5
Evapo_pot	3	3	4	3	5	4	3.6	3
Rainfall	6	5	6	8	3	8	6	6
Area	2	2	2	2	2	2	2	2
Production	1	1	1	1	1	1	1	1

TABLE 6.7: Performance of AI models on Dataset 2 using top selected features through proposed SHAP-based feature selection and ranking method.

		SHAP- RF	SHAP- AdaBoost	SHAP- CatBoost	SHAP- LGBM	SHAP- KNN	SHAP- DT
K=1	MSE	46381.12	111935.6	80470.02	3923.4	29334.06	32952.06
	RMSE	215.36	334.56	283.67	62.63	171.27	181.53
	MAE	155.57	258.97	218.07	35.32	104.74	125.34
	R^2	0.91	0.79	0.85	0.97	0.76	0.69
K=2	MSE	120915.02	34761.31	8738.3	7543.1	26715.02	21807.06
	RMSE	347.71	186.44	93.47	86.65	163.45	147.67
	MAE	10.42	242.2	308.38	214.69	285.9	192.14
	R^2	0.78	0.69	0.98	0.98	0.71	0.70
K=3	MSE	48133.42	28340.02	9351.3	3360.41	28627.42	21196.02
	RMSE	219.39	168.35	96.711	57.96	169.19	145.59
	MAE	161.07	93.57	35.749	38.68	109.29	94.93
	R^2	0.91	0.95	0.98	0.99	0.77	0.71
K=4	MSE	52058.61	109375.03	22184.02	30136.02	30520.02	22836.07
	RMSE	228.16	330.72	148.94	173.6	174.7	151.12
	MAE	155.16	248.11	81.08	125.79	123.85	102.64
	R^2	0.90	0.80	0.96	0.84	0.74	0.78
K=5	MSE	72641.02	31544.31	10640.18	26212.02	22651.02	22377.01
	RMSE	269.52	177.61	103.15	161.9	150.5	149.59
	MAE	193.46	118.03	39.03	115.19	104.43	100.29
	R^2	0.87	0.94	0.98	0.72	0.78	0.69
K=6	MSE	155285.03	93539.02	21744.03	31449.02	27389.02	21385.02
	RMSE	394.05	305.84	147.46	177.34	165.34	149.23
	MAE	302.44	231.2	84.75	125.6	108.26	97.48
	R^2	0.71	0.83	0.96	0.72	0.70	0.71
K=7	MSE	115995.03	42719.42	8941.4	26673.01	24903.02	21761.04
	RMSE	340.57	206.69	94.55	163.32	157.81	147.51
	MAE	250.68	150.45	58.13	115.88	104.07	89.34
	R^2	0.78	0.92	0.98	0.71	0.74	0.70
K=8	MSE	26778.02	107026.84	19337.78	29969.25	31020.39	33411.14
	RMSE	163.66	327.14	139.06	21.84	176.12	182.78
	MAE	114.6	264.55	100.151	113.33	170.13	126.98
	R^2	0.949	0.798	0.963	0.943	0.941	0.937

bold values indicate best results

6.3.4 Effects on tuning various hyper-parameters

Fine-tuning hyper-parameters often enhances the performance of AI models by regulating the learning process. Table 6.8 illustrates the impact of hyper-parameter tuning across different models. For RF, AdaBoost, CatBoost, LGBM, and DT models, the max_depth parameter is initially optimized. The max_depth is a critical hyper-parameter influencing over-fitting and

model complexity. It defines the maximum depth, a DT can grow. The models show an average improvement of $\approx 39.62\%$ in the MSE with a `max_depth` of 100 compared to 50. However, increasing `max_depth` to 200 yielded no significant improvement and led to a slight decline in performance, indicating `max_depth` of 100 as the optimum value for the better performance. The `min_samples_leaf` parameter requires the minimal number of samples needed at a leaf node. Better performance of the model's RF and DT is obtained with `min_samples_leaf` set of 5 compared to 2, similarly, for the `min_samples_split` parameter determines the minimal number of samples required for internal node. The best results are achieved with `min_samples_split` parameter value of 20. For ensemble methods like AdaBoost, LGBM, KNN, and RF, the `n_estimators` representing the total count of estimators is tuned. Results are evaluated for `n_estimators` set of 51, 101, and 151, and the best performance is obtained at `n_estimators` of 150. Additionally, for AdaBoost, CatBoost, and LGBM, the `learning_rate` parameter is optimized. For the case of AdaBoost, the best result is achieved at learning rates of 0.05, while for CatBoost and LGBM, learning rates of 0.1 provide the highest performance. In the case of KNN, the `n_neighbors` parameter gives the number of neighbouring samples for predictions. The optimization of the `n_neighbors` parameter provides the good performance with `n_neighbors` set to 10 while increasing `n_neighbors` set to 15 causes performance reduction. The `leaf_size` parameter of KNN specifies the threshold at which nodes stop splitting. The optimal value of `leaf_size` is found to be 10. Hyper-parameter tuning also significantly impacts the evaluation metrics for Dataset 2, as detailed in Table 6.9.

6.3.5 Comparison with various benchmark feature selection techniques across various models

To check the effect of the number of features selected using the proposed feature selection methods, we replicate the experimental setup to analyze the top-K features across five baseline feature selection method is replicated. These methods include Select K-Best [278], Recursive Feature Elimination (RFE) [279], Forward Selection (FS) [280], Lasso [281], and Principal Component Analysis (PCA) [282]. Evaluation metrics are generated for all six AI models across two datasets for different values of K with top-K features. Table 6.10 presents the comparative results of the six AI models for $K = 5$. The results demonstrate that the novel feature selection methods outperform the benchmark feature selection techniques across all AI models except KNN [283]. These findings highlight the overall superiority of our approach in most cases.

Table 6.11 presents the results comparison for $K = 8$, showcasing the results of six black-box AI methods using Select K-Best, RFE, FS, Lasso, and PCA as benchmark feature selection techniques. The results indicate that our proposed feature selection techniques outperform the benchmark methods across all AI models except for Adaboost.

6.3.6 Quantitative Validation of the Proposed XAI Framework

The proposed framework was quantitatively validated using two independent crop yield datasets. The validation considered three aspects: prediction performance, feature selection effectiveness, and interpretability support. The proposed SHAP-based feature selection method reduced the

TABLE 6.8: Effect of hyper-parameter tuning for the AI models on dataset 1

RF									
	n_estimator			max_depth			min_samp_leaf		
	50	100	150	50	100	200	2	5	10
MSE	11720.4	977101	9824.2	14782.6	10784.4	18692.2	11720.4	9924.3	9924.3
RMSE	108.25	98.84	99.11	121.58	103.84	136.7	108.25	99.62	99.62
MAE	7.10	6.49	6.50	7.97	6.81	8.96	7.10	6.52	6.52
R ²	0.93	0.98	0.98	0.82	0.95	0.79	0.93	0.97	0.97
AdaBoost									
	n_estimator			learning_rate			max_depth		
	50	100	150	0.01	0.05	0.1	50	100	200
MSE	120995.7	9842.1	9782.1	130924.6	10421.2	10824.3	118690.4	18419.8	123410.4
RMSE	347.84	99.20	98.90	361.83	102.08	104.03	344052	135.71	351.3
MAE	267.67	6.94	6.92	278.43	7.14	7.27	265.86	7.88	167.25
R ²	0.78	0.98	0.98	0.75	0.96	0.96	0.78	0.97	0.77
CatBoost									
	max_depth			learning_rate					
	50	100	200	0.01	0.05	0.1			
MSE	24611.33	19401.12	19612.31	62862.42	22721.32	19264.22			
RMSE	156.87	139.28	140.04	163.89	150.73	138.79			
MAE	12.18	10.82	10.87	12.72	11.69	10.77			
R ²	0.84	0.97	0.97	0.80	0.86	0.98			
LGBM									
	n_estimator			max_depth			learning_rate		
	50	100	150	50	100	200	0.01	0.05	0.1
MSE	72426.31	40226.81	36420.62	52416.3	39721.14	40783.4	38147.22	42163.21	32620.43
RMSE	269.12	200.56	190.84	228.9	199.30	201.9	195.31	205.33	180.61
MAE	22.11	16.48	15.69	18.81	16.38	16.59	16.02	16.84	14.84
R ²	0.69	0.91	0.95	0.87	0.92	0.91	0.91	0.90	0.96
KNN									
	n_neighbors			leaf_size					
	3	10	15	10	50				
MSE	62387.44	51284.32	60421.3	50926.02	68823.41				
RMSE	249.77	226.46	245.80	225.67	262.34				
MAE	180.63	163.78	177.7	163.21	189.73				
R ²	0.79	0.92	0.82	0.91	0.74				
DT									
	max_depth			min_sample_split			min_sample_leaf		
	50	100	200	5	10	20	2	5	10
MSE	21248.12	12486.02	12486.02	27241.61	3834.11	13283.42	21734.4	14834.8	14732.9
RMSE	145.76	111.74	111.74	165.05	117.61	115.25	147.42	121.79	121.37
MAE	99.68	76.42	76.42	112.87	80.43	78.82	100.82	83.29	83.00
R ²	0.77	0.97	0.97	0.75	0.93	0.94	0.76	0.89	0.88

bold values indicate best results

number of input features while maintaining or improving predictive performance. Additionally, the framework consistently outperformed benchmark feature selection techniques such as PCA, RFE, and traditional filter methods across multiple machine learning models. These results indicate that the proposed framework provides both accurate predictions and reliable explanations, thereby enhancing trustworthiness in AI-assisted agricultural decision-making.

TABLE 6.9: Effect of hyper-parameter tuning for the AI models on dataset 2

RF									
	n_estimator			max_depth			min_samp_leaf		
	50	100	150	05	100	200	2	5	10
MSE	32836.12	26778.02	26778.02	47268.12	27224.2	27241.2	31436.5	28215.42	28621.57
RMSE	181.20	163.66	163.66	217.41	164.9	165.04	177.3	167.9	169.17
MAE	126.88	114.6	114.6	152.14	115.4	115.5	124.08	117.5	118.3
R ²	0.89	0.94	0.94	0.78	0.93	0.93	0.89	0.92	0.92
AdaBoost									
	n_estimator			learning_rate			max_depth		
	50	100	150	0.01	0.05	0.1	50	100	200
MSE	36464.03	36286.4	28720.5	32624.05	28340.02	28340.02	112660.4	48452.6	66878
RMSE	190.95	190.48	169.4	180.82	168.35	168.35	335.64	220.12	258.61
MAE	106.3	106.04	94.31	100.39	93.57	93.57	258.84	160.98	188.61
R ²	0.87	0.88	0.93	0.91	0.95	0.95	0.79	0.91	0.88
CatBoost									
	max_depth			learning_rate					
	50	100	200	0.01	0.05	0.1			
MSE	16282.62	12461.25	8941.4	12446.4	12446.4	9421.6			
RMSE	127.60	111.62	94.55	111.56	111.56	97.06			
MAE	78.44	68.62	58.13	68.58	68.58	59.67			
R ²	0.84	0.90	0.98	0.90	0.90	0.96			
LGBM									
	n_estimator			max_depth			learning_rate		
	50	100	150	50	100	200	0.01	0.05	0.1
MSE	8641.02	3360.41	4682.2	9626.6	3360.41	7682.1	4267.4	4267.4	5664.1
RMSE	92.95	57.96	68.42	98.11	57.96	87.6	65.32	65.32	75.26
MAE	62.03	38.68	45.66	65.47	38.68	58.49	43.61	43.61	50.24
R ²	0.79	0.99	0.95	0.69	0.99	0.74	0.96	0.96	0.88
KNN									
	n_neighbors			leaf_size					
	3	10	15	10	50				
MSE	46421.13	33124.62	33240.5	31020.39	46427.54				
RMSE	215.4	182	182.3	176.12	215.4				
MAE	208.12	175.81	176.1	170.13	207.15				
R ²	0.83	0.92	0.92	0.94	0.83				
DT									
	max_depth			min_sample_split			min_sample_leaf		
	50	100	200	5	10	20	2	5	10
MSE	44421.27	38761.4	38625.3	46421.56	33411.14	38426.76	71426.3	34761.32	52614.8
RMSE	210.76	196.8	196.5	215.45	182.78	196.02	267.25	186.44	229.37
MAE	164.42	136.7	136.5	149.68	126.98	136.18	185.66	129.5	159.35
R ²	0.85	0.89	0.89	0.84	0.93	0.89	0.71	0.92	0.81

bold values indicate best results

6.4 The significance of features through XAI

The key features that influence the decision-making process of each AI model across the two datasets are now emphasized.

TABLE 6.10: Performance comparisons of proposed Shap-based feature selection with benchmark feature selection technique using all features from Dataset 1

Models	SHAP				Select K-best			
	MSE	RMSE	MAE	R^2	MSE	RMSE	MAE	R^2
RF	19771.1	98.84	6.49	0.98	118690.4	344.52	265.86	0.78
AdaBoost	63250.6	251.49	23.08	0.92	78373	279.95	210.62	0.85
CatBoost	19401.12	139.28	10.82	0.97	113640.5	362.82	262.7	0.76
LGBM	36210.52	190.29	15.56	0.95	77553	278.48	211.21	0.86
KNN	108593.07	329.53	18.68	0.86	53812	231.97	171.37	0.90
DT	18419.88	135.71	7.88	0.97	123410.4	351.3	267.25	0.77
Models	RFE				FS			
	MSE	RMSE	MAE	R^2	MSE	RMSE	MAE	R^2
RF	215940.2	464.5	366.29	0.60	195530.6	442.19	350.96	0.64
AdaBoost	235110.6	484.89	391.07	0.56	246250.7	496.24	393.42	0.54
CatBoost	219980.7	469.02	375.08	0.59	246430.2	496.42	392.72	0.54
LGBM	198060.4	445.04	354.87	0.63	227850.4	477.33	372.67	0.58
KNN	195070.7	441.67	350.57	0.64	442130.4	664.93	462.47	0.18
DT	189760.4	435.61	344.25	0.65	227850.4	477.33	372.67	0.58
Models	Lasso				PCA			
	MSE	RMSE	MAE	R^2	MSE	RMSE	MAE	R^2
RF	26.723.2	163.47	103.57	0.58	18247.2	135.08	89.74	0.66
AdaBoost	25773.4	160.54	106.22	0.53	24266	155.78	105.96	0.55
CatBoost	3106.2	176.24	125.53	0.43	314444	177.32	123.68	0.42
LGBM	24724	157.24	111.17	0.55	21917	148.05	95.818	0.60
KNN	27148.1	164.77	114.7	0.50	21206	145.62	97.22	0.61
DT	24266	155.78	105.96	0.55	16845	129.79	89.14	0.69

bold values indicate best results

6.4.1 Global summary plot (GSP)

The GSP for each AI model, highlighting the significant features influencing their decision-making processes is presented here. Fig. 6.2 and Fig. 6.3 illustrate the feature importance of the various AI models applied to Dataset 1 and 2, respectively. Fig. 6.2 (a-f) represents that for all AI methods, 'Production' is the most influential feature for Dataset 1. For Dataset 2, 'Production_tons' plays the most crucial role in prediction for all models, as shown in Fig. 6.3 (a-f). The importance of various features in effective crop yield prediction is represented by the bar length along the 'Mean Absolute SHAP Value' axis in Fig. 6.2 and Fig. 6.3.

6.4.2 Beeswarm summary plot (BSP)

In this section the BSP for each AI methods is presented. Fig. 6.4 and Fig. 6.5 show the BSP for Dataset 1 and Dataset 2, respectively. Each data point in the dataset is represented as a dot in a horizontal direction according to the SHAP value, indicating its impact on the prediction. Points with higher SHAP value are positioned further from the center, and densely packed points are vertically stacked. The color bar on the right side of the plots shows the raw value of each instance, where red and blue color represent high and low feature values, respectively. Like the GSP, the BSP also highlights the importance of features. Notably, the RF and AdaBoost models rely heavily on a single feature for decision-making, whereas, for the other models, multiple features significantly contribute to crop yield prediction.

TABLE 6.11: Performance comparisons of proposed Shap-based feature selection with benchmark feature selection technique using all features from Dataset 2

Models	SHAP				Select K-best			
	MSE	RMSE	MAE	R^2	MSE	RMSE	MAE	R^2
RF	26787.02	163.66	114.6	0.949	112660.4	335.64	258.84	0.79
AdaBoost	107026.84	327.14	264.55	0.798	121320.3	348.32	254.81	0.78
CatBoost	19337.78	139.06	100.151	0.963	47873.0	218.8	157.22	0.91
LGBM	29969.25	21.84	113.33	0.943	48452.6	220.12	160.98	0.91
KNN	31020.39	176.12	170.13	0.941	48368.2	219.93	151.18	0.91
DT	33411.14	182.78	126.98	0.937	66878	258.61	188.61	0.88
Models	RFE				FS			
	MSE	RMSE	MAE	R^2	MSE	RMSE	MAE	R^2
RF	301550.3	549.13	446.95	0.44	269090.2	518.74	418.54	0.50
AdaBoost	235110.4	484.89	391.07	0.56	263660.4	513.47	411.01	0.51
CatBoost	219980.2	469.02	375.04	0.59	242220.1	492.16	393.93	0.55
LGBM	198060.4	445.04	354.87	0.63	234070.1	483.81	385.77	0.57
KNN	195070.3	441.67	350.57	0.64	230220.4	479.82	378.78	0.57
DT	189760.2	435.61	345.25	0.65	235480.5	485.26	386.14	0.56
Models	Lasso				PCA			
	MSE	RMSE	MAE	R^2	MSE	RMSE	MAE	R^2
RF	360950.2	600.79	473.18	0.33	360530.2	600.44	496.17	0.33
AdaBoost	377660.4	614.54	489.03	0.30	372040.5	609.95	434.02	0.31
CatBoost	29763.4	545.53	421.29	0.45	363740.2	603.11	491.82	0.33
LGBM	255650.3	505.62	411.45	0.53	427300.4	653.68	445.97	0.21
KNN	301550.2	549.13	446.95	0.44	327340.2	572.14	430.44	0.39
DT	215940.3	464.7	366.29	0.60	325050.3	570.14	433.06	0.40

bold values indicate best results

6.5 Comparing XAI models with black-box AI methods

The clarifications for the accurate and inaccurate predictions in Dataset 1 and Dataset 2 are presented here. Specifically, the black-box model clarifications with the clarifications generated by the present XAI framework are compared. Best performance in terms of evaluation metrics for Dataset 1 and Dataset 2 is obtained for RF and CatBoost models, respectively, as represented in Section 6.3.1

6.5.1 Dataset 1

6.5.1.1 Accurate prediction

The LIME local plot in Fig. 6.6(a), generated using LIME, shows that ‘Production’ has the most significant negative contribution, while features like ‘Pesticide’, ‘Annual Rainfall’, ‘Area’, and ‘Fertilizer’ have minor positive or negative influences. SHAP waterfall plot in Fig. 6.6(b), generated using SHAP, gives a more detailed results of feature contributions as compared to LIME. In Fig. 6.6(a, b) x-axis shows the various features and y-axis shows the cumulative impacts of each features on the final prediction. It highlights ‘Production’ as the dominant factor with a negative SHAP value (-129.8), while features such as ‘Annual Rainfall’ (-8.53) and ‘Fertilizer’ (-1.31) provide a negative impact in the accurate prediction. Conversely, ‘Area’ (+1.81) and ‘Pesticide’ (+1.07) have minor positive effects on prediction.

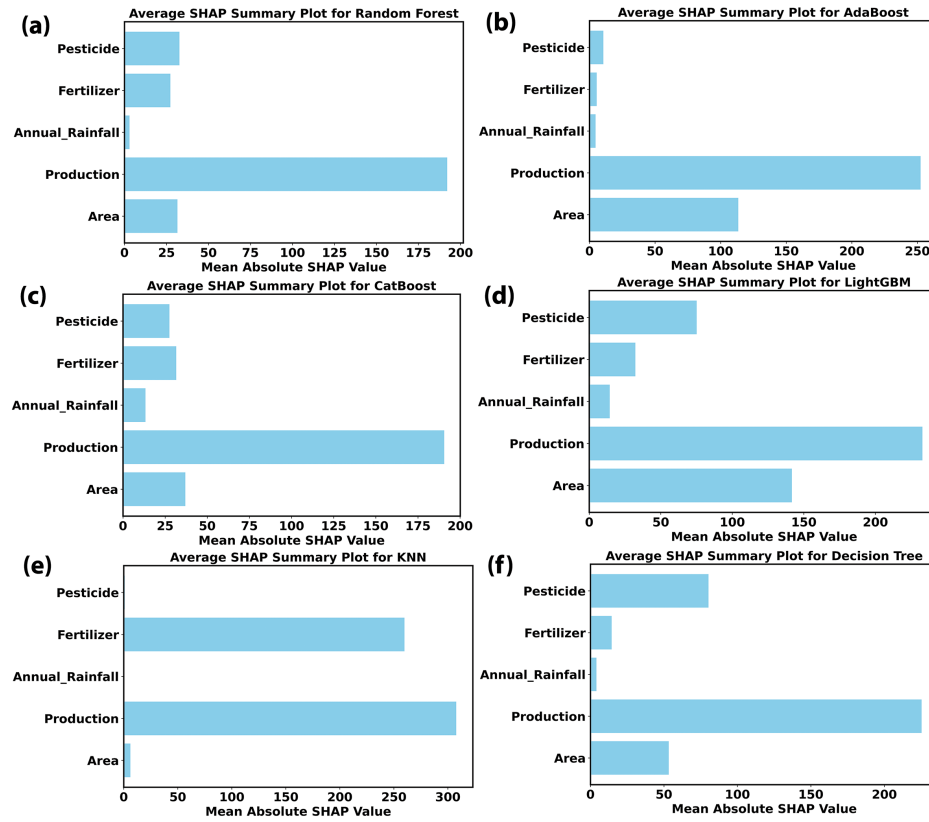


FIGURE 6.2: Global summary plot of feature importance for each AI models (generated Using SHAP) for dataset 1: (a) RF, (b) AdaBoost, (c) CatBoost, (d) LGBM, (e) KNN, (f)DT

6.5.1.2 Inaccurate prediction

LIME local plot in Fig. 6.6(c) highlights ‘Production’ as the most influential negative contributor with a smaller impact of ‘Area’, ‘Fertilizer’, ‘Pesticide’, and ‘Annual_Rainfall’. Similarly, the SHAP waterfall plot in Fig. 6.6(d) reveals ‘Production’ as the dominant factor with a substantial negative SHAP value (-143.22). However, it also shows positive contributions from ‘Area’ (+11.09), ‘Pesticide’ (+5.46), and ‘Fertilizer’ (+4.31), which skew the prediction upwards, while ‘Annual_Rainfall’ (-6.58) adds a minor negative influence.

6.5.2 Dataset 2

6.5.2.1 Accurate prediction

The LIME local plot in Fig. 6.7(a) highlights that ‘Production_tons’ and ‘Area_ha’ are the dominant negative contributors, followed by other features like ‘Evapo_Poten_mm’ and ‘Water_Deficit_mm’. Positive contributions from ‘Rainfall_mm’ and ‘Evapo_Actual_mm’ features balance the prediction. SHAP waterfall plot in Fig. 6.7(b) provides a more detailed breakdown of these contributions. It shows that ‘Production_tons’ (-876.32), ‘Evapo_Poten_mm’ (-2.133),

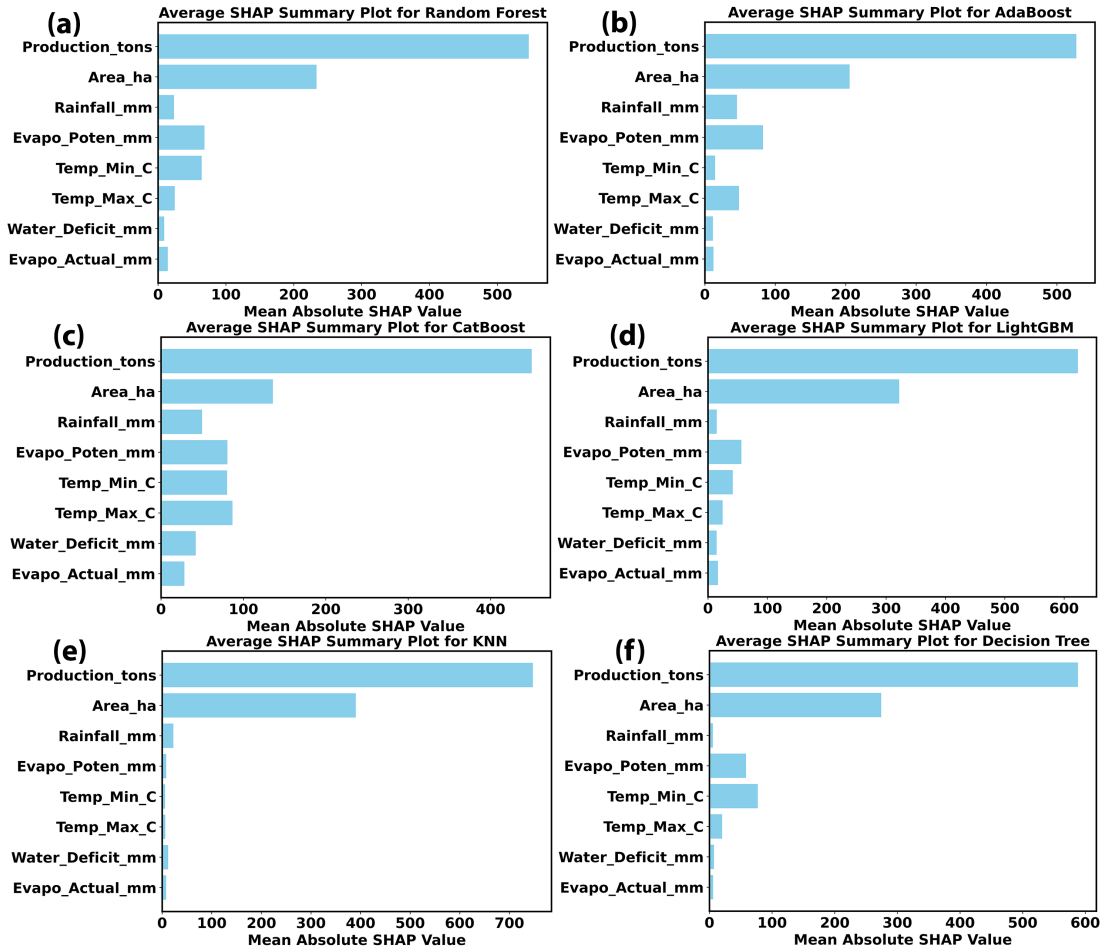


FIGURE 6.3: Global Summary Plot of Feature Importance for each AI models (Generated Using SHAP) for dataset 2: (a) RF, (b) AdaBoost, (c) CatBoost, (d) LightGBM, (e) KNN, (f)DT

and ‘Area_ha’ (-1.403) provide the primary negative influence, while positive contributions from ‘Rainfall_mm’ (+59.22) and ‘Evapo_Actual_mm’ (+37.94) add smaller impacts.

6.5.2.2 Inaccurate prediction

LIME local plot in Fig. 6.7(c) identifies ‘Production_tons’ and ‘Area_ha’ as the most significant positive contributors, which lead to the overestimation, while smaller negative contributions come from features like ‘Water_Deficit_mm’ and ‘Temp_Max_C’. SHAP waterfall plot in Fig. 6.7(d) further details this imbalance, showing ‘Production_tons’ (+672.79) and ‘Area_ha’ (+2.567) as dominant positive contributors, while features like ‘Evapo_Poten_mm’ (-113.2) and ‘Temp_Max_C’ (-36.77) have smaller negative effects.

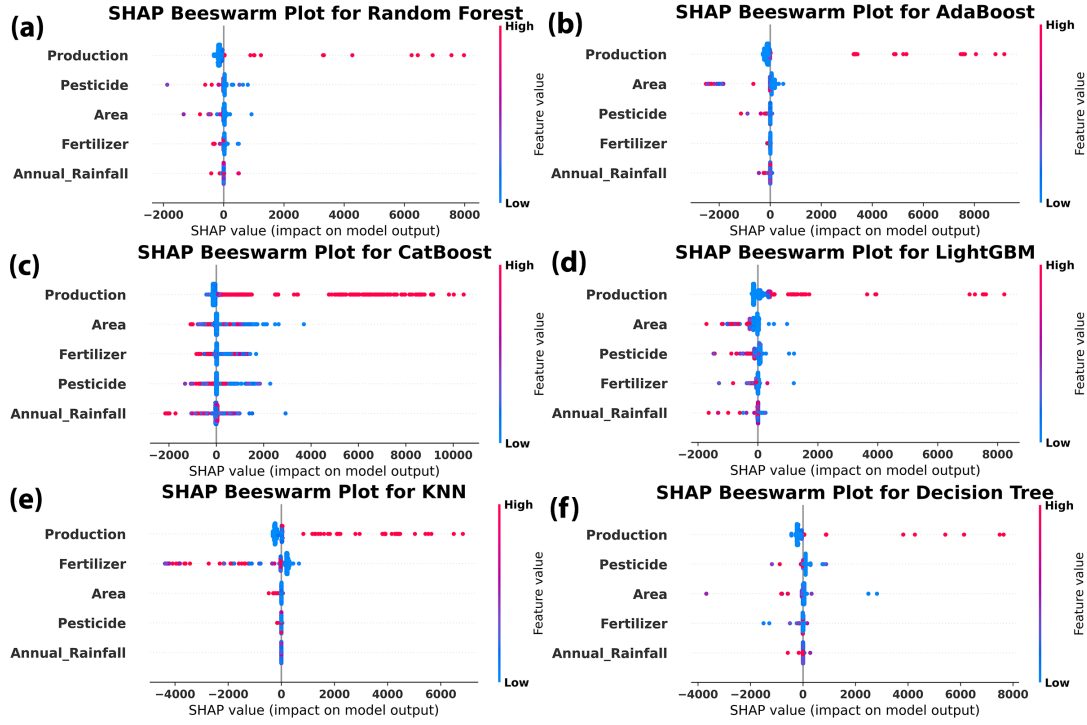


FIGURE 6.4: Beeswarm summary plot of feature importance for each AI models (generated using SHAP) for dataset 1: (a) RF, (b) AdaBoost, (c) CatBoost, (d) LGBM, (e) KNN, (f) DT

6.6 Counterfactual analysis of influential features for prediction adjustment

Counterfactual analysis investigate the ‘what-if’ scenarios by modifying specific input features and observing how the changes impact the model’s predictions. Fig. 6.8(a) shows the actual and predicted value of crop yield to be 75.31 and 70.13, respectively, where the target yield is 76.51 for Dataset 1. The counterfactual explanation in Fig. 6.8(a) suggests significant changes in the ‘Production’ and ‘Fertilizer’ features to achieve the target prediction. The counterfactual values of ‘Production’ and ‘Fertilizer’ being higher than the original values. This indicates an important role in the prediction. A higher ‘Production’ value in counterfactual analysis agrees with SHAP analysis, confirming its significant role in crop yield forecasting. Other features such as ‘Area’, ‘Annual_Rainfall’, and ‘Pesticide’ remain fixed to the original value.

Fig. 6.8(b) represents the exact and forecasted yield of 3092.52 and 2763.59, respectively, with a target yield of 4092.52 for Dataset 2. The counterfactual explanation highlights ‘Production_ton’ and ‘Area_ha’ as the most important features in reaching the target yield. The counterfactual values indicate the contribution of ‘Rainfall_mm’ and ‘Area_ha’ need to increase, while the ‘Production_tons’ feature is required to be reduced than the original values for achieving the target prediction. Other features, such as ‘Evapo_Actual_mm’, ‘Water_Deficit_mm’, ‘Temp_Max_C’, ‘Temp_Min_C’ and ‘Evapo_Poten_mm’ remain fixed to the original value.

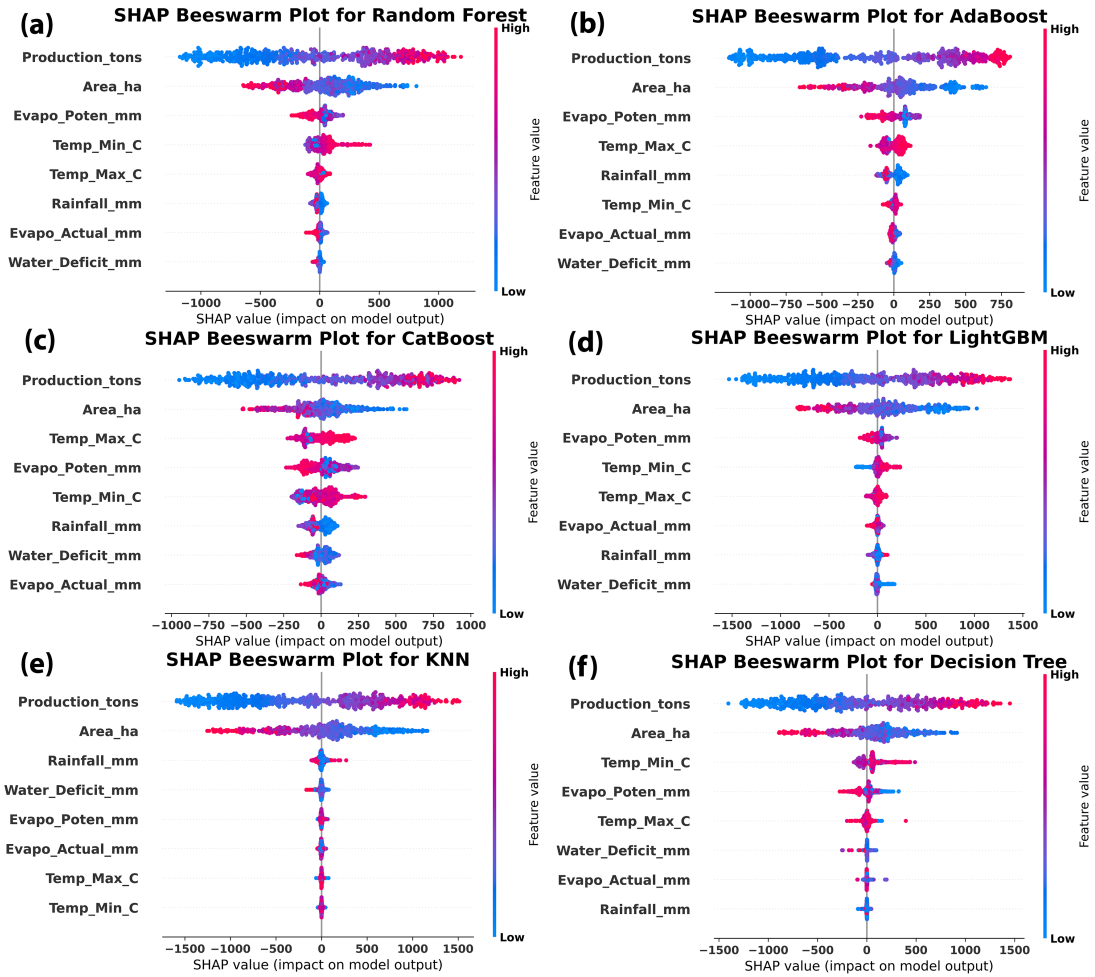


FIGURE 6.5: Beeswarm Summary Plot of Feature Importance for each AI models (Generated Using SHAP) for dataset 2: (a) RF, (b) AdaBoost, (c) CatBoost, (d) LightGBM, (e) KNN, (f) DT

6.7 Chapter Summary

This present work addresses critical gaps in AI applications in agriculture by developing an XAI framework to be used for crop yield forecasting. The framework increases the interpretability of AI models by providing global and local explanations, as well as counterfactual scenarios, for a detailed understanding of the decision-making process. The framework shows outstanding result as compared to previously reported XAI frameworks to date. Moreover, a novel XAI-based feature selection method is proposed here for the identification, ranking, and evaluation of significant features. This proposed feature selection method surpasses existing benchmark feature selection methods. An in-depth sensitivity analysis is done to evaluate the effects of each features and hyper-parameter tuning for model performance. By bridging the gaps of interpretability, feature selection, and sensitivity analysis, the work presented in this chapter is aimed at achieving better and advanced AI applications in agriculture. The proposed XAI framework paves the way for transparent and efficient precision agriculture in real-field applications by enhancing decision-making and improving model interpretability.

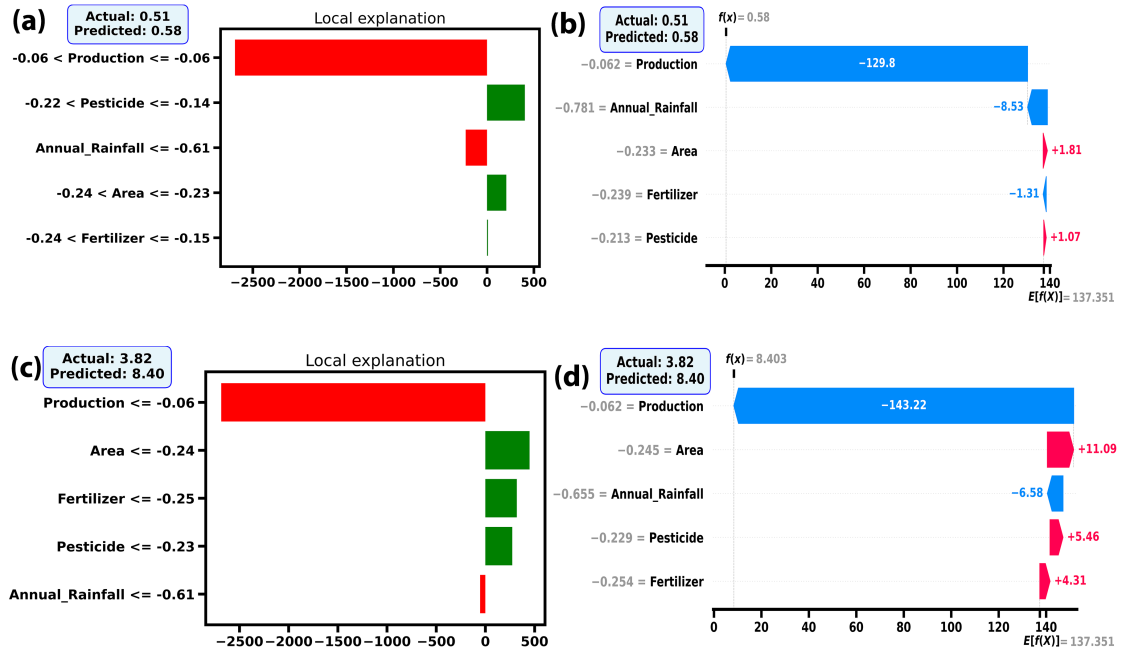


FIGURE 6.6: LIME local and waterfall Plot for local explanations on dataset 1: (a, b) Accurate Prediction, (c, d) Inaccurate Prediction

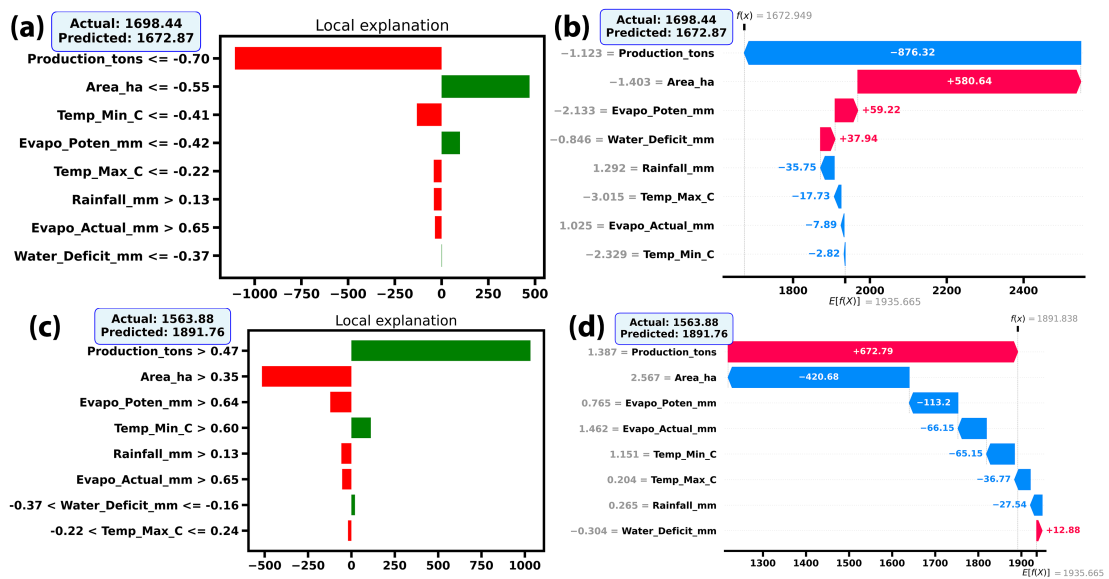


FIGURE 6.7: LIME local and waterfall Plot for local explanations on Dataset 2: (a, b) Accurate Prediction, (c, d) Inaccurate Prediction

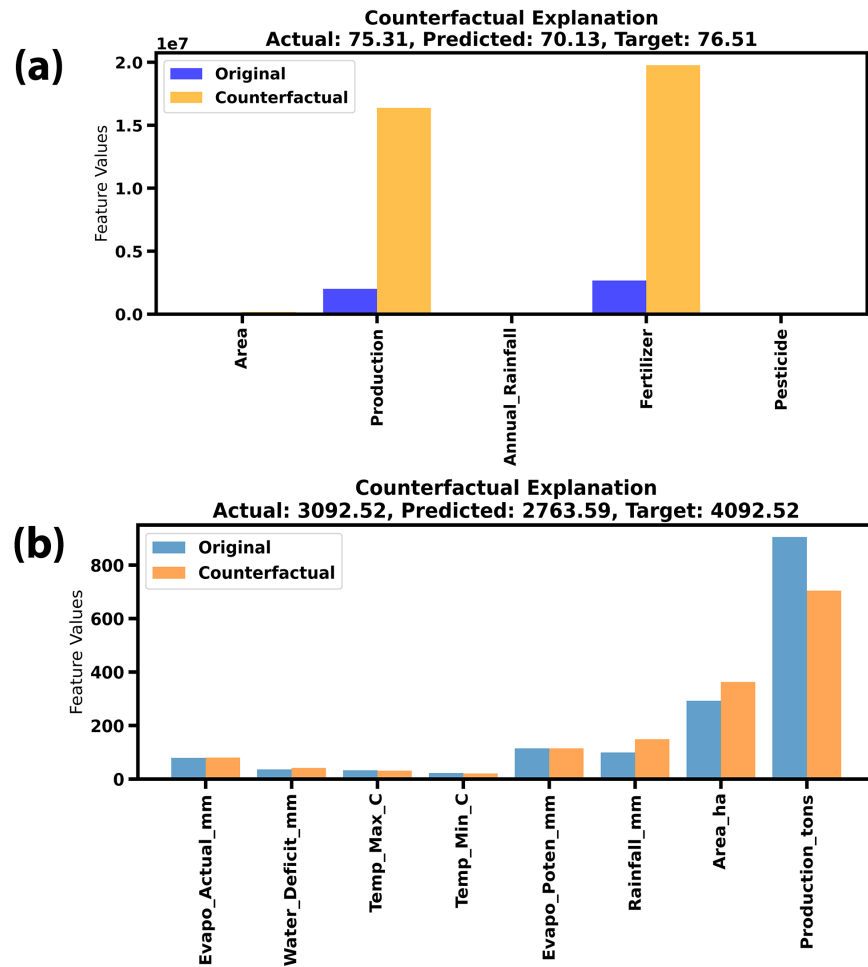


FIGURE 6.8: (a) Counterfactual explanations plot for Dataset 1, (b) Counterfactual explanations plot for Dataset 2

Chapter 7

Conclusion

The concluding chapter brings concurrently the research presented in the course of this thesis, outlining the significant results and highlighting their broader suggestions. The work in this thesis initiates with reviewing the study process, from the initial environmental analysis to the development of an XAI framework, emphasizing how the main objectives were achieved. This chapter highlights the key contributions to precision agriculture through systemic methodological, and practical advancements. It also explores the effects of environment on agricultural industry. This chapter also addresses the limitations associated with the current study and suggests possible avenues for future investigation. Finally, the chapter concludes by outlining a pathway for upcoming research opportunities in precision agriculture.

The primary objective of this thesis is to develop, construct, and verify a framework for PA, using AI models and XAI to establish a sustainable, efficient, and reliable decision-support system for farmers. The goal is to achieve by performing several interdependent studies, each targeting a specific stage of the agricultural cycle. The key findings from each chapter highlight the accomplishment of the research objectives.

The work begins with Chapter 1, which highlights the significance of precision agriculture in responding to problems like food security at a global scale due to rising populations and climate change. It introduced how AI and IoT can be used to enhance farm management, improve efficiency, and reduce environmental impacts. The chapter outlines the research objectives, including developing models for rainfall prediction, smart irrigation, soil fertility classification, and crop recommendations. The aim is to support farmers with data-driven, sustainable farming practices using AI and IoT technologies.

Chapter 2 provided the foundation for precision agriculture by focusing on environmental understanding through rainfall prediction. This chapter demonstrated the ACO-ANN hybrid model for daily rainfall forecasting in Alipurduar, West Bengal. The ACO-ANN model achieve a RMSE of 2.8402 and a MAE of 1.3978, which outperforms the traditional regression models and other hybrid configurations. This confirme the capability of hybrid ML systems to handle non-linear meteorological data.

Building on the foundation, Chapter 3 convert knowledge into action by developing a smart irrigation system. This chapter addresses the critical issue of water management by creating an intelligent irrigation system for coriander crop. The system utilize ANN optimized with GA-ANN to predict reference EV_{T0} with good accuracy. The Hybrid model achieve the lowest RMSE values across all crop growth stages. Utilizing IoT sensors for live data acquisition and Android app for remote operation help the system to minimize water destruction.

The primary goal of Chapter 4 is to classify soil fertility based on soil nutrients. It established ACO-XgBoost, a hybrid ML model with 98% accuracy in classifying soil fertility. Additionally, the chapter uses XAI tool to promote transparency and trust in the model's outputs for the farmers. The system gained farmers' trust by highlighting the most essential nutrients, such as NPK from the set of micro and macro soil nutrients.

Chapter 5 develops a system for recommending suitable crops and appropriate fertilizer applications for the local conditions of West Bengal. The highest-ranking model in the evaluation, CatBoost-SSA, achieved $\approx 99\%$ in recommending the crop for specific land. It also calculates the optimal amount of fertilizer to improve yield while protecting the environment. To support the "green AI" approach, this chapter evaluates the environmental impact of AI models.

Chapter 6 presents an XAI framework to find the proper reason behind a specific crop yield prediction. Using tools like SHAP and LIME, the framework explains the implications of diverse elements like precipitation levels and soil characteristics on crop yield. This transparency helps to build farmers' faith and assurance in using modern AI tools for their agricultural decisions.

7.1 Principal Contributions

7.1.1 Methodological Contribution

The work presented in the thesis developed a hybrid AI model particularly designed for farming challenges. Instead of using traditional algorithms, this research combined them with advanced optimization techniques to get better results. Key new models include: ACO-ANN for rainfall prediction, GA-ANN for optimized use of water in irrigation, ACO-XgBoost for soil fertility classification, and CatBoost-SSA for providing sustainable crop recommendations.

7.1.2 System-level Contribution

A significant contribution is the creation of an end-to-end framework. Rather than solving a single problem, this thesis built a connected system for guiding a farmer through the whole agricultural process from checking weather to making final crop choices. It shows how different AI tools can logically work together to create a smart agrarian ecosystem.

7.2 Implications of the Research

7.2.1 For Farmers

The framework provides empowering farmers through data-driven technologies that help them make more intelligent and logical decisions. Accurate forecasts and recommendations reduce risks linked to weather and soil conditions. Automated systems, like the smart irrigation, also reduces manual labor. By simplifying AI recommendations, the system enables farmers to adopt and utilize advanced technology confidently.

7.2.2 For Environment

The research offers a clear path towards sustainable farming. Automated irrigation system is designed to cut down the waste of water significantly. The targeted fertilizer recommendations helps to prevent the overuse of chemicals, protecting soil health and preventing water pollution. The work also introduces methods to measure the ecological footprint associated with AI-based systems, encouraging future researchers for the development of eco-friendly technology.

7.3 Limitations of the Work

- The studies were mainly conducted in specific regions of West Bengal, India. The models were trained on local soil and climate data. Thus they may not work as well in other parts of the world without adjustment.
- The smart irrigation study focused only on coriander, and the recommendation system included a specific list of crops relevant to West Bengal. More research is needed to apply these systems to a broader range of crops.
- The research did not include long-term, multi-season field trials. Farming is a cyclical process, and validating the system's performance and benefits over several years is the next step to prove its long-term value.

7.4 Recommendations for Future Research

The models are required to be tested and adapted for various climates, soil varieties, and a much wider variety of crops. This would make the system a more universal tool for farmers.

The recommendation system could be made even more powerful by adding real-time economic data. Information such as crop market prices and fertilizer costs can support farmers in formulating decisions that benefit not only their land but also their financial well-being.

The farmer-friendly web and mobile apps, can be improved with more interactive features and better data visualizations. This will make it even easier for farmers to understand and use AI advice.

The most important next step is to conduct large-scale, multi-year studies on real farms. This would provide strong evidence for long-term benefits in crop yield, resource savings, and farm profitability.

This work is set to reshape farming decisions with the help of AI. The work has transitioned from basic predictions to developing a full-fledged, intelligent, and easily understandable ecosystem for smart agriculture. By addressing the key challenges of forecasting, irrigation, soil health, recommendations, and trust, this research provides a solid foundation for advancing PA. This journey from complex data to clear, actionable insights marks a significant step toward an era where agriculture is more creative, efficient, and environmental friendly.

Bibliography

- [1] Belay Simane, Thandi Kapwata, Natasha Naidoo, Guéladio Cissé, Caradee Y Wright, and Kiros Berhane. Ensuring africa’s food security by 2050: the role of population growth, climate-resilient strategies, and putative pathways to resilience. *Foods*, 14(2):262, 2025.
- [2] Bernard Gilland. World population and food supply: can food production keep pace with population growth in the next half-century? *Food policy*, 27(1):47–63, 2002.
- [3] Sara Hertog, Patrick Gerland, John Wilmoth, et al. India overtakes china as the world’s most populous country. 2023.
- [4] Mark J Koetse and Piet Rietveld. The impact of climate change and weather on transport: An overview of empirical findings. *Transportation Research Part D: Transport and Environment*, 14(3):205–221, 2009.
- [5] Ministry of Finance and Government of India. *Economic Survey 2017-18 (Volume I and Volume II)*. Oxford University Press, 2018.
- [6] Bruce Erickson and Scott W Fausti. The role of precision agriculture in food security. *Agronomy Journal*, 113(6):4455–4462, 2021.
- [7] Robert E Evenson and Douglas Gollin. Assessing the impact of the green revolution, 1960 to 2000. *science*, 300(5620):758–762, 2003.
- [8] David Mulla and Raj Khosla. Historical evolution and recent advances in precision farming. *Soil-specific farming precision agriculture*, pages 1–35, 2016.
- [9] James E Addicott. The precision farming revolution. In *The Precision Farming Revolution: Global Drivers of Local Agricultural Methods*, pages 1–35. Springer, 2019.
- [10] Sourabh Sagar and Mahantesh N Birje. Precision agriculture using internet of things and cloud computing: A review. *SN Computer Science*, 6(4):1–18, 2025.
- [11] Abhishek Thakur, Ankit Kumar, and Sudhansu Kumar Mishra. Control techniques for vision-based autonomous vehicles for agricultural applications: A meta-analytic review. In *International Conference on Artificial Intelligence on Textile and Apparel*, pages 1–14. Springer, 2023.
- [12] Ghulam Mohyuddin, Muhammad Adnan Khan, Abdul Haseeb, Shahzadi Mahpara, Muhammad Waseem, and Ahmed Mohammed Saleh. Evaluation of machine learning

- approaches for precision farming in smart agriculture system: a comprehensive review. *IEEE access*, 12:60155–60184, 2024.
- [13] Chandni Khandelwal, Mridul Singhal, Gaurav Gaurav, GS Dangayach, and ML Meena. Agriculture supply chain management: a review (2010–2020). *Materials Today: Proceedings*, 47:3144–3153, 2021.
- [14] Sneha Kumari, VG Venkatesh, Felix Ter Chian Tan, S Vijayakumar Bharathi, M Ramasubramanian, and Yangyan Shi. Application of machine learning and artificial intelligence on agriculture supply chain: a comprehensive review and future research directions. *Annals of Operations Research*, 348(3):1573–1617, 2025.
- [15] Vaidyeswaran Rajaraman. Johnmccarthy—father of artificial intelligence. *Resonance*, 19(3):198–207, 2014.
- [16] Marly van Assen, Emanuele Muscogiuri, Giovanni Tessarin, and Carlo N De Cecco. Artificial intelligence: A century-old story. In *Artificial Intelligence in Cardiothoracic Imaging*, pages 3–13. Springer, 2022.
- [17] Abhinav Sharma, Arpit Jain, Prateek Gupta, and Vinay Chowdary. Machine learning applications for precision agriculture: A comprehensive review. *IEEE Access*, 9:4843–4873, 2020. doi: 10.1109/ACCESS.2020.3048415.
- [18] T Saranya, C Deisy, S Sridevi, and Kalaiarasi Sonai Muthu Anbananthen. A comparative study of deep learning and internet of things for precision agriculture. *Engineering Applications of Artificial Intelligence*, 122:106034, 2023.
- [19] Saurabh Bhattacharya and Manju Pandey. Developing an agriculture ontology for extracting relationships from texts using natural language processing to enhance semantic understanding. *International Journal of Information Technology*, pages 1–20, 2024.
- [20] Haider Mahmood Jawad, Aqeel Mahmood Jawad, Rosdiadee Nordin, Sadik Kamel Gharghan, Nor Fadzilah Abdullah, Mahamod Ismail, and Mahmood Jawad Abu-AlShaeer. Accurate empirical path-loss model based on particle swarm optimization for wireless sensor networks in smart agriculture. *IEEE Sensors Journal*, 20(1):552–561, 2019.
- [21] Xiao Jian Tan, Wai Loon Cheor, Kwok Shien Yeo, and Wai Zhe Leow. Expert systems in oil palm precision agriculture: A decade systematic review. *Journal of King Saud University-Computer and Information Sciences*, 34(4):1569–1594, 2022.
- [22] Moomal Bukhari, Syed Owais Athar, Mukhtar Ullah, and Muhammad Naveed Aman. A fuzzy logic-based smart irrigation controller for precision agriculture. *IEEE Internet of Things Journal*, 2024.
- [23] Margarida Almeida, Eva Sousa, Carolina Rodrigues, Maria Beatriz Candeias, and Manuel Au-Yong-Oliveira. Samsung vs. apple: How different communication strategies affect consumers in portugal. *Administrative Sciences*, 11(1):19, 2021.

- [24] Nurzaman Ahmed, Debashis De, and Iftekhhar Hussain. Internet of things (iot) for smart precision agriculture and farming in rural areas. *IEEE internet of things journal*, 5(6): 4890–4899, 2018.
- [25] Jian Gao, Wenzhi Zeng, Zhipeng Ren, Chang Ao, Guoqing Lei, Thomas Gaiser, and Amit Kumar Srivastava. A fertilization decision model for maize, rice, and soybean based on machine learning and swarm intelligent search algorithms. *Agronomy*, 13(5):1400, 2023.
- [26] Jyoti A Dhanke, Diksha Srivastava, D Menaga, Roop Raj, Kambala Vijaya Kumar, Pradeep Jangir, and P Mani. Climate-based ai-powered precision irrigation: Sustainably smart agriculture frameworks for maximum crop yields. *Remote Sensing in Earth Systems Sciences*, 8(1):161–172, 2025.
- [27] Anna Chlingaryan, Salah Sukkarieh, and Brett Whelan. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and electronics in agriculture*, 151:61–69, 2018.
- [28] Mahlatse Kganyago, Clement Adjorlolo, Paidamwoyo Mhangara, and Lesiba Tsoeleng. Optical remote sensing of crop biophysical and biochemical parameters: An overview of advances in sensor technologies and machine learning algorithms for precision agriculture. *Computers and Electronics in Agriculture*, 218:108730, 2024.
- [29] Arthur L Samuel. Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3):210–229, 1959.
- [30] Alan M Turing. Computing machinery and intelligence. In *Parsing the Turing test: Philosophical and methodological issues in the quest for the thinking computer*, pages 23–65. Springer, 2007.
- [31] Tiago Domingues, Tomás Brandão, and João C Ferreira. Machine learning for detection and prediction of crop diseases and pests: A comprehensive survey. *Agriculture*, 12(9): 1350, 2022.
- [32] Sean D Holcomb, William K Porter, Shaun V Ault, Guifen Mao, and Jin Wang. Overview on deepmind and its alphago zero ai. In *Proceedings of the 2018 international conference on big data and education*, pages 67–71, 2018.
- [33] Carlos F Gaitán. Machine learning applications for agricultural impacts under extreme events. In *Climate extremes and their implications for impact and risk assessment*, pages 119–138. Elsevier, 2020.
- [34] Emmanuelle Garrigues, Michael S Corson, Denis A Angers, Hayo MG van der Werf, and Christian Walter. Soil quality in life cycle assessment: Towards development of an indicator. *Ecological indicators*, 18:434–442, 2012.
- [35] Ian C Grieve. Human impacts on soil properties and their implications for the sensitivity of soil systems in scotland. *Catena*, 42(2-4):361–374, 2001.

- [36] KS Karthika, I Rashmi, and MS Parvathi. Biological functions, uptake and transport of essential nutrients in relation to plant growth. In *Plant nutrients and abiotic stress tolerance*, pages 1–49. Springer, 2018.
- [37] Bhuwan Kashyap and Ratnesh Kumar. Sensing methodologies in agriculture for soil moisture and nutrient monitoring. *IEEE Access*, 9:14095–14121, 2021.
- [38] ACAR Emrullah, Mehmet Sirac OZERDEM, and Burak Berk USTUNDAG. Machine learning based regression model for prediction of soil surface humidity over moderately vegetated fields. In *2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, pages 1–4. IEEE, 2019.
- [39] Xiaogang Wang, Wenjin Hu, Kaishu Li, Lepeng Song, and Luqing Song. Modeling of soft sensor based on dbn-elm and its application in measurement of nutrient solution composition for soilless culture. In *2018 IEEE International Conference of Safety Produce Informatization (IICSPI)*, pages 93–97. IEEE, 2018.
- [40] Seonyoung Park, Jungho Im, Sumin Park, and Jinyoung Rhee. Amsr2 soil moisture down-scaling using multisensor products through machine learning approach. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1984–1987. IEEE, 2015.
- [41] Mark A Friedl, Damien Sulla-Menashe, Bin Tan, Annemarie Schneider, Navin Ramankutty, Adam Sibley, and Xiaoman Huang. Modis collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote sensing of Environment*, 114(1):168–182, 2010.
- [42] Rabie Reda, Taoufiq Saffaj, Bouzida Ilham, Ouadi Saidi, Kadmiri Issam, Lakssir Brahim, and El Mestafa El Hadrami. A comparative study between a new method and other machine learning algorithms for soil organic carbon and total nitrogen prediction using near infrared spectroscopy. *Chemometrics and Intelligent Laboratory Systems*, 195:103873, 2019.
- [43] Vahid Farhangmehr, Juan Cobo, Abdolmajid Mohammadian, Pierre Payeur, Hamidreza Shirkhani, and Hanifeh Imanian. A convolutional neural network model for soil temperature prediction under ordinary and hot weather conditions: Comparison with a multi-layer perceptron model. *Sustainability*, 15:7897, 2023. doi: 10.3390/su15107897. URL <https://www.mdpi.com/2071-1050/15/10/7897/pdf?version=1683854674>.
- [44] Priyanka Priyanka, Praveen Kumar, Sucheta Panda, Tejinder Thakur, K. V. Uday, and Varun Dutt. Can machine learning models predict soil moisture evaporation rates? an investigation via novel feature selection techniques and model comparisons. *Frontiers in Earth Science*, 12, 2024. doi: 10.3389/feart.2024.1344690. URL <https://www.frontiersin.org/articles/10.3389/feart.2024.1344690/pdf?isPublishedV2=False>.
- [45] Jinyang Du, John S. Kimball, K. G. Jencso, Zachary Hoylman, Colin Brust, David Ketchum, Y. Xu, Hui Lü, and Justin Sheffield. Machine-learning based multi-layer soil

- moisture forecasts—an application case study of the montana 2017 flash drought. *Water Resources Research*, 60, 2024. doi: 10.1029/2023wr036973.
- [46] Yulan Chen, Jianjun Li, Ziqi Zhang, Juying Jiao, Nan Wang, Leichao Bai, Yue Liang, Qian Xu, and Shijie Zhang. Modeling soil loss under rainfall events using machine learning algorithms. *Journal of Environmental Management*, 352:120004, 2024. doi: 10.1016/j.jenvman.2023.120004.
- [47] Lei Xu, Hong wei Yu, Zeqiang Chen, Wenying Du, Nengcheng Chen, and Min Huang. Hybrid deep learning and s2s model for improved sub-seasonal surface and root-zone soil moisture forecasting. *Remote sensing*, 15:3410, 2023. doi: 10.3390/rs15133410. URL <https://www.mdpi.com/2072-4292/15/13/3410/pdf?version=1688555394>.
- [48] Qin Jiang, Weiyue Li, Zedong Fan, Xiaogang He, Weiwei Sun, Sheng Chen, Jiahong Wen, Jun Gao, and Jun Wang. Evaluation of the era5 reanalysis precipitation dataset over chinese mainland. *Journal of hydrology*, 595:125660, 2021.
- [49] Shannon Sterling and Agnès Ducharne. Comprehensive data set of global land cover change for land surface model applications. *Global Biogeochemical Cycles*, 22(3), 2008.
- [50] Lu Li, Yongjiu Dai, Wei Shangguan, Nan Wei, and Zhongwang Wei. Multistep forecasting of soil moisture using spatiotemporal deep encoder–decoder networks. *Journal of Hydrometeorology*, 2022. doi: 10.1175/jhm-d-21-0131.1. URL <https://journals.ametsoc.org/downloadpdf/journals/hydr/23/3/JHM-D-21-0131.1.pdf>.
- [51] Xiaojie Li, Jian Tang, and Chengxiang Yin. Sequence-to-sequence learning for prediction of soil temperature and moisture. *IEEE Geoscience and Remote Sensing Letters*, 19:1 – 5, 2022. doi: 10.1109/lgrs.2022.3141072.
- [52] Atta ur Rahman, Sagheer Abbas, Mohammed Gollapalli, Rashad Ahmed, Shabib Aftab, Munir Uddin Ahmad, Muhammad Adnan Khan, and Amir Mosavi. Rainfall prediction system using machine learning fusion for smart cities. *Sensors*, 22:3504, 2022. doi: 10.3390/s22093504. URL <https://www.mdpi.com/1424-8220/22/9/3504/pdf?version=1652085887>.
- [53] Renze Dong, Hongze Leng, Chengwu Zhao, Junqiang Song, Juan Zhao, and Xiaoqun Cao. A hybrid data assimilation system based on machine learning. *Frontiers in Earth Science*, 10, 2023. doi: 10.3389/feart.2022.1012165. URL <https://doi.org/10.3389/feart.2022.1012165>.
- [54] Yu Li, Wei Lv, and Zhiyao Liang. Prediction of daily temperature based on the robust machine learning algorithms. *Sustainability*, 15:9289, 2023. doi: 10.3390/su15129289. URL <https://doi.org/10.3390/su15129289>.
- [55] K. Venkatachalam, P. Trojovský, Dragan Pamučar, Nebojsa Bacanin, and Vladimir Simic. Dwfh: An improved data-driven deep weather forecasting hybrid model using transductive long short term memory (t-lstm). *Expert systems with applications*, 213:119270, 2022. doi: 10.1016/j.eswa.2022.119270.

- [56] Hongchao Xu and Zhenyu Wu. Mcx-yolov5: Efficient helmet detection in complex power warehouse scenarios. *Journal of Real-Time Image Processing*, 21(2):27, 2024.
- [57] Fu Jie Tey, Tin-Yu Wu, and Jiann-Liang Chen. Machine learning-based short-term rainfall prediction from sky data. *ACM Transactions on Knowledge Discovery From Data*, 16:1 – 18, 2022. doi: 10.1145/3502731.
- [58] Yangyang Li, Harianto Rahardjo, Alfrendo Satyanaga, Saranya Rangarajan, and Daryl Tsen-Tieng Lee. Soil database development with the application of machine learning methods in soil properties prediction. *Engineering Geology*, 306:106769, 2022.
- [59] Javier I Rivera and Carlos A Bonilla. Predicting soil aggregate stability using readily available soil properties and machine learning techniques. *Catena*, 187:104408, 2020.
- [60] J Padarian, B Minasny, and AB McBratney. Using deep learning to predict soil properties from regional spectral data. *Geoderma Regional*, 16:e00198, 2019.
- [61] Kathrin J Ward, Sabine Chabrilat, Carsten Neumann, and Saskia Foerster. A remote sensing adapted approach for soil organic carbon prediction based on the spectrally clustered lucas soil database. *Geoderma*, 353:297–307, 2019.
- [62] Sami Khanal, John Fulton, Andrew Klopfenstein, Nathan Douridas, and Scott Shearer. Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield. *Computers and electronics in agriculture*, 153:213–225, 2018.
- [63] Antonios Morellos, Xanthoula-Eirini Pantazi, Dimitrios Moshou, Thomas Alexandridis, Rebecca Whetton, Georgios Tziotzios, Jens Wiebenson, Ralf Bill, and Abdul M Mouazen. Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using vis-nir spectroscopy. *Biosystems Engineering*, 152:104–116, 2016.
- [64] Akash Manish Lad, K Mani Bharathi, B Akash Saravanan, and R Karthik. Factors affecting agriculture and estimation of crop yield using supervised learning algorithms. *Materials Today: Proceedings*, 62:4629–4634, 2022.
- [65] SK Singh, Rajat Saxena, Akhilesh Porwal, Neetu Ray, and SS Ray. Assessment of hailstorm damage in wheat crop using remote sensing. *Current Science*, pages 2095–2100, 2017.
- [66] Elisa Kamir, François Waldner, and Zvi Hochman. Estimating wheat yields in australia using climate records, satellite image time series and machine learning methods. *ISPRS Journal of Photogrammetry and Remote Sensing*, 160:124–135, 2020.
- [67] Hossein Aghighi, Mohsen Azadbakht, Davoud Ashourloo, Hamid Salehi Shahrabi, and Soheil Radiom. Machine learning regression techniques for the silage maize yield prediction using time-series images of landsat 8 oli. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(12):4563–4577, 2018.
- [68] Kentaro Kuwata and Ryosuke Shibasaki. Estimating crop yields with deep learning and remotely sensed data. In *2015 IEEE international geoscience and remote sensing symposium (IGARSS)*, pages 858–861. IEEE, 2015.

- [69] Chaofa Bian, Hongtao Shi, Suqin Wu, Kefei Zhang, Meng Wei, Yindi Zhao, Yaqin Sun, Huifu Zhuang, Xuewei Zhang, and Shuo Chen. Prediction of field-scale wheat yield using machine learning method and multi-spectral uav data. *Remote Sensing*, 14(6):1474, 2022.
- [70] Sandeep Gupta, Angelina Geetha, K Sakthidasan Sankaran, Abu Sarwar Zamani, Mahyudin Ritonga, Roop Raj, Samrat Ray, and Hussien Sobahi Mohammed. Machine learning-and feature selection-enabled framework for accurate crop yield prediction. *Journal of Food Quality*, 2022(1):6293985, 2022.
- [71] Chandan Kumar, Partson Mubvumba, Yanbo Huang, Jagman Dhillon, and Krishna Reddy. Multi-stage corn yield prediction using high-resolution uav multispectral data and machine learning models. *Agronomy*, 13(5):1277, 2023.
- [72] Saiara Samira Sajid, Mohsen Shahhosseini, Isaiah Huber, Guiping Hu, and Sotirios V Archontoulis. County-scale crop yield prediction by integrating crop simulation with machine learning models. *Frontiers in Plant Science*, 13:1000224, 2022.
- [73] Marzhan Sadenova, Nail Beisekenov, Petar Sabev Varbanov, and Ting Pan. Application of machine learning and neural networks to predict the yield of cereals, legumes, oilseeds and forage crops in kazakhstan. *Agriculture*, 13(6):1195, 2023.
- [74] S Pavani and Augusta Sophy Beulet P. Improved precision crop yield prediction using weighted-feature hybrid svm: Analysis of ml algorithms. *IETE Journal of Research*, 70(3):2628–2640, 2024.
- [75] Ravindra Yadav, Anita Seth, and Naresh Dembla. Optimizing crop yield prediction: Data-driven analysis & machine learning modeling using usda datasets. *Current Agriculture Research Journal*, 12(1):272–285, 2024.
- [76] Mohamed Bouni, Badr Hssina, Khadija Douzi, and Samira Douzi. Integrated iot approaches for crop recommendation and yield-prediction using machine-learning. *Internet of Things (IoT)*, 5(4), 2024.
- [77] Nazifi Shuaibu, GN Obunadike, and Bashir Ahmad Jamilu. Crop yield prediction using selected machine learning algorithms. *FUDMA journal of Sciences*, 8(1):61–68, 2024.
- [78] Kavita Jhajharia and Pratistha Mathur. Machine learning approaches to predict crop yield using integrated satellite and climate data. *International Journal of Ambient Computing and Intelligence (IJACI)*, 13(1):1–17, 2022.
- [79] Farhat Abbas, Hassan Afzaal, Aitazaz A Farooque, and Skylar Tang. Crop yield prediction through proximal sensing and machine learning algorithms. *Agronomy*, 10(7):1046, 2020.
- [80] Martin Kuradusenge, Eric Hitimana, Damien Hanyurwimfura, Placide Rukundo, Kam-bombo Mtonga, Angelique Mukasine, Claudette Uwitonze, Jackson Ngabonziza, and Angelique Uwamahoro. Crop yield prediction using machine learning models: Case of irish potato and maize. *Agriculture*, 13(1):225, 2023.

- [81] Shruti Kulkarni, Shah Nawaz Mandal, G Srivatsa Sharma, Monica R Mundada, et al. Predictive analysis to improve crop yield using a neural network model. In *2018 international conference on advances in computing, communications and informatics (ICACCI)*, pages 74–79. IEEE, 2018.
- [82] Zheng Chu and Jiong Yu. An end-to-end model for rice yield prediction using deep learning fusion. *Computers and Electronics in Agriculture*, 174:105471, 2020.
- [83] Puyu Feng, Bin Wang, De Li Liu, Cathy Waters, Dengpan Xiao, Lijie Shi, and Qiang Yu. Dynamic wheat yield forecasts are improved by a hybrid approach using a biophysical model and machine learning technique. *Agricultural and Forest Meteorology*, 285:107922, 2020.
- [84] Yaping Cai, Kaiyu Guan, David Lobell, Andries B Potgieter, Shaowen Wang, Jian Peng, Tianfang Xu, Senthil Asseng, Yongguang Zhang, Liangzhi You, et al. Integrating satellite and climate data to predict wheat yield in australia using machine learning approaches. *Agricultural and forest meteorology*, 274:144–159, 2019.
- [85] Abdülkadir Gümüşçü, Mehmet Emin Tenekeci, and Ali Volkan Bilgili. Estimation of wheat planting date using machine learning algorithms based on available climate data. *Sustainable Computing: Informatics and Systems*, 28:100308, 2020.
- [86] Louis Kouadio, Ravinesh C Deo, Vivekananda Byrareddy, Jan F Adamowski, Shahbaz Mushtaq, and Van Phuong Nguyen. Artificial intelligence approach for the prediction of robusta coffee yield using soil fertility properties. *Computers and electronics in agriculture*, 155:324–338, 2018.
- [87] Andrea A Gamboa, Paula A Cáceres, Henry Lamos, Diego A Zárate, and David E Puentes. Predictive model for cocoa yield in santander using supervised machine learning. In *2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*, pages 1–5. IEEE, 2019.
- [88] Qiong Su and Vijay P Singh. Advancing irrigation management: integrating technology and sustainability to address global food security. *Environmental monitoring and assessment*, 196(11):1018, 2024.
- [89] Muhammed Fuhad, Stenin K George, Manu Elappila, Shamanth Nagaraju, Vandana Reddy, Addapalli VN Krishna, and Sachin Malayath Jose. Cost effective and energy efficient drip irrigation system for iot enabled smart agriculture. In *International Conference on Advances in Distributed Computing and Machine Learning*, pages 199–210. Springer, 2024.
- [90] Robert G Evans, Jake LaRue, Kenneth C Stone, and Bradley A King. Adoption of site-specific variable rate sprinkler irrigation systems. *Irrigation science*, 31(4):871–887, 2013.
- [91] Enrique Playán and Luciano Mateos. Modernization and optimization of irrigation systems to increase water productivity. *Agricultural water management*, 80(1-3):100–116, 2006.

- [92] Michael J Roberts, Noah O Braun, Thomas R Sinclair, David B Lobell, and Wolfram Schlenker. Comparing and combining process-based crop models and statistical models with some implications for climate change. *Environmental Research Letters*, 12(9):095010, 2017.
- [93] Jiusheng Li, RE Yoder, LO Odhiambo, and J Zhang. Simulation of nitrate distribution under drip irrigation using artificial neural networks. *Irrigation science*, 23(1):29–37, 2004.
- [94] G Kavianand, VM Nivas, R Kiruthika, and S Lalitha. Smart drip irrigation system for sustainable agriculture. In *2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)*, pages 19–22. IEEE, 2016.
- [95] Emmanuel Abiodun Abioye, Mohammad Shukri Zainal Abidin, Mohd Saiful Azimi Mahmud, Salinda Buyamin, Muhammad Khairie Idham AbdRahman, Abdulrahaman Okino Otuoze, Muhammad Shahrul Azwan Ramli, and Ona Denis Ijike. Iot-based monitoring and data-driven modelling of drip irrigation system for mustard leaf cultivation experiment. *Information Processing in Agriculture*, 8(2):270–283, 2021.
- [96] Amin Seyedzadeh, Saman Maroufpoor, Eisa Maroufpoor, Jalal Shiri, Omid Bozorg-Haddad, and Farnoosh Gavazi. Artificial intelligence approach to estimate discharge of drip tape irrigation based on temperature and pressure. *Agricultural Water Management*, 228:105905, 2020.
- [97] Yaoqi Peng, Yingxin Xiao, Zetian Fu, Yuhong Dong, Yongjun Zheng, Haijun Yan, and Xinxing Li. Precision irrigation perspectives on the sustainable water-saving of field crop production in china: Water demand prediction and irrigation scheme optimization. *Journal of cleaner production*, 230:365–377, 2019.
- [98] Jalal Shiri, Bakhtiar Karimi, Nazir Karimi, Mohammad Hossein Kazemi, and Sepideh Karimi. Simulating wetting front dimensions of drip irrigation systems: Multi criteria assessment of soft computing models. *Journal of Hydrology*, 585:124792, 2020.
- [99] MN Elnesr and AA Alazba. Simulation of water distribution under surface dripper using artificial neural networks. *Computers and Electronics in Agriculture*, 143:90–99, 2017.
- [100] Yu-Chuan Chang, Ting-Wei Huang, and Nen-Fu Huang. A machine learning based smart irrigation system with lora p2p networks. In *2019 20th Asia-Pacific Network Operations and Management Symposium (APNOMS)*, pages 1–4. IEEE, 2019.
- [101] Gursimran Singh, Deepak Sharma, Amarendra Goap, Sugandha Sehgal, AK Shukla, and Satish Kumar. Machine learning based soil moisture prediction for internet of things based smart irrigation system. In *2019 5th International conference on signal processing, computing and control (ISPCC)*, pages 175–180. IEEE, 2019.
- [102] Roque Torres-Sanchez, Honorio Navarro-Hellin, Antonio Guillamon-Frutos, Ruben San-Segundo, Maria Carmen Ruiz-Abellón, and Rafael Domingo-Miguel. A decision support system for irrigation management: Analysis and implementation of different learning techniques. *Water*, 12(2):548, 2020.

- [103] Mehdi Jamei, Farahnaz Karimi, Mumtaz Ali, Bakhtiar Karimi, Masoud Karbasi, and Younes Aminpour. Experimental and computational assessment of wetting pattern for two-layered soil profiles in pulse drip irrigation: Designing a novel optimized bidirectional deep learning paradigm. *Journal of Hydrology*, 614:128496, 2022.
- [104] Fatma Trabelsi and Salsebil Bel Hadj Ali. Exploring machine learning models in predicting irrigation groundwater quality indices for effective decision making in medjerda river basin, tunisia. *Sustainability*, 14(4):2341, 2022.
- [105] Ali Mokhtar, Nadhir Al-Ansari, Wessam El-Ssawy, Renata Graf, Pouya Aghelpour, Hongming He, Salma M Hafez, and Mohamed Abuarab. Prediction of irrigation water requirements for green beans-based machine learning algorithm models in arid region. *Water resources management*, 37(4):1557–1580, 2023.
- [106] Rodrigo Togneri, Diego Felipe Dos Santos, Glauber Camponogara, Hitoshi Nagano, Gilliard Custodio, Ronaldo Prati, Stenio Fernandes, and Carlos Kamienski. Soil moisture forecast for smart irrigation: The primetime for machine learning. *Expert Systems with Applications*, 207:117653, 2022.
- [107] Shiqi Wei, Tianfang Xu, Guo-Yue Niu, and Ruijie Zeng. Estimating irrigation water consumption using machine learning and remote sensing data in kansas high plains. *Remote Sensing*, 14(13):3004, 2022.
- [108] Ezgi Kurtulmuş, Bilge Arslan, and Ferhat Kurtulmuş. Deep learning for proximal soil sensor development towards smart irrigation. *Expert systems with applications*, 198:116812, 2022.
- [109] Huaijie He, Ling Liu, and Xiuqun Zhu. Optimization of extreme learning machine model with biological heuristic algorithms to estimate daily reference evapotranspiration in hetao irrigation district of china. *Engineering Applications of Computational Fluid Mechanics*, 16(1):1939–1956, 2022.
- [110] Zhehan Tang, Yufang Jin, Patrick H Brown, and Meerae Park. Estimation of tomato water status with photochemical reflectance index and machine learning: Assessment from proximal sensors and uav imagery. *Frontiers in Plant Science*, 14:1057733, 2023.
- [111] KyeongHwan Lee, Naiqian Zhang, and Sanjoy Das. Comparing adaptive neuro-fuzzy inference system (anfis) to partial least-squares (pls) method for simultaneous prediction of multiple soil properties. In *2003 ASAE Annual Meeting*, page 1. American Society of Agricultural and Biological Engineers, 2003.
- [112] Renkuan Liao, Shirui Zhang, Xin Zhang, Mingfei Wang, Huarui Wu, and Lili Zhangzhong. Development of smart irrigation systems based on real-time soil moisture data in a greenhouse: Proof of concept. *Agricultural Water Management*, 245:106632, 2021.
- [113] Zhijun Chen, Shijun Sun, Yixin Wang, Qiuyao Wang, and Xudong Zhang. Temporal convolution-network-based models for modeling maize evapotranspiration under mulched drip irrigation. *Computers and electronics in agriculture*, 169:105206, 2020.

- [114] Abhishek Khanna and Sanmeet Kaur. Evolution of internet of things (iot) and its significant impact in the field of precision agriculture. *Computers and electronics in agriculture*, 157:218–231, 2019.
- [115] R Nikhil, BS Anisha, and Ramakanth Kumar. Real-time monitoring of agricultural land with crop prediction and animal intrusion prevention using internet of things and machine learning at edge. In *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, pages 1–6. IEEE, 2020.
- [116] Haider Mahmood Jawad, Rosdiadee Nordin, Sadik Kamel Gharghan, Aqeel Mahmood Jawad, and Mahamod Ismail. Energy-efficient wireless sensor networks for precision agriculture: A review. *Sensors*, 17(8):1781, 2017.
- [117] Dongbo Xie, Liang Chen, Lichao Liu, Liqing Chen, and Hai Wang. Actuators and sensors for application in agricultural robots: A review. *Machines*, 10(10):913, 2022.
- [118] METER Group. Ech2o ec-5 soil moisture sensor, 2025. URL <https://metergroup.com/products/ech2o-ec-5-soil-moisture-sensor/>. Accessed: 2025-08-09.
- [119] Robu.in. Dht22 digital temperature and humidity sensor module (am2302), 2025. URL <https://robu.in/product/dht22-digital-temperature-humidity-sensor-temperature-humidity-module-am2302/>. Accessed: 2025-08-09.
- [120] Robu.in. Adafruit bh1750 light sensor – stemma qt/qwiic, 2025. URL <https://robu.in/product/adafruit-bh1750-light-sensor-stemma-qt-qwiic/>. Accessed: 2025-08-09.
- [121] Robu.in. Mh-z19 infrared co₂ sensor module for co₂ monitor, 2025. URL <https://robu.in/product/mh-z19-infrared-co2-sensor-module-co2-monitor/>. Accessed: 2025-08-09.
- [122] DFRobot. Gravity analog ph sensor / meter pro kit v2, 2025. URL <https://robu.in/product/dfrobot-gravity-analog-ph-sensor-meter-pro-kit-v2/>. Accessed: 2025-08-09.
- [123] Robu.in. Ds18b20 waterproof temperature probe (1m), 2025. URL <https://robu.in/product/ds18b20-water-proof-temperature-probe-black-1m/>. Accessed: 2025-08-09.
- [124] METTLER TOLEDO. Ion meters, 2025. URL https://www.mt.com/in/en/home/products/Laboratory_Analytics_Browse/pH-meter/pH-meters/ion-meters.html. Accessed: 2025-08-09.
- [125] Hydreon Corporation. Rg-11 optical rain gauge, 2025. URL <https://rainsensors.com/products/rg-11/>. Accessed: 2025-08-09.
- [126] Davis Instruments. Anemometer for vantage pro2 and vantage pro, 2025. URL <https://www.davisinstruments.com/products/anemometer-for-vantage-pro2-vantage-pro>. Accessed: 2025-08-09.

- [127] LI-COR Environmental. Li-250a light meter, 2025. URL <https://www.licor.com/products/light/light-meter>. Accessed: 2025-08-09.
- [128] Robu.in. Gps neo-6m satellite positioning module development board for arduino, stm32, 8051, 2025. URL <https://robu.in/product/gps-neo-6m-satellite-positioning-module-development-board-for-arduino-stm32-51/>. Accessed: 2025-08-09.
- [129] Robu.in. Mics-5524 carbon monoxide, hydrogen, methane, and alcohol gas sensor detection module, 2025. URL <https://robu.in/product/mics-5524-carbon-monoxide-hydrogen-methane-alcohol-gas-sensor-detection-module/>. Accessed: 2025-08-09.
- [130] Robu.in. Unwelded gy-sht31-d digital temperature and humidity sensor, 2025. URL <https://robu.in/product/unwelded-gy-sht31-d-digital-temperature-and-humidity-sensor/>. Accessed: 2025-08-09.
- [131] Robu.in. Bmp180 digital barometric pressure sensor module, 2025. URL <https://robu.in/product/bmp180-digital-barometric-pressure-sensor-module/>. Accessed: 2025-08-09.
- [132] Ravesa Akhter and Shabir Ahmad Sofi. Precision agriculture using iot data analytics and machine learning. *Journal of King Saud University-Computer and Information Sciences*, 34(8):5602–5618, 2022.
- [133] Sewnet Getahun, Habtamu Kefale, and Yohannes Gelaye. Application of precision agriculture technologies for sustainable crop production and environmental sustainability: A systematic review. *The Scientific World Journal*, 2024(1):2126734, 2024.
- [134] Martin Smith. The application of climatic data for planning and management of sustainable rainfed and irrigated crop production. *Agricultural and Forest meteorology*, 103(1-2):99–108, 2000.
- [135] Aondover Tarhule. Damaging rainfall and flooding: the other sahel hazards. *Climatic change*, 72(3):355–377, 2005.
- [136] Milind Mujumdar, Preethi Bhaskar, MVS Ramarao, Umakanth Uppara, Mangesh Goswami, Hemant Borgaonkar, Supriyo Chakraborty, Somaru Ram, Vimal Mishra, M Rajeevan, et al. Droughts and floods. *Assessment of climate change over the Indian region: a report of the Ministry of Earth Sciences (MoES), Government of India*, pages 117–141, 2020.
- [137] Malay K Bhowmick, Ashish K Srivastava, Sudhanshu Singh, Madhab C Dhara, Santanu S Aich, Sampad R Patra, and Abdelbagi M Ismail. Realizing the potential of coastal flood-prone areas for rice production in west bengal: prospects and challenges. *New frontiers in stress management for durable agriculture*, pages 543–577, 2020.

- [138] Moitrayee Das. An appraisal of spatio—temporal occurrence of fluvial hazard: Case study of flood in the districts of jalpaiguri and alipurduar (west bengal, india). In *Fluvial Systems in the Anthropocene: Process, Response and Modelling*, pages 265–275. Springer, 2022.
- [139] Sachchidanand Singh and Mitthan Lal Kansal. Cloudburst—a major disaster in the indian himalayan states. *Civil Engineering for Disaster Risk Reduction*, pages 115–126, 2022.
- [140] Mahdi Sharifzadeh, Alexandra Sikinioti-Lock, and Nilay Shah. Machine-learning methods for integrated renewable power generation: A comparative study of artificial neural networks, support vector regression, and gaussian process regression. *Renewable and Sustainable Energy Reviews*, 108:513–538, 2019.
- [141] Mohammed HI Dore. Climate change and changes in global precipitation patterns: what do we know? *Environment international*, 31(8):1167–1181, 2005.
- [142] E Toth, A Brath, and A Montanari. Comparison of short-term rainfall prediction models for real-time flood forecasting. *Journal of hydrology*, 239(1-4):132–147, 2000.
- [143] Jana Sillmann, Thordis Thorarinsdottir, Noel Keenlyside, Nathalie Schaller, Lisa V Alexander, Gabriele Hegerl, Sonia I Seneviratne, Robert Vautard, Xuebin Zhang, and Francis W Zwiers. Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities. *Weather and climate extremes*, 18:65–74, 2017.
- [144] Yaser Sheikhi, Seyed Mohammad Ashrafi, Mohammad Reza Nikoo, and Ali Haghighi. Enhancing daily rainfall prediction in urban areas: a comparative study of hybrid artificial intelligence models with optimization algorithms. *Applied Water Science*, 13(12):232, 2023.
- [145] Chandni Singh, Peter Dorward, and Henny Osbahr. Developing a holistic approach to the analysis of farmer decision-making: Implications for adaptation policy and practice in developing countries. *Land use policy*, 59:329–343, 2016.
- [146] Zhigang Ye, Shan Yin, Yin Cao, and Yong Wang. Ai-driven optimization of agricultural water management for enhanced sustainability. *Scientific Reports*, 14(1):25721, 2024.
- [147] Yujie Fang and Lizhong Xiong. General mechanisms of drought response and their application in drought resistance improvement in plants. *Cellular and molecular life sciences*, 72(4):673–689, 2015.
- [148] Abhinav Kumar Singh, Pankaj Kumar, Rawshan Ali, Nadhir Al-Ansari, Dinesh Kumar Vishwakarma, Kuldeep Singh Kushwaha, Kanhu Charan Panda, Atish Sagar, Ehsan Mirzania, Ahmed Elbeltagi, et al. An integrated statistical-machine learning approach for runoff prediction. *Sustainability*, 14(13):8209, 2022.
- [149] Javier Diez-Sierra and Manuel Del Jesus. Long-term rainfall prediction using atmospheric synoptic patterns in semi-arid climates with statistical and machine learning methods. *Journal of Hydrology*, 586:124789, 2020.
- [150] Binh Thai Pham, Lu Minh Le, Tien-Thinh Le, Kien-Trinh Thi Bui, Vuong Minh Le, Hai-Bang Ly, and Indra Prakash. Development of advanced artificial intelligence models for daily rainfall prediction. *Atmospheric Research*, 237:104845, 2020.

- [151] Yazid Tikhamarine, Doudja Souag-Gamane, Ali Najah Ahmed, Saad Sh Sammen, Ozgur Kisi, Yuk Feng Huang, and Ahmed El-Shafie. Rainfall-runoff modelling using improved machine learning methods: Harris hawks optimizer vs. particle swarm optimization. *Journal of Hydrology*, 589:125133, 2020.
- [152] Pengcheng Zhang, Yangyang Jia, Jerry Gao, Wei Song, and Hareton Leung. Short-term rainfall forecasting using multi-layer perceptron. *IEEE Transactions on Big Data*, 6(1): 93–106, 2018.
- [153] Adel Mellit, A Massi Pavan, and Mohamed Benghanem. Least squares support vector machine for short-term prediction of meteorological time series. *Theoretical and applied climatology*, 111(1):297–307, 2013.
- [154] Wanie M Ridwan, Michelle Sapitang, Awatif Aziz, Khairul Faizal Kushiar, Ali Najah Ahmed, and Ahmed El-Shafie. Rainfall forecasting model using machine learning methods: Case study terengganu, malaysia. *Ain Shams Engineering Journal*, 12(2):1651–1663, 2021.
- [155] Ahmed M Elshewey, Amel Ali Alhussan, Doaa Sami Khafaga, Marwa Radwan, El-Sayed M El-Kenawy, and Nima Khodadadi. An enhanced adaptive dynamic metaheuristic optimization algorithm for rainfall prediction depends on long short-term memory. *PLoS One*, 20(6):e0317554, 2025.
- [156] Saad Sh Sammen, Ozgur Kisi, Mohammad Ehteram, Ahmed El-Shafie, Nadhir Al-Ansari, Mohammad Ali Ghorbani, Shakeel Ahmad Bhat, Ali Najah Ahmed, and Shamsuddin Shahid. Rainfall modeling using two different neural networks improved by metaheuristic algorithms. *Environmental Sciences Europe*, 35(1):112, 2023.
- [157] Shashank Shetty, Diya Dharmendra, Sanjay Bankapur, Prabhu Prasad, et al. Hydrostack: A hybrid meta-ensemble machine learning framework for accurate annual rainfall prediction. In *2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, pages 1926–1934. IEEE, 2024.
- [158] El-Sayed M Elkenawy, Amel Ali Alhussan, Marwa M Eid, and Abdelhameed Ibrahim. Rainfall classification and forecasting based on a novel voting adaptive dynamic optimization algorithm. *Frontiers in Environmental Science*, 12:1417664, 2024.
- [159] Saumyajit Ghosh, Dipankar Das, Shasanka Kumar Gayen, and Pankaj Bhardwaj. Delineation of groundwater potential zones using the ahp technique: a case study of alipurduar district, west bengal. *Modeling Earth Systems and Environment*, pages 1–31, 2023.
- [160] ANKITA MALLICK, LALU DAS, JAVED AKHTER, and GAUTAM SAHA. Symbolic representation of district wise climate change scenarios of west bengal state, india, using thornthwaite classification scheme. 2023.
- [161] Lopamudra Bagchi and Somali Roy Chakraborty. Impact of climate change and resultant urban flooding in alipurduar town, west bengal-its vulnerability and mitigation. *International Journal of Innovative Research in Science Engineering and Technology*, 180(1): 335–363, 2023.

- [162] Alfred Dawson Quansah, Felicia Dogbey, Prince Junior Asilevi, Patrick Boakye, Lawrence Darkwah, Sampson Oduro-Kwarteng, Yen Adams Sokama-Neuyam, and Patrick Mensah. Assessment of solar radiation resource from the nasa-power reanalysis products for tropical climates in ghana towards clean energy application. *Scientific Reports*, 12(1):10684, 2022.
- [163] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5:115–133, 1943.
- [164] Simon Haykin. Neural networks expand sp’s horizons. *IEEE Signal Processing Magazine*, 13(2):24–49, 2002.
- [165] Andries P Engelbrecht. *Computational intelligence: an introduction*. John Wiley & Sons, 2007.
- [166] M Monjezi and H Dehghani. Evaluation of effect of blasting pattern parameters on back break using neural networks. *International Journal of Rock Mechanics and Mining Sciences*, 45(8):1446–1453, 2008.
- [167] Mary M Poulton. Neural networks as an intelligence amplification tool: A review of applications. *Geophysics*, 67(3):979–993, 2002.
- [168] Howard Demuth, Mark Beale, and Martin Hagan. *Neural network toolbox*. Mathworks, 1994.
- [169] Marco Dorigo. Optimization, learning and natural algorithms. *Ph. D. Thesis, Politecnico di Milano*, 1992.
- [170] Yasin Hajizadeh, Mike Christie, and Vasily Demyanov. Ant colony optimization for history matching and uncertainty quantification of reservoir models. *Journal of Petroleum Science and Engineering*, 77(1):78–92, 2011.
- [171] Bafe Betela and Kebede Wolka. Evaluating soil erosion and factors determining farmers’ adoption and management of physical soil and water conservation measures in bachire watershed, southwest ethiopia. *Environmental Challenges*, 5:100348, 2021.
- [172] Abhirup Paria, Arindam Giri, Subrata Dutta, and Sarmistha Neogy. Smart irrigation for coriander plant: Saving water with ai and iot. *Water Resources Management*, pages 1–17, 2025. doi: <https://doi.org/10.1007/s11269-025-04112-x>.
- [173] Fatemeh Javanbakht-Sheikhahmad, Farahnaz Rostami, Hossein Azadi, Hadi Veisi, Farzad Amiri, and Frank Witlox. Agricultural water resource management in the socio-hydrology: A framework for using system dynamics simulation. *Water Resources Management*, pages 1–20, 2024. <https://doi.org/10.1007/s11269-024-03786-z>.
- [174] George do Nascimento Araújo Júnior, José Edson Florentino de Moraes, Luciana Sandra Bastos de Souza, Antonio José Steidle Neto, Gherman Garcia Leal de Araujo, and Thieres George Freire da Silva. Use of lower quality water in irrigated agriculture and effects on forages with productive potential in semiarid regions: a review. *Environmental Processes*, 10(3):44, 2023. <https://doi.org/10.1007/s40710-023-00655-6>.

- [175] Richa Kumar. Degrowth, diversity and decentralisation: Building sustainable food systems for food and nutrition security. In *Reimagining Prosperity: Social and Economic Development in Post-COVID India*, pages 171–187. Springer, 2023. https://doi.org/10.1007/978-981-19-7177-8_10.
- [176] Stephen Luo Sheng Yong, Jing Lin Ng, Yuk Feng Huang, Chun Kit Ang, Norashikin Ahmad Kamal, Majid Mirzaei, and Ali Najah Ahmed. Enhanced daily reference evapotranspiration estimation using optimized hybrid support vector regression models. *Water Resources Management*, pages 1–29, 2024. <https://doi.org/10.1007/s11269-024-03860-6>.
- [177] Francesco Maria De Filippi and Giuseppe Sappa. The simulation of bracciano lake (central italy) levels based on hydrogeological water budget: A tool for lake water management when climate change and anthropogenic impacts occur. *Environmental Processes*, 11(1):8, 2024. <https://doi.org/10.1007/s40710-024-00688-5>.
- [178] Yazid Tikhamarine, Anurag Malik, Doudja Souag-Gamane, and Ozgur Kisi. Artificial intelligence models versus empirical equations for modeling monthly reference evapotranspiration. *Environmental Science and Pollution Research*, 27:30001–30019, 2020. <https://doi.org/10.1007/s11356-020-08792-3>.
- [179] Qicong Chen, Wenqing Zhang, Yali Wang, Weifeng Cai, Qian Ni, Cuiping Jiang, Jiyu Li, and Chunyan Shen. Genetic algorithm-back propagation neural network model-and response surface methodology-based optimization of polysaccharide extraction from cinnamonomum cassia presl, isolation, purification and bioactivities. *Foods*, 14(4):686, 2025.
- [180] Masoud Karbasi, Mumtaz Ali, Sayed M Bateni, Changhyun Jun, Mehdi Jamei, and Zaher Mundher Yaseen. Boruta extra tree-bidirectional long short-term memory model development for pan evaporation forecasting: Investigation of arid climate condition. *Alexandria Engineering Journal*, 86:425–442, 2024. <https://doi.org/10.1016/j.aej.2023.11.061>.
- [181] Marta García-Mollá, Rosa Puertas Medina, Virginia Vega-Carrero, and Carles Sanchis-Ibor. Economic efficiency of drip and flood irrigation. comparative analysis at farm scale using dea. *Agricultural Water Management*, 309:109314, 2025.
- [182] Mahmoud A Abdelhamid, Tarek Kh Abdelkader, Hassan AA Sayed, Zhao Zhang, Xiaohui Zhao, and Mohamed F Atia. Design and evaluation of a solar powered smart irrigation system for sustainable urban agriculture. *Scientific Reports*, 15(1):11761, 2025.
- [183] Youssouf Ouzani, Fatima Hiouani, Mirza Junaid Ahmad, and Kyung-Sook Choi. Impact of climate change on wheat production in algeria and optimization of irrigation scheduling for drought periods. *Water*, 17(11):1658, 2025.
- [184] S Ajith, S Vijayakumar, and N Elakkiya. Yield prediction, pest and disease diagnosis, soil fertility mapping, precision irrigation scheduling, and food quality assessment using machine learning and deep learning algorithms. *Discover Food*, 5(1):1–23, 2025.

- [185] Rab Nawaz Bashir, Mahlaqa Saeed, Mohammed Al-Sarem, Rashid Marie, Muhammad Faheem, Abdelrahman Elsharif Karrar, and Bahaeldein Elhussein. Smart reference evapotranspiration using internet of things and hybrid ensemble machine learning approach. *Internet of Things*, 24:100962, 2023. <https://doi.org/10.1016/j.iot.2023.100962>.
- [186] Reham R Mostafa, Ozgur Kisi, Rana Muhammad Adnan, Tayeb Sadeghifar, and Alban Kuriqi. Modeling potential evapotranspiration by improved machine learning methods using limited climatic data. *Water*, 15(3):486, 2023. <https://doi.org/10.3390/w15030486>.
- [187] Demet Yildirim, Erdem Küçüktopcu, Bilal Cemek, and Halis Simsek. Comparison of machine learning techniques and spatial distribution of daily reference evapotranspiration in türkiye. *Applied Water Science*, 13(4):107, 2023. <https://doi.org/10.1007/s13201-023-01912-7>.
- [188] Jayashree TR, NV Subba Reddy, and U Dinesh Acharya. Modeling daily reference evapotranspiration from climate variables: Assessment of bagging and boosting regression approaches. *Water Resources Management*, 37(3):1013–1032, 2023. <https://doi.org/10.1007/s11269-022-03399-4>.
- [189] Alexander Kocian, Giulia Carmassi, Fatjon Cela, Stefano Chessa, Paolo Milazzo, and Luca Incrocci. Iot based dynamic bayesian prediction of crop evapotranspiration in soilless cultivations. *Computers and Electronics in Agriculture*, 205:107608, 2023. <https://doi.org/10.1016/j.compag.2022.107608>.
- [190] Juan Dong, Liwen Xing, Ningbo Cui, Li Guo, Chuan Liang, Lu Zhao, Zhihui Wang, and Daozhi Gong. Estimating reference crop evapotranspiration using optimized empirical methods with a novel improved grey wolf algorithm in four climatic regions of china. *Agricultural Water Management*, 291:108620, 2024. <https://doi.org/10.1016/j.agwat.2023.108620>.
- [191] Mohaddeseh Bidabadi, Hossein Babazadeh, Jalal Shiri, and Ali Saremi. Estimation reference crop evapotranspiration (et₀) using artificial intelligence model in an arid climate with external data. *Applied Water Science*, 14(1):1–10, 2024. <https://doi.org/10.1007/s13201-023-02058-2>.
- [192] Jitendra Rajput, Man Singh, K Lal, Manoj Khanna, A Sarangi, J Mukherjee, and Shrawan Singh. Data-driven reference evapotranspiration (et₀) estimation: a comparative study of regression and machine learning techniques. *Environment, Development and Sustainability*, pages 1–28, 2023. <https://doi.org/10.1007/s10668-023-03978-4>.
- [193] Zehuai Chen, Chunlei Chen, Yingpan Yang, Xingpeng Wang, Hanghai Zhou, and Chunfang Zhang. Rhamnolipids supplement in salinized soils improves cotton growth through ameliorating soil properties and modifying rhizosphere communities. *Applied Soil Ecology*, 194:105–174, 2024. <https://doi.org/10.1016/j.apsoil.2023.105174>.
- [194] Yunus Ziya Kaya, Martina Zelenakova, Fatih Üneş, Mustafa Demirci, Helena Hlavata, and Peter Mesáros. Estimation of daily evapotranspiration in košice city (slovakia) using several soft computing techniques. *Theoretical and Applied Climatology*, 144:287–298, 2021.

- [195] Lei Zhang, Xin Zhao, Ge Zhu, Jun He, Jian Chen, Zhicheng Chen, Seydou Traore, Junguo Liu, and Vijay P Singh. Short-term daily reference evapotranspiration forecasting using temperature-based deep learning models in different climate zones in china. *Agricultural Water Management*, 289:108498, 2023. <https://doi.org/10.1016/j.agwat.2023.108498>.
- [196] Amninder Singh and Amir Haghverdi. Development and evaluation of temperature-based deep learning models to estimate reference evapotranspiration. *Artificial Intelligence in Agriculture*, 9:61–75, 2023. <https://doi.org/10.1016/j.aiaa.2023.08.003>.
- [197] Satish Bhaurao More, Paresch Chandra Deka, Amit Prakash Patil, and Sujay Raghavendra Naganna. Machine learning-based modeling of saturated hydraulic conductivity in soils of tropical semi-arid zone of india. *Sādhana*, 47(1):26, 2022. <https://doi.org/10.1007/s12046-022-01805-6>.
- [198] Ozgur Kisi, Behrooz Keshtegar, Mohammad Zounemat-Kermani, Salim Heddad, and Nguyen-Thoi Trung. Modeling reference evapotranspiration using a novel regression-based method: radial basis m5 model tree. *Theoretical and Applied Climatology*, 145(1-2):639–659, 2021. <https://doi.org/10.1007/s00704-021-03645-6>.
- [199] Richard G Allen, LS Pereira, D Raes, and M Smith. Crop evapotranspiration: Fao irrigation and drainage paper 56. *FAO, Rome, Italy*, 20, 1998.
- [200] Jiabing Cai, Yu Liu, Tingwu Lei, and Luis Santos Pereira. Estimating reference evapotranspiration with the fao penman–monteith equation using daily weather forecast messages. *Agricultural and Forest Meteorology*, 145(1-2):22–35, 2007. <https://doi.org/10.1016/j.agrformet.2007.04.012>.
- [201] Petra Sieber, Niclas Ericsson, Torun Hammar, and Per-Anders Hansson. Albedo impacts of current agricultural land use: Crop-specific albedo from modis data and inclusion in lea of crop production. *Science of the Total Environment*, 835:155455, 2022.
- [202] Andreas Matzarakis, Frank Rutz, and Helmut Mayer. Modelling radiation fluxes in simple and complex environments—application of the rayman model. *International journal of biometeorology*, 51:323–334, 2007. <https://doi.org/10.1007/s00484-006-0061-8>.
- [203] Houshang Ghamarnia, Maryam Jafarizade, Elham Meri, and Mohamad Aghbal Gobadei. Lysimetric determination of coriandrum sativum l. water requirement and single and dual crop coefficients in a semiarid climate. *Journal of Irrigation and Drainage Engineering*, 139(6):447–455, 2013. [https://doi.org/10.1061/\(ASCE\)IR.1943-4774.0000561](https://doi.org/10.1061/(ASCE)IR.1943-4774.0000561).
- [204] Haitham Abdulmohsin Afan, Wan Hanna Melini Wan Mohtar, Muammer Aksoy, Ali Najah Ahmed, Faidhahrahman Khaleel, Md Munir Hayet Khan, Ammar Hatem Kamel, Mohsen Sherif, and Ahmed El-Shafie. A multi-functional genetic algorithm-neural network model for predicting suspended sediment loads. *Water Resources Management*, pages 1–16, 2024. <https://doi.org/10.1007/s11269-024-04054-w>.

- [205] Kunal Sandip Garud, Simon Jayaraj, and Moo-Yeon Lee. A review on modeling of solar photovoltaic systems using artificial neural networks, fuzzy logic, genetic algorithm and hybrid models. *International Journal of Energy Research*, 45(1):6–35, 2021.
- [206] Abhirup Paria and Susovan Jana. Prediction of crops production using random forest regression. In *Proceedings of International Conference on Computational Intelligence, Data Science and Cloud Computing: IEM-ICDC 2021*, pages 97–106. Springer, 2022. https://doi.org/10.1007/978-981-19-1657-1_8.
- [207] Takaaki Inoue, Fukashi Yamamichi, Shinsuke Okada, Shuzo Hamamoto, Masato Fujisawa, and SMART Study Group. Change in irrigation flow through a flexible ureterscope with various devices in the working channel: Comparison between an automatic irrigation pump and gravity-based irrigation. *International Journal of Urology*, 27(4):333–338, 2020. <https://doi.org/10.1111/iju.14197>.
- [208] Waheb A Jabbar, Tan Mei Ting, M Fikri I Hamidun, Ajwad H Che Kamarudin, Wenyan Wu, Jamil Sultan, AbdulRahman A Alsewari, and Mohammed AH Ali. Development of lorawan-based iot system for water quality monitoring in rural areas. *Expert systems with applications*, 242:122862, 2024. <https://doi.org/10.1016/j.eswa.2023.122862>.
- [209] Xu Annie Wang, Julie Tang, and Mark Whitty. Side-view apple flower mapping using edge-based fully convolutional networks for variable rate chemical thinning. *Computers and Electronics in Agriculture*, 178:105673, 2020.
- [210] Surili Agarwal, Neha Bhangale, Kameya Dhanure, Shreeya Gavhane, VA Chakkarwar, and MB Nagori. Application of colorimetry to determine soil fertility through naive bayes classification algorithm. In *2018 9th International conference on computing, communication and networking technologies (ICCCNT)*, pages 1–6. Ieee, 2018.
- [211] Muh Syahlan Natsir, Elyas Palantei, Ingrid Nurtanio, and Ansar Suyuti. Classification of soil fertility level based on texture with convolutional neural network (cnn) algorithm. In *2023 5th International Conference on Cybernetics and Intelligent System (ICORIS)*, pages 1–6. IEEE, 2023.
- [212] Sandeep Kumar Sunori, Lalit Mohan, Mohit Pant, and Pradeep Juneja. Classification of soil fertility using lvq and pnn techniques. In *2023 8th International Conference on Communication and Electronics Systems (ICCES)*, pages 1441–1446. IEEE, 2023.
- [213] M Sujatha and CD Jaidhar. Classification of soil fertility using machine learning-based classifier. In *2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC)*, pages 138–143. IEEE, 2021.
- [214] Rahul Jaiswal. Soil fertility dataset. <https://www.kaggle.com/datasets/rahuljaiswalonkaggle/soil-fertility-dataset>, 2025. Accessed: 2025-04-07.
- [215] Ali Aldrees, Majid Khan, Abubakr Taha Bakheit Taha, and Mujahid Ali. Evaluation of water quality indexes with novel machine learning and shapley additive explanation (shap) approaches. *Journal of Water Process Engineering*, 58:104789, 2024.

- [216] R Jayalakshmi et al. Mining agricultural data to predict soil fertility using ensemble boosting algorithm. *International Journal of Information Communication Technologies and Human Development (IJICTHD)*, 14(1):1–10, 2022.
- [217] Harshiv Chandra, Pranav M Pawar, R Elakkiya, PS Tamizharasan, Raja Muthalagu, and Alavikunhu Panthakkan. Explainable ai for soil fertility prediction. *IEEE Access*, 11: 97866–97878, 2023.
- [218] TG Keerthan Kumar, C Shubha, , and SA Sushma. Random forest algorithm for soil fertility prediction and grading using machine learning. *Int. J. Innov. Technol. Explor. Eng*, 9:1301–1304, 2019.
- [219] Sk Al Zaminur Rahman, Kaushik Chandra Mitra, and SM Mohidul Islam. Soil classification using machine learning methods and crop suggestion based on soil series. In *2018 21st international conference of computer and information technology (ICCIT)*, pages 1–4. IEEE, 2018.
- [220] Somnath Bera, Tanushree Dey, Anwesha Mukherjee, and Rajkumar Buyya. E-cropreco: a dew-edge-based multi-parametric crop recommendation framework for internet of agricultural things. *The Journal of Supercomputing*, 79(11):11965–11999, 2023. doi: <https://doi.org/10.1007/s11227-023-05131-4>.
- [221] Tanushree Dey, Somnath Bera, Lakshman Prasad Latua, Milan Parua, Anwesha Mukherjee, and Debashis De. icrop: An intelligent crop recommendation system for agriculture 5.0. *IEEE Transactions on AgriFood Electronics*, 2024. doi: 10.1109/TAFE.2024.3454109.
- [222] Yashashree Mahale, Nida Khan, Kunal Kulkarni, Shivali Amit Wagle, Preksha Pareek, Ketan Kotecha, Tanupriya Choudhury, and Ashutosh Sharma. Crop recommendation and forecasting system for maharashtra using machine learning with lstm: A novel expectation-maximization technique. *Discover Sustainability*, 5(1):134, 2024. doi: <https://doi.org/10.1007/s43621-024-00292-5>.
- [223] Yaganteeswarudu Akkem, Saroj Kumar Biswas, and Aruna Varanasi. Streamlit-based enhancing crop recommendation systems with advanced explainable artificial intelligence for smart farming. *Neural Computing and Applications*, pages 1–15, 2024. doi: <https://doi.org/10.1007/s00521-024-10208-z>.
- [224] Junjun Lu, Erfu Dai, Yuxin Miao, and Krzysztof Kusnerek. Developing a new active canopy sensor-and machine learning-based in-season rice nitrogen status diagnosis and recommendation strategy. *Field Crops Research*, 317:109540, 2024. doi: <https://doi.org/10.1016/j.fcr.2024.109540>.
- [225] Murali Krishna Senapaty, Abhishek Ray, and Neelamadhab Padhy. A decision support system for crop recommendation using machine learning classification algorithms. *Agriculture*, 14(8):1256, 2024. doi: <https://doi.org/10.3390/agriculture14081256>.
- [226] Md Anisur Rahman, Narayan Ranjan Chakraborty, Abu Sufiun, Sumit Kumar Banshal, and Fowzia Rahman Tajnin. An aiot-based hydroponic system for crop recommendation

- and nutrient parameter monitorization. *Smart Agricultural Technology*, 8:100472, 2024. doi: <https://doi.org/10.1016/j.atech.2024.100472>.
- [227] Gunaganti Sravanthi and Nageswara Rao Moparthi. An efficient iot based crop disease prediction and crop recommendation for precision agriculture. *Cluster Computing*, pages 1–28, 2024. doi: <https://doi.org/10.1007/s10586-023-04246-w>.
- [228] Feyera Liben, Wuletawu Abera, Meklit Tariku Chernet, Mohammed Ebrahim, Amsalu Tilaye, Teklu Erkossa, Degefie Tibebe Degefie, Powell Mponela, Job Kihara, and Lulseged Tamene. Site-specific fertilizer recommendation using data driven machine learning enhanced wheat productivity and resource use efficiency. *Field Crops Research*, 313:109413, 2024. doi: <https://doi.org/10.1016/j.fcr.2024.109413>.
- [229] Chang Xu, Lei Zhao, Haojie Wen, Yiding Zhang, and Lingxian Zhang. A novel cascaded multi-task method for crop prescription recommendation based on electronic medical record. *Computers and Electronics in Agriculture*, 219:108790, 2024. doi: <https://doi.org/10.1016/j.compag.2024.108790>.
- [230] Amit Bhola and Prabhat Kumar. Farm-level smart crop recommendation framework using machine learning. *Annals of Data Science*, pages 1–24, 2024. doi: <https://doi.org/10.1007/s40745-024-00534-3>.
- [231] Mahmoud Y Shams, Samah A Gamel, and Fatma M Talaat. Enhancing crop recommendation systems with explainable artificial intelligence: a study on agricultural decision-making. *Neural Computing and Applications*, 36(11):5695–5714, 2024. doi: <https://doi.org/10.1007/s00521-023-09391-2>.
- [232] Farhad Soleimani Gharehchopogh, Mohammad Namazi, Laya Ebrahimi, and Benyamin Abdollahzadeh. Advances in sparrow search algorithm: a comprehensive survey. *Archives of Computational Methods in Engineering*, 30(1):427–455, 2023. doi: <https://doi.org/10.1007/s11831-022-09804-w>.
- [233] Mojtaba Ghasemi, Ebrahim Akbari, Abolfazl Rahimnejad, Seyed Ehsan Razavi, Sahand Ghavidel, and Li Li. Phasor particle swarm optimization: a simple and efficient variant of pso. *Soft Computing*, 23:9701–9718, 2019. doi: <https://doi.org/10.1007/s00500-018-3536-8>.
- [234] Yangchen Lu, Xiaobing Yu, Zhengpeng Hu, and Xuming Wang. Convolutional neural network combined with reinforcement learning-based dual-mode grey wolf optimizer to identify crop diseases and pests. *Swarm and Evolutionary Computation*, 94:101874, 2025. doi: <https://doi.org/10.1016/j.swevo.2025.101874>.
- [235] Chilin Wei, Zhu Li, Delan Zhu, Tong Xu, Zhichao Liang, Yuhan Liu, and Nana Zhao. Regulation of the physicochemical properties of nutrient solution in hydroponic system based on the catboost model. *Computers and Electronics in Agriculture*, 229:109729, 2025. doi: <https://doi.org/10.1016/j.compag.2024.109729>.

- [236] Hadeeqa Afzal, Madiha Amjad, Ali Raza, Kashif Munir, Santos Gracia Villar, Luis Alonso Dzul Lopez, and Imran Ashraf. Incorporating soil information with machine learning for crop recommendation to improve agricultural output. *Scientific Reports*, 15(1): 8560, 2025. doi: <https://doi.org/10.1038/s41598-025-88676-z>.
- [237] Federico Marini and Beata Walczak. Particle swarm optimization (pso). a tutorial. *Chemo-metrics and intelligent laboratory systems*, 149:153–165, 2015.
- [238] Department of Agriculture & Farmers Welfare, Government of India. Soil health card portal. Online, 2025. URL <https://soilhealth.dac.gov.in/home>. Accessed: 1 March 2025.
- [239] ICAR-Indian Institute of Soil Science (ICAR-IISS). Four decades of stcr research - crop wise recommendations. Online PDF, 2025. URL <https://iiss.icar.gov.in/downloads/stcr%20Crop%20wise%20Recommendations.pdf>. Accessed: 1 March 2025.
- [240] Ruma Das, Abhirup Paria, and PK Giri. Machine learning framework for selective and sensitive metal ion sensing with nitrogen-doped graphene quantum dots heterostructure. *Carbon*, page 119800, 2024.
- [241] Siti Nurul Afiah Mohd Johari, Siti Khairunniza-Bejo, Abdul Rashid Mohamed Shariff, Nur Azuan Husin, Mohamed Mazmira Mohd Masri, and Noorhazwani Kamarudin. Effect of datasets size on the machine learning performance of the bagworm, metisa plana (walker) infestation using uav remote sensing. *Journal of Plant Diseases and Protection*, 132(1):1–17, 2025. doi: <https://doi.org/10.1007/s41348-024-01020-x>. URL <https://link.springer.com/article/10.1007/s41348-024-01020-x>.
- [242] Mohammed A Awadallah, Mohammed Azmi Al-Betar, Iyad Abu Doush, Sharif Naser Makhadmeh, and Ghazi Al-Naymat. Recent versions and applications of sparrow search algorithm. *Archives of Computational Methods in Engineering*, 30(5):2831–2858, 2023.
- [243] DeBao Chen and ChunXia Zhao. Particle swarm optimization with adaptive population size and its application. *Applied Soft Computing*, 9(1):39–48, 2009.
- [244] Shahrzad Saremi, Seyedeh Zahra Mirjalili, and Seyed Mohammad Mirjalili. Evolutionary population dynamics and grey wolf optimizer. *Neural Computing and Applications*, 26(5): 1257–1263, 2015.
- [245] Seema Rameshwarrao Wankhade and Anjali B Raut. Development of a model for estimation of soil parameters using deep learning. *International Journal of Information Technology*, pages 1–17, 2024. doi: <https://doi.org/10.1007/s41870-023-01675-z>.
- [246] PSS Gopi and M Karthikeyan. Red fox optimization with ensemble recurrent neural network for crop recommendation and yield prediction model. *Multimedia Tools and Applications*, 83(5):13159–13179, 2024. doi: <https://doi.org/10.1007/s11042-023-16113-2>.
- [247] Mohamed Omar Abdullahi, Abdukadir Dahir Jimale, Yahye Abukar Ahmed, and Abdulaziz Yasin Nageye. Revolutionizing somali agriculture: harnessing machine learning and iot for optimal crop recommendations. *Discover Applied Sciences*, 6(3):77, 2024. doi: <https://doi.org/10.1007/s42452-024-05739-y>.

- [248] F. N. Ortatas, U. Ozkaya, M. E. Sahin, and et al. Sugar beet farming goes high-tech: a method for automated weed detection using machine learning and deep learning in precision agriculture. *Neural Comput & Applic*, 36:4603–4622, 2024. doi: <https://doi.org/10.1007/s00521-023-09320-3>.
- [249] Emily Sullivan. Understanding from machine learning models. *The British Journal for the Philosophy of Science*, 2022.
- [250] B. Keswani, A. G. Mohapatra, A. Mohanty, and et al. Adapting weather conditions based iot enabled smart irrigation technique in precision agriculture mechanisms. *Neural Comput & Applic*, 31(Suppl 1):277–292, 2019. doi: <https://doi.org/10.1007/s00521-018-3737-1>.
- [251] Patrick Filippi, Si Yang Han, and Thomas FA Bishop. On crop yield modelling, predicting, and forecasting and addressing the common issues in published studies. *Precision Agriculture*, 26(1):1–19, 2025. doi: <https://doi.org/10.1007/s11119-024-10212-2>.
- [252] RNV Jagan Mohan, Pravallika Sree Rayanoothala, and R Praneetha Sree. Next-gen agriculture: integrating ai and xai for precision crop yield predictions. *Frontiers in Plant Science*, 15:1451607, 2025.
- [253] Md Toukir Ahmed, Md Wadud Ahmed, and Mohammed Kamruzzaman. A systematic review of explainable artificial intelligence for spectroscopic agricultural quality assessment. *Computers and Electronics in Agriculture*, 235:110354, 2025.
- [254] Daniel Enemona Mathew, Deborah Uzoamaka Ebem, Anayo Chukwu Ikegwu, Pamela Eberechukwu Ukeoma, and Ngozi Fidelia Dibiazue. Recent emerging techniques in explainable artificial intelligence to enhance the interpretable and understanding of ai models for human. *Neural Processing Letters*, 57(1):16, 2025.
- [255] Masahiro Ryo, Boyan Angelov, Stefano Mammola, Jamie M Kass, Blas M Benito, and Florian Hartig. Explainable artificial intelligence enhances the ecological interpretability of black-box species distribution models. *Ecography*, 44(2):199–205, 2021.
- [256] Antonios Mamalakis, Elizabeth A Barnes, and Imme Ebert-Uphoff. Investigating the fidelity of explainable artificial intelligence methods for applications of convolutional neural networks in geoscience. *Artificial Intelligence for the Earth Systems*, 1(4):e220012, 2022.
- [257] Hakan Başağaoğlu, Debaditya Chakraborty, Cesar Do Lago, Lilianna Gutierrez, Mehmet Arif Şahinli, Marcio Giacomoni, Chad Furl, Ali Mirchi, Daniel Moriasi, and Sema Sevinç Şengör. A review on interpretable and explainable artificial intelligence in hydroclimatic applications. *Water*, 14(8):1230, 2022.
- [258] Debjani Sihi, Biswanath Dari, Abraham Peedikayil Kuruvila, Gaurav Jha, and Kanad Basu. Explainable machine learning approach quantified the long-term (1981–2015) impact of climate and soil properties on yields of major agricultural crops across conus. *Frontiers in Sustainable Food Systems*, 6:847892, 2022.
- [259] Pooja Goyal, Rakesh Sharda, Mukesh Saini, and Mukesh Siag. A deep learning approach for early detection of drought stress in maize using proximal scale digital images. *Neural Computing and Applications*, 36(4):1899–1913, 2024.

- [260] Abhirup Paria and Ruma Das. Implementation of iot and machine learning techniques in smart irrigation systems. In *IoT Sensors, ML, AI and XAI: Empowering A Smarter World*, pages 143–151. Springer, 2024.
- [261] Sadia Kamal, Parth Sharma, PK Gupta, Mohammad Khubeb Siddiqui, Ankush Singh, and Abhijit Dutt. Dvtxai: a novel deep vision transformer with an explainable ai-based framework and its application in agriculture. *The Journal of Supercomputing*, 81(1):1–32, 2025.
- [262] Ying Wang, Pengxin Wang, Kevin Tansey, Junming Liu, Bethany Delaney, and Wenting Quan. An interpretable approach combining shapley additive explanations and lightgbm based on data augmentation for improving wheat yield estimates. *Computers and Electronics in Agriculture*, 229:109758, 2025.
- [263] Nisha P Shetty, Balachandra Muniyal, Ketavarapu Sriyans, Kunyalik Garg, Shiv Pratap, Aman Priyanshu, and Dhruthi Kumar. Counterfactual based approaches for feature attributions of stress factors affecting rice yield. *Engineering Reports*, 7(1):e13085, 2025.
- [264] Thilina Abekoon, Hirushan Sajindra, Namal Rathnayake, Imesh U Ekanayake, Anuradha Jayakody, and Upaka Rathnayake. A novel application with explainable machine learning (shap and lime) to predict soil n, p, and k nutrient content in cabbage cultivation. *Smart Agricultural Technology*, 11:100879, 2025.
- [265] Andrea Cartolano, Alfredo Cuzzocrea, and Giovanni Pilato. Analyzing and assessing explainable ai models for smart agriculture environments. *Multimedia Tools and Applications*, 83(12):37225–37246, 2024.
- [266] Ivan Malashin, Vadim Tynchenko, Andrei Gantimurov, Vladimir Nelyub, Aleksei Borodulin, and Yadviga Tynchenko. Predicting sustainable crop yields: Deep learning and explainable ai tools. *Sustainability*, 16(21):9437, 2024.
- [267] Jie Pei, Shaofeng Tan, Yaopeng Zou, Chunhua Liao, Yinan He, Jian Wang, Huabing Huang, Tianxing Wang, Haifeng Tian, Huajun Fang, et al. The role of phenology in crop yield prediction: Comparison of ground-based phenology and remotely sensed phenology. *Agricultural and Forest Meteorology*, 361:110340, 2025.
- [268] Food and Agriculture Organization of the United Nations. FAO Home, 2025. URL <http://www.fao.org/home/en/>. Accessed: 2025-01-12.
- [269] The World Bank Group. World Development Indicators Database. <https://data.worldbank.org/>, 2025. URL <https://data.worldbank.org/>. Accessed: 2025-01-12.
- [270] International Crops Research Institute for the Semi-Arid Tropics (ICRISAT). Digital Library on Crop, Soil, and Water Management Practices in Semi-Arid Tropics. Online, 2025. URL <http://data.icrisat.org/dld/src/about-dld.html>. Accessed: 2025-01-12.
- [271] Ruma Das, Abhirup Paria, and PK Giri. Machine learning framework for selective and sensitive metal ion sensing with nitrogen-doped graphene quantum dots heterostructure. *Carbon*, 232:119800, 2025.

- [272] Vijay Mahore, Peeyush Soni, Arpita Paul, Prakhar Patidar, and Rajendra Machavaram. Machine learning-based draft prediction for mouldboard ploughing in sandy clay loam soil. *Journal of Terramechanics*, 111:31–40, 2024. ISSN 0022-4898. doi: <https://doi.org/10.1016/j.jterra.2023.09.002>. URL <https://www.sciencedirect.com/science/article/pii/S0022489823000836>.
- [273] Ravi Shekhar Tiwari. Hate speech detection using lstm and explanation by lime (local interpretable model-agnostic explanations). In *Computational Intelligence Methods for Sentiment Analysis in Natural Language Processing Applications*, pages 93–110. Elsevier, 2024.
- [274] Javier Del Ser, Alejandro Barredo-Arrieta, Natalia Díaz-Rodríguez, Francisco Herrera, Anna Saranti, and Andreas Holzinger. On generating trustworthy counterfactual explanations. *Information Sciences*, 655:119898, 2024.
- [275] Dunwang Qin, Zhen Peng, and Lifeng Wu. Interpretable predictive modeling of non-stationary long time series. *Computers & Industrial Engineering*, 194:110412, 2024.
- [276] Pengfei Yu and Xuesong Yan. Stock price prediction based on deep neural networks. *Neural Computing and Applications*, 32(6):1609–1628, 2020.
- [277] R John Martin, Ruchi Mittal, Varun Malik, Fathe Jeribi, Shams Tabrez Siddiqui, Mohammad Alamgir Hossain, and SL Swapna. Xai-powered smart agriculture framework for enhancing food productivity and sustainability. *IEEE Access*, 2024.
- [278] Sita Rani, Amit Kumar Mishra, Aman Kataria, Saurav Mallik, and Hong Qin. Machine learning-based optimal crop selection system in smart agriculture. *Scientific Reports*, 13(1):15997, 2023.
- [279] G Mariammal, A Suruliandi, SP Raja, and E Poongothai. Prediction of land suitability for crop cultivation based on soil and environmental characteristics using modified recursive feature elimination technique with various classifiers. *IEEE transactions on computational social systems*, 8(5):1132–1142, 2021.
- [280] Sahameh Shafiee, Lars Martin Lied, Ingunn Burud, Jon Arne Dieseth, Muath Alsheikh, and Morten Lillemo. Sequential forward selection and support vector regression in comparison to lasso regression for spring wheat yield prediction based on uav imagery. *Computers and Electronics in Agriculture*, 183:106036, 2021.
- [281] Jasmin Heilemann, Christian Klassert, Luis Samaniego, Stephan Thober, Andreas Marx, Friedrich Boeing, Bernd Klauer, and Erik Gawel. Projecting impacts of extreme weather events on crop yields using lasso regression. *Weather and Climate Extremes*, 46:100738, 2024.
- [282] Virendra Singh, HM Agrawal, GC Joshi, M Sudershan, and AK Sinha. Elemental profile of agricultural soil by the edxrf technique and use of the principal component analysis (pca) method to interpret the complex data. *Applied Radiation and Isotopes*, 69(7):969–974, 2011.

- [283] S áP Raja, Barbara Sawicka, Zoran Stamenkovic, and G Mariammal. Crop prediction based on characteristics of the agricultural environment using various feature selection techniques and classifiers. *IEEe Access*, 10:23625–23641, 2022.

Abhinav Paria 14-07-2026

Sarmistha Neogy 14.07.2026

Professor
Computer Sc. & Engg. Department
Jadavpur University
Kolkata-700032

Arindam Guu
14.07.26
Associate Professor
Dept. of Computer Science & Engineering
HALDIA INSTITUTE OF TECHNOLOGY