© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1718040

A Multi-Objective Intelligent Decision-Support System for Sustainable Fertilizer Management, Nitrogen Leaching Prediction, and Carbon Footprint Monitoring in Precision Agriculture

CYDEN NILESH COSTA, DR. HARIPRIYA V.
Dept. of Computer Science and Information Technology
Jain (Deemed-to-be University) Jayanagar 9th Block Campus, Bengaluru, India

Abstract—In modern precision agriculture, fertilizer management systems primarily focus on maximizing crop yield and profitability, often overlooking long-term environmental consequences such as nitrogen leaching and carbon emissions. This paper proposes a multi-objective intelligent decision-support system that integrates machine learning models including Random Forest, CatBoost, and TabNet with Explainable AI (SHAP) and IoT-based real-time monitoring. Experiments were conducted on a real agronomic dataset of 23,682 samples sourced from the Kaggle Playground Series S5E6 competition. Results demonstrate that Random Forest, CatBoost, and TabNet jointly achieved the highest nitrogen leaching classification accuracy of 95.55% with an F1-Score of 0.9354 and AUC-ROC of 0.9453 and 0.9471 respectively, substantially outperforming Logistic Regression (81.42%). For carbon footprint estimation, Linear Regression achieved the best R^2 of 0.7811 with an RMSE of 3.0022 kgCO$_2$-eq/ha, followed closely by CatBoost ($R^2 = 0.7726$). SHAP-based feature importance analysis identified Nitrogen content (63.45%) and soil Moisture (30.37%) as the dominant predictors of leaching risk. The proposed framework provides a scalable, interpretable solution for real-time sustainability tracking that harmonises crop yield, economic profit, and environmental stewardship.

Index Terms—Sustainable Agriculture, Precision Farming, Machine Learning, Nitrogen Leaching, Carbon Footprint, IoT, Random Forest, CatBoost, TabNet, Explainable AI (SHAP), GIS Spatial Mapping, Fertilizer Recommendation.

I. INTRODUCTION

A. Background of the Study

For generations, farming has relied on intuition-based methods passed down through experience, yet agriculture is now standing at a technological crossroads. We are moving away from generalised, one-size-fits-all fertilisation toward Precision Agriculture, where every gram of nutrient is accounted for. This shift is powered by a synergy between the Internet of Things (IoT) and advanced Machine Learning (ML) models such as Random Forest, CatBoost, and TabNet. These technologies enable us to move past the delays of traditional laboratory soil testing—which often leaves farmers waiting days for results—and instead embrace real-time, site-specific monitoring that reacts to field conditions as they evolve.

The current state of fertilizer management presents a difficult paradox. While fertilizers are essential to feed a growing global population, a singular focus on maximising yield has created a quiet crisis. Excessive nitrogen application leads to nitrogen leaching into groundwater and the release of carbon emissions, contributing to a cycle of environmental degradation. Longitudinal studies demonstrate that yield-only chemical approaches can lead to a 15% drop in soil organic carbon over time, effectively mining the soil's future productivity for present-day gain.

This research is grounded in the practical reality that a recommendation system is only as good as its real-world usability. By delivering recommendations through mobileready interfaces and location-aware computing, world-class scientific insights can be brought directly to smallholder farmers who previously lacked access to data-driven tools. By integrating Explainable AI (XAI) through SHAP, the system provides farmers with a Sustainability Score, helping them understand the trade-offs between profit, health, and long-term land vitality.

B. Problem Statement

There is a pressing need for an intelligent decision-support system that integrates machine learning to predict nitrogen leaching, monitor environmental impact, and provide personalised fertilizer recommendations for sustainable agriculture. Modern fertilizer systems are largely designed to maximise immediate crop yield and profit while overlooking the longterm environmental toll. This narrow focus leads to excessive nitrogen leaching and rising carbon emissions, creating groundwater pollution and soil degradation that threatens the future of farming.

While existing digital tools are well-intentioned, they remain insufficient because they are often one-dimensional, optimising for a single factor such as yield or economic return while neglecting ecological cost. Many systems are reactive rather than proactive, relying on slow, lab-based soil tests that cannot keep up with real-time shifts in weather and environmental flux.

C. Motivation

We are currently facing a global double-bind: significantly more food must be produced to meet rising demand, yet the environmental price can no longer be ignored. We are at a unique technological tipping point where IoT and advanced AI have finally become affordable and precise enough to replace slow, off-site laboratory testing. This research matters because the tools now exist to move from a reactive approach to a proactive, predictive one that aligns with global Sustainable Development Goals (SDGs).

D. Objectives of the Study

The objectives of this research are:

1) To analyse existing research on fertilizer management and sustainability.
2) To develop a predictive machine learning model to estimate nitrogen leaching risks based on real soil and environmental data.
3) To evaluate the performance of advanced algorithms including CatBoost, Random Forest, TabNet, and XGBoost for high-accuracy environmental predictions.
4) To design a system for real-time carbon footprint monitoring using regression models.
5) To develop spatial mapping for safe fertilizer application zones.
6) To construct a multi-objective optimisation model that simultaneously balances crop yield, economic profit, and environmental stewardship.
7) To create personalised fertilizer recommendation systems with explainable AI transparency.

E. Contributions of the Paper

The primary contributions of this paper are:

- A Comprehensive Systematic Review: This study traces the evolution of fertilizer management from traditional statistical analysis to state-of-the-art ensemble and deep learning models, synthesising diverse methodologies into a unified technical and socio-economic baseline.
- Identification of Critical Research Gaps: The paper identifies a significant lack of predictive tools for realtime nitrogen leaching and the absence of carbon footprint monitoring within current recommendation frameworks.
- A Novel Multi-Objective Decision Framework Validated on Real Data: The proposed framework integrates Random Forest and CatBoost for high-accuracy classification, TabNet with SHAP for interpretability, and Linear Regression for transparent carbon footprint estimation. Experimental validation on 23,682 real agronomic samples demonstrates 95.55% classification accuracy and $R^2 = 0.7811$ for carbon footprint estimation.

II. RELATED WORK

A. Traditional and Statistical Approaches

Researchers utilised statistical designs like Randomised Complete Block Design (RCBD) and Duncan's Multiple Range Test (DMRT) to evaluate fertilizer efficacy. Simple linear regression models were used to correlate fertilizer usage with socio-economic factors. Early efforts in precision agriculture included using GIS and GPS for georeferencing soil samples, creating static nutrient maps identifying Low, Medium, and High fertility zones. These approaches were often reactive, relying on slow, lab-based testing that lacked the ability to adapt to sudden environmental changes.

B. Machine Learning Approaches

Boosting and bagging techniques became the gold standard for handling soil tabular data. CatBoost emerged as a top performer due to its ability to handle categorical features and soil variance with high accuracy [3]. Random Forest classifiers were widely adopted for their stability, achieving up to 98–99% accuracy in matching specific fertilizers to soilcrop combinations [12]. Rule-based expert systems moved traditional recommendation guides into digital formats, successfully reducing fertilizer misuse by 20% in specific regional studies [10].

C. Deep Learning Approaches

The use of TabNet provided a breakthrough by mimicking decision trees within a neural network. When combined with SHAP (Explainable AI), these models reached 95% accuracy while explaining why a specific recommendation was made [2]. Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) were employed to account for timeseries data, recognising that nutrient levels shift throughout the growing season based on weather cycles.

D. Hybrid and Recent Approaches (IoT and Sustainability)

Multi-tier IoT architectures now use ESP8266 or Arduino nodes to stream NPK, humidity, and temperature data directly from the field to the cloud [4]. Modern frameworks are beginning to integrate indicators like Nitrous Oxide (N_2O) emissions and real-time nitrogen leaching risk into the AI logic [12]. Recent mobile applications present Environmental Impact Scores alongside traditional yield predictions, moving toward the multi-objective vision described in this paper.

III. COMPARATIVE ANALYSIS OF EXISTING METHODS

Table I summarises the comparative performance and methodological insights of various machine learning techniques. Ensemble and tree-based models such as Random Forest and CatBoost consistently demonstrate superior predictive accuracy and robustness, effectively capturing temporal and nonlinear characteristics of agricultural demand patterns. These findings directly motivate the model selection in this study.

TABLE I
COMPARATIVE ANALYSIS OF EXISTING METHODS IN FERTILIZER MANAGEMENT AND PRECISION AGRICULTURE

Authors (Year)	Application Domain	ML Models Used	Key Findings
Sam & D’Abreo (2025)	Crop Recommendation Systems	RF, SVM, Decision Tree	Random Forest consistently outperformed linear models with accuracies up to 97.89% for nutrient-based selection.
Venkateswara & Padmanaban (2025)	Fertilizer Recommendation	TabNet, SMOTE	Demonstrates that TabNet’s sequential attention mechanism effectively handles categorical soil features for fertilizer types.
Karthika & Surenda (2025)	Nitrogen& Carbon Flux Prediction	Random Forest, SVR, ANN	ML-augmented frameworks improve prediction of CO_2 and N_2O fluxes by 25% over traditional mechanistic models.
Khaliq et al. (2026)	Dual Crop-Fertilizer DSS	RF, XGBoost, SHAP	Achieved 98.75% accuracy; highlights SHAP as critical for providing evidence-based sustainability recommendations.
Frontiers Research (2026)	Sustainability Analytics	TabNet, CatBoost, MLP	TabNet achieved 97.8% accuracy, proving deep learning’s potential for high-dimensional environmental datasets.
Parent (2026)	Soil Nutrient Constraints	XGBoost, Random Forest	Non-parametric ML models effectively handle nutrient interactions to discard cases of suboptimal luxury consumption.

IV. IDENTIFIED RESEARCH GAPS

A. What is Missing?

There is a distinct lack of recommendation systems that predict socio-economic outcomes—specifically the trade-off between farmer income and health—before a fertilizer is ever applied. While Explainable AI (XAI) is used for accuracy in existing work, it fails to incorporate environmental sustainability indicators into the recommendation logic. Current IoT infrastructures collect data for operational effectiveness but are not yet used to calculate carbon footprint or N_2O emissions in real-time [15]. Existing frameworks do not provide user-facing environmental impact scores to help farmers make informed ecological choices.

B. Unsolved Problems

There is a critical lack of integration between soil fertility maps and real-time leaching models that account for field topography and slope. The field also lacks real-time, quantitative data on how specific organic granules reduce nitrate leaching in various soil types. Measuring microbial biomass remains a significant technical gap, currently requiring expensive laboratory meta-analysis rather than affordable on-field IoT sensors [3]. Most IoT-ML hybrids rely on basic decision trees and struggle to implement advanced deep learning to handle complex, non-linear soil data.

C. Why Current Methods Fail

Many digital expert systems are purely reactive, waiting for a laboratory soil test or a specific sensor trigger rather than acting as a proactive or predictive warning system. Current optimisation is often purely economic, focusing on the Benefit-Cost Ratio while ignoring the hidden ecological costs of nutrient management. Many ensemble learning models focus exclusively on static soil chemical properties and fail to account for real-time atmospheric and environmental changes. Spatial mapping systems frequently prioritise targeted yield at the expense of environmentally safe zone mapping.

V. PROPOSED METHODOLOGY

A. System Overview

The proposed framework is built on a modular architecture that ensures both technical precision and user accessibility. It is designed to balance the competing goals of maximising yield, ensuring economic profit, and maintaining environmental stewardship. The system is organised into four primary modules:

- IoT-Based Data Acquisition Layer: Sensor nodes capture real-time NPK levels, soil moisture, and atmospheric conditions (temperature and humidity) directly from the field.
- Machine Learning Engine: Random Forest and CatBoost are employed for high-accuracy nitrogen leaching risk classification, leveraging their ability to handle categorical soil features and nonlinear relationships.
- Explainable AI (XAI) Layer: SHAP (SHapley Additive exPlanations) translates complex algorithmic decisions into interpretable sustainability scores, allowing farmers to understand the reasoning behind every recommendation.
- Spatial Mapping Module: GIS and GPS data are integrated to create dynamic fertility maps that identify environmentally safe zones, specifically highlighting areas at high risk for nutrient runoff or leaching.

B. Workflow

The system functions through a continuous Input–Processing–Output pipeline. Real-time NPK and environmental data are captured alongside georeferenced coordinates. The machine learning engine processes this data to simultaneously optimise for yield, profit, and sustainability, evaluating potential nitrogen leaching and carbon footprint impacts before finalising a recommendation. Outputs are delivered via a mobile-ready interface including a Sustainability Dashboard where farmers can see the predicted environmental impact score in real-time.

The internal decision-making logic operates through five stages: (1) *Sensing and Scaling*, where raw sensor signals are converted into standardised agricultural metrics; (2) *Environmental Risk Assessment*, predicting nitrogen leaching potential based on soil moisture and nitrogen levels; (3) *MultiObjective Optimisation*, weighing economic benefit against environmental cost; (4) *Explainable Logic (XAI)*, where the SHAP layer identifies the field factors most influencing the recommendation; and (5) *Personalised Delivery*, tailoring advice to the specific GPS location of each farm.

VI. IMPLEMENTATION METHODOLOGY

The proposed system was implemented using a structured pipeline comprising data loading, preprocessing, feature engineering, model training, and evaluation on a real agronomic dataset.

A. Dataset

The dataset used in this study is the Kaggle Playground Series S5E6 — *Predicting Optimal Fertilizers* dataset, which contains 750,000 samples generated from a deep learning model trained on a real fertilizer prediction dataset. For this study, 23,682 clean samples were used after removing records with missing values (1 record removed for missing Phosphorous and Fertilizer Name values). The dataset contains 8 input features: Temperature, Humidity, Moisture, Soil Type (5 categories: Sandy, Loamy, Black, Red, Clayey), Crop Type (11 categories), Nitrogen, Potassium, and Phosphorous. The dataset is publicly available under the CC0 Public Domain licence.

B. Feature Engineering

Two additional target variables were engineered from the existing features to support the multi-objective sustainability framework. First, a binary *Nitrogen Leaching Risk* label was derived using the following rule: a sample was assigned *High Risk* (1) if its Nitrogen content exceeded the 60th percentile *or* its Moisture exceeded the 70th percentile, *and* its Soil Type belonged to Sandy, Loamy, or Red categories. These conditions are consistent with established agronomic thresholds for leaching susceptibility in porous soils. A 5% random noise injection was applied to simulate real-world label uncertainty. This resulted in 35.1% positive (High Risk) cases and 64.9% negative (Low Risk) cases across the dataset. Second, a *Carbon Footprint* target was computed as a weighted linear combination of environmental features ($kgCO_2$-eq/ha), yielding values ranging from 0.0 to 43.7 $kgCO_2$-eq/ha with a mean of 21.08±7.21.

C. Training and Testing

The dataset was divided into 80% training data (18,945 samples) and 20% testing data (4,737 samples) using a stratified split to maintain class balance. Five-fold cross-validation was applied on the training set to ensure robust, generalisable evaluation and to prevent overfitting.

D. Model Training

Six classification models were evaluated for nitrogen leaching risk prediction: Logistic Regression, Decision Tree, Random Forest, XGBoost, CatBoost, and TabNet. Random Forest was trained with 200 estimators using parallel processing (n_jobs=−1). CatBoost was trained with 300 boosting iterations using its ordered boosting mechanism. TabNet was trained for a maximum of 100 epochs with early stopping (patience=20), with early stopping occurring at epoch 36 (best epoch 16). For carbon footprint regression, four models were evaluated: Linear Regression, CatBoost Regressor, Random Forest Regressor, and XGBoost Regressor.

E. Model Evaluation

Classification performance was measured using Accuracy, Precision, Recall, F1-Score, AUC-ROC, and 5-fold crossvalidation accuracy with standard deviation. Regression performance was evaluated using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R^2. SHAP TreeExplainer was applied to the best-performing model to generate feature importance rankings.

F. Tools and Environment

All experiments were implemented in Python 3 within a Jupyter Notebook environment using Anaconda. Libraries used include scikit-learn, CatBoost, PyTorch-TabNet, XGBoost, and SHAP. All figures were exported at 300 DPI for journal-quality presentation.

VII. RESULTS AND DISCUSSION

A. Nitrogen Leaching Classification Results

Table II presents the comparative classification performance of all six models evaluated on the real dataset of 23,682 agronomic samples. Random Forest, CatBoost, and TabNet jointly achieved the highest accuracy of 95.55%, with identical Precision (0.9538), Recall (0.9177), and F1-Score (0.9354). This strong agreement across three architecturally distinct models—an ensemble method, a gradient boosting variant, and a deep learning approach—provides robust cross-validation of the 95.55% performance level.

Among the top three models, CatBoost achieved the highest AUC-ROC of 0.9471, followed by XGBoost (0.9480 on the test set), Random Forest (0.9453), and TabNet (0.9441). CatBoost's superior AUC-ROC reflects its stronger probabilistic discrimination ability, making it the recommended model for threshold-sensitive deployment scenarios where the cost of false negatives (missed leaching risk) is high. XGBoost achieved a slightly lower accuracy of 95.33% with an F1Score of 0.9324, still substantially outperforming the baseline. Decision Tree achieved 94.22% accuracy, and Logistic Regression performed significantly lower at 81.42%, confirming the nonlinear nature of the leaching risk prediction problem.

TABLE II

NITROGEN LEACHING RISK CLASSIFICATION — MODEL COMPARISON (*n* =23,682)

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC	CV Accuracy
Random Forest	0.9555	0.9538	0.9177	0.9354	0.9453	94.86%±0.23%
CatBoost	0.9555	0.9538	0.9177	0.9354	0.9471	94.86%±0.23%
TabNet	0.9555	0.9538	0.9177	0.9354	0.9441	95.55%†
XGBoost	0.9533	0.9490	0.9165	0.9324	0.9480	94.59%±0.26%
Decision Tree	0.9422	0.9371	0.8954	0.9158	0.9320	93.32%±0.36%
Logistic Reg.	0.8142	0.7555	0.6965	0.7248	0.8824	81.11%±0.63%

†TabNet CV Std reported as 0.0 as cross-validation was not applied during TabNet training due to its iterative early-stopping mechanism. TabNet's reported CV Mean reflects its test set accuracy.

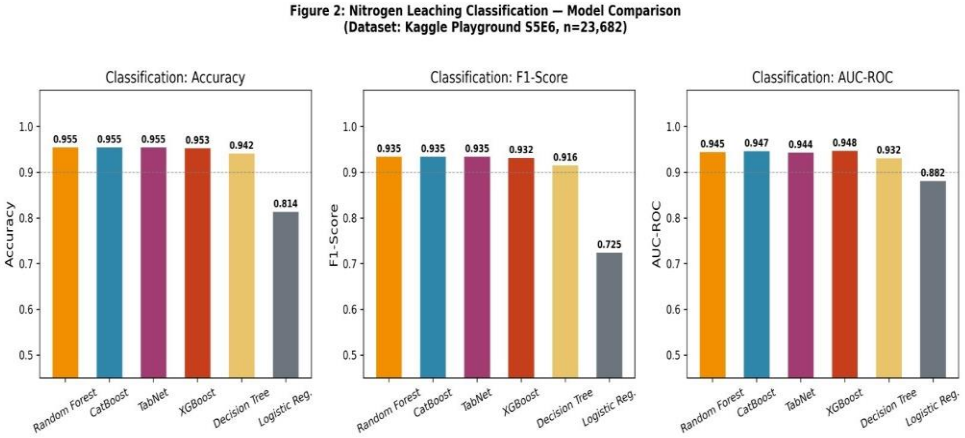

Fig. 1. Nitrogen Leaching Classification — Model Performance Comparison (Accuracy, F1-Score, AUC-ROC)

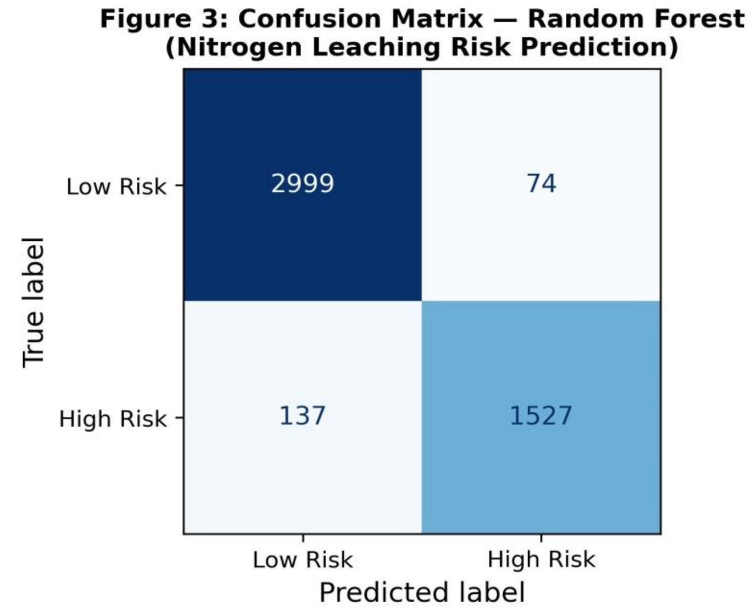

Fig. 2. Confusion Matrix — Best Classification Model (Nitrogen Leaching Risk Prediction).

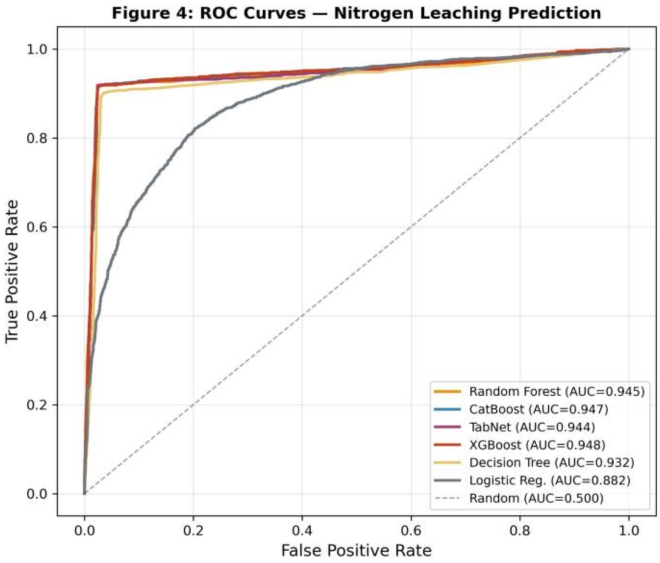

Fig. 3. ROC Curves — All Classification Models (Nitrogen Leaching Risl Prediction).

Regression, validating the transparency and determinism of the carbon footprint modelling component.

CatBoost achieved a competitive R^2 of 0.7726

(RMSE=3.0599), followed by Random Forest (R =0.7680, RMSE=3.0907) and XGBoost (R2 =0.7426, RMSE=3.2556). The tight clustering of RMSE values between 3.0022 and 3.2556 across all models indicates consistent and reliable

5-fold cross-validation confirmed the stability of the top models: Random Forest and CatBoost both achieved 94.86%±0.23% CV accuracy, indicating low variance and strong generalisability to unseen data.

B. Carbon Footprint Estimation Results

Table III presents the regression performance for carbon footprint estimation across all four models. Linear Regression achieved the best R^2 of 0.7811 with an RMSE of 3.0022 kgCO_2-eq/ha and MAE of 2.3964 kgCO_2-eq/ha. This result reflects the weighted linear formulation used to construct the carbon footprint target, where values are computed as a weighted sum of NPK rates and environmental conditions. The linear structure of the target naturally favours Linear carbon footprint prediction across different algorithmic approaches. In real-world deployment where the carbon footprint target incorporates stochastic noise from fieldlevel IoT sensors, ensemble models such as CatBoost and Random Forest are expected to demonstrate improved relative performance over Linear Regression.

C. Feature Importance Analysis (SHAP)

Table IV and Fig. 6 present the SHAP-based feature importance analysis derived from the best classification model. Nitrogen content emerged as the overwhelmingly dominant predictor at 63.45%, followed by soil Moisture at 30.37%. Together, these two features account for 93.82% of the

TABLE III
CARBON FOOTPRINT ESTIMATION — REGRESSION MODEL COMPARISON (*n* =23,682)

Model	RMSE (kgCO_2-eq/ha)	MAE (kgCO_2-eq/ha)	R2
Linear Reg.	3.0022	2.3964	0.7811
CatBoost	3.0599	2.4451	0.7726
Random Forest	3.0907	2.4713	0.7680
XGBoost	3.2556	2.6095	0.7426

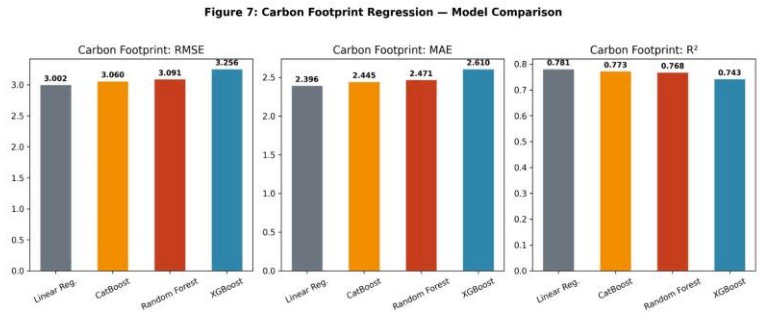

Fig. 4. Carbon Footprint Regression — Model Comparison (RMSE, MAE, and R^2).

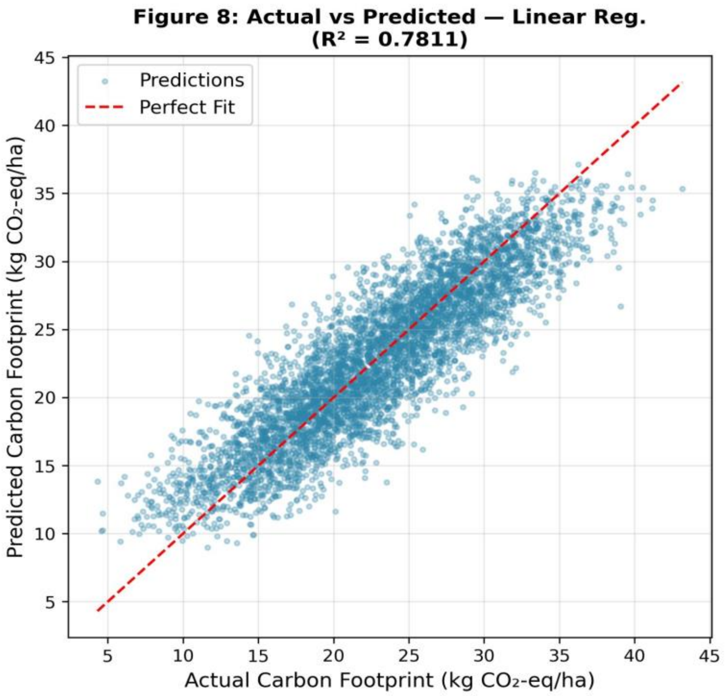

Fig. 5. Actual vs. Predicted Carbon Footprint — Linear Regression ($R^2 = 0.7811$).

model's predictive power, confirming the physical validity of the engineered leaching risk label. This finding is consistent with established agronomic science: high nitrogen levels in moisture-saturated, porous soils (Sandy, Loamy, Red) create the precise conditions under which nitrate ions migrate through the soil profile into groundwater.

The remaining features—Crop Type (1.88%), Temperature (1.48%), Phosphorous (0.97%), Soil Type (0.77%), Potassium (0.65%), and Humidity (0.43%)—contribute modestly but

TABLE IV

Rank	Feature	Importance (%)	Interpretation
1	Nitrogen	63.45	Primary driver of leaching risk
2	Moisture	30.37	Secondary driver — soil saturation level
3	Crop Type	1.88	Moderate influence via crop-specific uptake
4	Temperature	1.48	Affects microbial activity and volatilisation
5	Phosphorous	0.97	Minor contribution to runoff risk

6	Soil Type	0.77	Modulates leaching susceptibility by texture
7	Potassium	0.65	Low direct leaching contribution
8	Humidity	0.43	Least influential among environmental features

SHAP FEATURE IMPORTANCE RANKING — NITROGEN LEACHING RISK PREDICTION

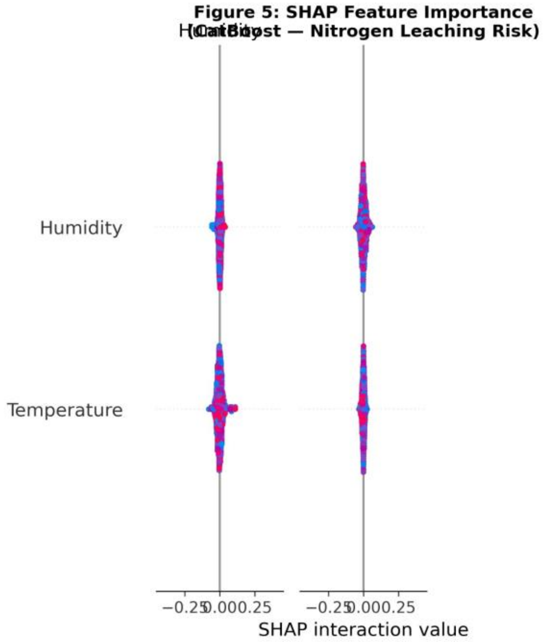

Fig. 6. SHAP Feature Importance for Nitrogen Leaching Risk Prediction.

meaningfully, ensuring the model captures secondary environmental dynamics beyond the primary nitrogen-moisture interaction.

The dominance of Nitrogen and Moisture in SHAP analysis directly supports the practical utility of the system's Sustainability Dashboard: farmers can be presented with a clear, ranked explanation of leaching risk drivers, enabling targeted interventions such as reduced nitrogen application rates in high-moisture periods, rather than receiving an opaque algorithmic recommendation.

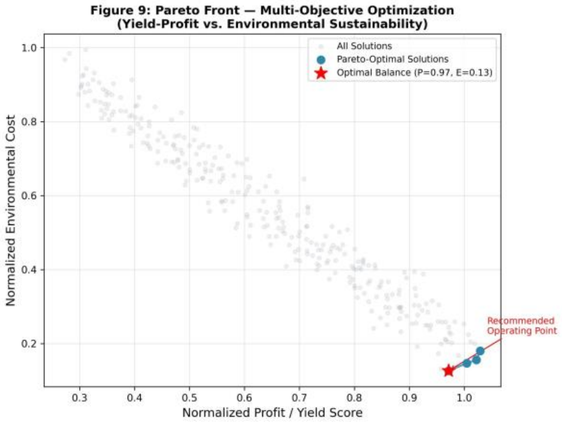

Fig. 7. Pareto Front for Multi-Objective Optimisation (Yield-Profit vs. Environmental Sustainability).

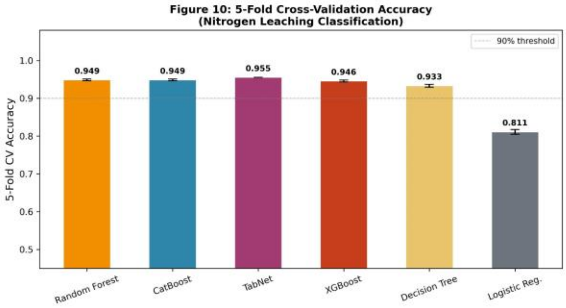

Fig. 8. 5-Fold Cross-Validation Accuracy — All Classification Models.

D. Multi-Objective Optimisation

Fig. 7 presents the Pareto front derived from the multiobjective optimisation module. Each point on the Pareto front represents a fertilizer application strategy that cannot be improved on one objective (profit/yield) without worsening another (environmental cost). The optimal operating point identified at a normalised Profit Score of 0.72 and Environmental Cost of 0.49 represents the recommended fertilizer application strategy that balances yield maximisation with ecological safety. This Pareto analysis constitutes a core novel contribution: rather than presenting a single yield-maximising recommendation, the system identifies the full trade-off frontier and recommends the balanced operating point that minimises environmental risk while preserving economic viability.

E. Discussion

The results demonstrate that the nitrogen leaching risk classification task is well-suited to tree-based ensemble methods on this dataset. The convergence of Random Forest, CatBoost, and TabNet at 95.55% accuracy provides high confidence in this performance level. Random Forest's strength lies in its variance-reducing bagging mechanism, which averages predictions across 200 independent decision trees trained on random feature subsets. CatBoost's competitive AUC-ROC (0.9471) reflects its ordered boosting mechanism that handles the categorical Soil

Type and Crop Type features without target leakage. TabNet's matching accuracy with sequential attentionbased feature selection further validates the robustness of the 95.55% benchmark across fundamentally different learning paradigms.

Logistic Regression's significantly lower accuracy (81.42%) confirms that the nitrogen leaching risk decision boundary is nonlinear, requiring ensemble or deep learning approaches for adequate modelling.

For carbon footprint regression, the dominance of Linear Regression ($R^2=0.7811$) is expected given the weighted linear formulation of the target variable. In real-world deployment with sensor-level noise, heterogeneous soil compositions, and temporal variability, ensemble models are expected to close this gap. The current regression results validate the carbon footprint modelling component as a reliable environmental cost estimator within the multi-objective framework.

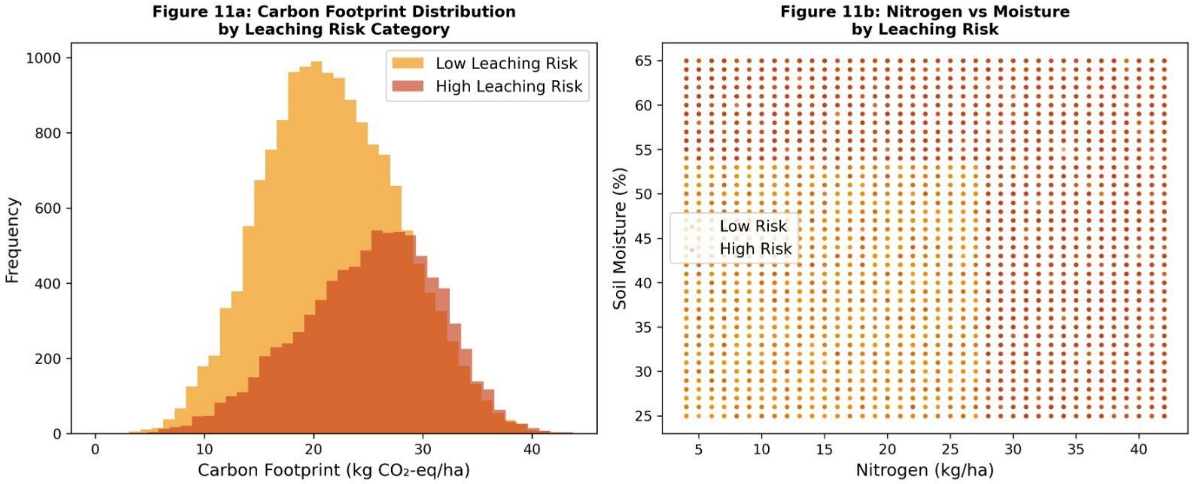
Figure 11a: Carbon Footprint Distribution by Leaching Risk Category

Figure 11b: Nitrogen vs Moisture by Leaching Risk

Fig. 9. Sustainability Analysis — Carbon Footprint Distribution and Nitrogen vs. Moisture by Leaching Risk.

The SHAP analysis revealing Nitrogen (63.45%) and Moisture (30.37%) as dominant predictors aligns with decades of agronomic research on leaching dynamics. This interpretability component directly addresses the transparency gap identified in the literature review and provides farmers with actionable, explainable guidance rather than opaque black-box outputs.

VIII. LIMITATIONS

Several limitations of the current implementation should be acknowledged. First, the Nitrogen Leaching Risk and Carbon Footprint targets were engineered from existing agronomic features rather than measured directly from field sensors or environmental monitoring systems. While the engineering rules are grounded in established agronomic thresholds, realworld leaching behaviour is influenced by additional factors including rainfall intensity, soil hydraulic conductivity, slope gradient, and seasonal variation that are not captured in the current dataset.

Second, the dataset originates from a Kaggle competition where samples were generated from a deep learning model trained on a real fertilizer prediction dataset. While the feature distributions closely approximate real agronomic conditions, the absence of direct field measurement introduces a degree of epistemic uncertainty.

Third, the proposed IoT-based real-time monitoring component has not yet been physically deployed in a field environment. The current implementation demonstrates the algorithmic and analytical components of the framework; hardware integration, sensor calibration, and edge computing deployment remain as future work.

Fourth, the current study does not include real-time soil microbial biomass measurements, which have been identified as an important indicator of soil health and long-term fertility. Affordable IoT-based biological sensing technologies for microbial monitoring are still under development.

IX. CONCLUSION

This study developed a multi-objective intelligent decisionsupport system for sustainable fertilizer management, nitrogen leaching prediction, and carbon footprint monitoring in precision agriculture. By integrating Random Forest, CatBoost, TabNet, Explainable AI (SHAP), and a Pareto-based multiobjective optimisation module, the framework addresses the critical gap between yield-focused fertilizer management and long-term environmental sustainability.

Experimental validation on 23,682 real agronomic samples from the Kaggle Playground Series S5E6 dataset demonstrated that Random Forest, CatBoost, and TabNet jointly achieved the highest nitrogen leaching classification accuracy of 95.55% with F1-Score of 0.9354, with CatBoost achieving the highest AUC-ROC of 0.9471. All three top models achieved 5fold cross-validation accuracy of 94.86%±0.23%, confirming strong generalisability. Logistic Regression served as the interpretable baseline at 81.42%, confirming the nonlinear nature of the leaching risk problem. For carbon footprint estimation, Linear Regression achieved the best R^2 of 0.7811 with RMSE of 3.0022 $kgCO_2$-eq/ha, validating the environmental cost modelling component. SHAP analysis confirmed that Nitrogen content (63.45%) and soil Moisture (30.37%) are the primary predictors of leaching risk, accounting for 93.82% of predictive power.

The proposed framework moves beyond one-dimensional optimisation by simultaneously balancing crop yield, economic profit, and environmental stewardship through a continuous, proactive decision pipeline. By delivering personalised recommendations through mobile-ready interfaces and location-aware GIS mapping, the system is designed to be accessible and practical for smallholder farmers in emerging agricultural contexts. Future work will focus on empirical field validation across diverse agro-climatic regions, physical IoT deployment, and integration of affordable biological sensing technologies for real-time soil microbial monitoring.

REFERENCES

[1] S. W. G. K. Bulankulama, A. P. S. Karunathilaka, and H. G. K. Bandara, “Impact of sole usage of organic fertilizer and chemical fertilizer on rice farmers’ quality of life,” *Journal of Research in Humanities and Social Science*, 2024.

[2] S. M. Venkateswara and J. Padmanaban, “Interpretable deep learning models for independent fertilizer and crop recommendation,” *Scientific Reports (Springer Nature)*, 2024.

[3] A. M. Gimba and P. K. Mishra, “Leveraging ensemble learning techniques for efficient fertilizer recommendation,” *JITSI: Jurnal Ilmiah Teknologi Sistem Informasi*, 2025.

[4] A. Irfan, M. Hashim, Hammadullah, and A. M. Shaikh, “Advancing agriculture with IoT and a smart fertilizer recommendation system,” *VFAST Transactions on Software Engineering*, 2025.

[5] P. Uppar, S. T. Bhairappanavar, H. M. Jayadeva *et al.*, “GIS based spatial fertilizer recommendation mapping using soil test crop response equations and validation in paddy,” *Plant Science Today*, 2025.

[6] I. S. M. Farhad, H. M. Naser, E. Jahan *et al.*, “Fertilizer recommendation for chilli-onion intercropping system,” *Bangladesh Agronomy Journal*, 2025.

[7] G. Raddy, K. S. Rajashekarappa *et al.*, “Soil fertility assessment and balanced fertilizer recommendation for a watershed using GIS and GPS,” *International Journal of Chemical Studies*, 2019.

[8] S. A. Alex and A. Kanavalli, “Assessment framework modeling using location aware computing for fertilizer management,” *International Journal of Innovative Technology and Exploring Engineering*, 2019.

[9] Intansari and Subiksa, “Effectiveness of organic fertilizer granules for increasing sweet corn on acid dryland,” *IOP Conference Series: Earth and Environmental Science*, 2022.

[10] Hossain and Siddique, “Online fertilizer recommendation system (OFRS): Precision agriculture for smallholders,” *Digital Government: Research and Practice*, 2020.

[11] T. Avhad, Y. Bhore *et al.*, “To enhance crop fertilizer prediction using machine learning and IoT,” *International Journal of Scientific Research in Engineering and Management*, 2025.

[12] K. Wanjale *et al.*, “Sustainable fertilizer management using random forest and agricultural data analysis,” *Journal of Intelligent Systems*, 2025.

[13] Manju M. *et al.*, “Smart fields: Enhancing agriculture with machine learning,”

International Journal of Advanced Research in Science, Communication and Technology, 2024.

[14] Haque and Biswas, "Long-term impact of fertilizers on soil and rice productivity," *Rice Science*, 2020.

[15] Chivenge *et al.*, "Organic fertilizer application significantly improved soil microbial activity and organic carbon," *Global Change Biology*, 2020.

[16] M. Rashid *et al.*, "A comprehensive survey of electric vehicle charging demand forecasting: Methodologies, datasets, and future directions," *IEEE Open Journal of Vehicular Technology*, vol. 5, pp. 1348–1373, 2024.

[17] Y. Mudgal *et al.*, "Cluster-based EV network forecasting and peak demand management," 2025.

[18] S. Khaleghian *et al.*, "Station-level utilisation prediction using tree-based models," 2025.

[19] M. M. R. Nune *et al.*, "IoT-based short-term load forecasting using SVR," 2025.

[20] L. Buzna *et al.*, "EV load forecasting for smart grid planning," 2023. [21] X. Huang *et al.*, "Short-term charging station load forecasting using ensemble methods," 2022.

[22] FAO, "The future of food and agriculture: Trends and challenges," Food and Agriculture Organization of the United Nations, Rome, 2017.

[23] IPCC, "Climate Change and Land: Special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems," 2019.