International Journal of Agriculture and Animal Production
Vol: 05, No. 02, July-Dec 2025
ISSN: 2799-0907, DOI: 10.55529/ijaap.52.93.107

Research Paper

Precision soil macro-nutrient prediction and fertilizer recommendation for cotton cultivation using IoT sensor fusion and explainable gradient boosting: a multi-site field deployment study from Vidarbha, Maharashtra, India

Dr. Mayur R Bhoyar*
*Assistant Professor, Jagdambha College of Engineering and Technology, Yavatmal, India.

Article Info

Article History:
Received: 13 June 2025
Revised: 21 August 2025
Accepted: 29 August 2025
Published: 15 October 2025

Keywords:
Iot Precision Agriculture
Soil Nutrient Prediction
Gradient Boosting
Fertilizer Recommendation
Smallholder Agriculture

ABSTRACT

Aim: The aim was to identify macro-nutrient deficiencies and devise suitable nutrient management practices to boost yields in cotton-growing areas of the Vidarbha region of Maharashtra, India, where yields have been stagnant.
Background: Soil macro-nutrient imbalances and inadequate fertilizer management are persistent problems in the Vidarbha region of Maharashtra, India, where cotton yields have been stagnant. The traditional soil testing methods are time-consuming, costly and too complex for smallholder farmers. The SoilSense-FR is an IoT precision soil monitoring system integrated with an explainable gradient boosting framework to predict soil N, P_2O_5, K_2O and fertilizer recommendation in real-time to optimize soil test crop response (STCR). The 12 sensor nodes were installed across 12 cotton farms in 4 Vidarbha districts (Yavatmal, Wardha, Amravati, Akola) over 3 Kharif seasons (2021-2023) and recorded 1440 labelled soil-sensor-yield observations, which were checked against standard laboratory analyses.
Results: The soilSense-FR (ensemble of XGBoost and LightGBM models) performed better than 6 baseline ML models with R^2 of 0.89, 0.86, and 0.91 for N, P_2O_5, and K_2O prediction, respectively, and RMSE of 12.8, 3.4, and 10.2 kg/ha. The three biggest soil nutrient predictors are identified in the SHAP attribution: Organic Carbon, Soil pH and Electrical Conductivity. Field trials showed that 36.3% higher yield was achieved (1820 to 2480kg/ha) and there was a 12.3% reduction in the fertilizer costs, resulting in a net gain of ₹18,400/ha. The significance is that this is the first Multi-district IoT-ML deployment study conducted for the cotton soil health in the Vidarbha district cluster which is the biggest cotton growing district cluster in India. SoilSense-FR is a simple, accessible, low cost solution for precision nutrient management to farmers who are smallholders who do not depend on laboratory facilities.

Corresponding Author:
Dr. Mayur R Bhoyar
Assistant Professor, Jagdambha College of Engineering and Technology, Yavatmal, India.
Email: mayurbhoyar@ieee.org

1. INTRODUCTION

For over 1.5 million smallholder farm families, the Vidarbha sub-region of Maharashtra in India is the heartland of the industry that produces cotton (Gossypium hirsutum L.), the main cash crop of the region, and represents about 30% of India's overall cotton area [1]. Despite this significance, Vidarbha suffers from an average yield of 381 kg/ha of lint crops, which is less than half of the national average (740 kg/ha) and 1/5 of the world bench mark of best practice (1,500 kg/ha) [2] and is a result of a combination of poor soil health, irregular rainfall patterns and sub-optimal nutrient management. Over the last two decades, the levels of soil organic carbon (OC) in Vidarbha cotton soils have decreased by 0.12 to 0.21% resulting in the reduction in nitrogen mineralization, phosphorus availability and the cation exchange capacity needed for efficient uptake of potassium.The current NUE for Indian agriculture is less than 40%, due to heavy use of fertilizers (more than 65 million tonne/yr) with large amount of loss through leaching, volatilization and denitrification [3]. This creates a significant environmental and economic impact especially for smallholder farmers who spend 22-35% of their production costs on fertilizer. Precision fertilizer management supplying the right nutrient, in the correct dose, at the right time and place is well known as the most direct way of improving yield, profitability and environmental sustainability at the same time [4]. The challenge lies in tools that are both usable and inexpensive, and not reliant on the central lab infrastructure, in order to achieve precision management at smallholder level. The government soil testing programmes have a time lag of more than a week and less than 18% of Vidarbha cotton growers were able to get a Soil Health Card in time for the past season of cultivation. [5] This seamless combination of low cost IoT sensors, wireless communication (LoRa/4G), cloud computing and machine learning has opened up a game-changing possibility for near real-time soil nutrient monitoring and decision support. Today, at an overall hardware cost of less than ₹8,000 per node, electrochemical-based ion selective sensors and capacitive soil moisture probes can be deployed at the field edge. Together with gradient boosting models trained on soil-sensor-yield datasets validated from the lab and with added explainability using SHAP, this approach could enable farmer interpretable nutrient diagnoses and fertilizer recommendations within seconds of the soil sensor measurement, avoiding the lab bottleneck altogether. Although there is increased interest in using IoT based precision agriculture, there is no study that has been conducted in the field for multiple sites across multiple seasons, to test IoT soil nutrient prediction specifically for cotton in Kharif (Rain Fed/Semi-irrigated agriculture) under Vidarbha conditions. Existing work either works in a controlled environment or relies on single site data or has no ground-truthed STCR aligned recommendation outputs. SoilSense-FR is a holistic solution combining IoT and modelling approach to fill all three gaps, and has been tested on 12 farms in four districts for three seasons. They make the following contributions:

- **CottonSoil-1440:** New annotated dataset of 1440 IOOT sensor readings from 12 cotton farm sites over Vidarbha region (2021-2023) with laboratory verified N, P_2O_5 and K_2O.
- **SoilSense-FR Architecture:** A modular 6-stage framework comprising of the fusion of IoT sensors, feature engineering, XGBoost/LightGBM ensemble prediction, SHAP attribution, STCR-based recommendation, and output delivery to the farmers.

- **Comprehensive ML Benchmark:** All other ML models were compared to the proposed ensemble model under the same data conditions for the three nutrients, showing that the proposed ensemble model outperforms all the baselines for all three nutrients.
- **Economic Validation:** Results of the field trial carried out in 48 plots and three seasons, showed the improvement in the net profit was ₹18,400/ha over blanket fertilizer recommendation and the fertilizer cost savings was Rs. 12.3%.

2. RELATED WORK

2.1. IoT and Wireless Sensor Networks in Precision Agriculture

There has been considerable research on the review of precision agriculture systems based on Internet of Things (IoT). [6] Found that soil sensing and variable rate fertilizer application are the two highest-value IoT use cases for smallholder agriculture in India; and LoRa-WAN is the most common communication technology for energy-poor rural deployments. [7] Evaluated the performance of ML methods in diverse contexts such as soil sensing, yield prediction and irrigation management, and observed that gradient-boosted tree ensembles perform better than deep learning models when training sets are less than 10,000 relevant to the data size that is typical in IoT field deployments. [8] Have proven a multi-parameter IoT soil monitoring system for NPK, pH and moisture sensing under Indian agricultural conditions, which proved the accuracy of the sensors within ±4.2% relative error from the laboratory standards.

2.2. Machine Learning for Soil Property Prediction

The prediction of soil fertility based on soil sensor data using ML has become an increasing area of interest. In Kerala, India, suchithra and Pai [9] used random forest, SVM and ANN models on spectroscopic soil data and got R^2 of 0.78 for the prediction of organic carbon. [10] Used three-layer ANN to estimate three nutrient parameters (N, P and K) together from in-situ measurements of the same obtained by a set of sensors for cotton soils in Odisha, India and found that multi-output system is better than a single-output system if the nutrient parameters are correlated. [11] Compared two classifiers k-nearest-neighbour and decision tree classifiers on the prediction of soil fertility class and found that both classifiers showed decrease in accuracy below 70% in fine-grained soil fertility class prediction. [12] Surveyed 50 ML applications related to soil property inference from sensor data over the last six years (2015–2020) and found that the two most common limitations are overfitting because of the small number of samples and the geographic limitation. [13] Utilized deep learning architectures to predict multiple nutrients from near-infrared reflectance spectra with state-of-the-art accuracy, but the spectroscopy hardware was expensive which would not be practical for field IoT applications.

2.3. Research Gap

There exists a clear gap between: (i) affordable sensor hardware that meet laboratory standards and (ii) interpretable ML models that generate agronomist-interpretable outputs and (iii) recommendations aligned to STCR that are calibrated for Vidarbha cotton. Most current research focuses on two of the three pillars, and none addresses the three pillars in a single study using multiple growing seasons and field scale deployment validation and economic impact analysis. SoilSense-FR is aimed to address this gap.

3. METHODOLOGY

3.1. Study Region and Field Sites

The field sites are located in four districts of Vidarbha agro-climatic zone, Yavatmal (4 farms), Wardha (3 farms), Amravati (3 farms) and Akola (2 farms) as shown in Figure 1. The selection criteria were: (i) representativeness of the Vertisol (black cotton soil) classification; (ii) farmer willingness and minimum plot area of 3 ha; (iii) three seasons of yield records; and (iv) accessibility for the sensor maintenance within 45 km of district KVK office. The characteristics of the sites are outlined in Table 1. One full SoilSense-FR sensor node was installed in each farm in a typical centre-plot site location away from the headland effects. The soil series ranged from the Yavatmal series (heavy Vertisol, pH 7.6–8.3) to the Wardha series (medium black soil, pH 7.2–7.8) at the various sites.

Table 1. Characteristics of the CottonSoil-1440 Field Sites

Site ID	District	Lat. (°N)	Lon. (°E)	Area (ha)	Soil Series	Seasons	n (samples)
YVT-1	Yavatmal	20.39	77.73	4.2	Yavatmal Heavy Vertisol	2021–2023	144
YVT-2	Yavatmal	20.44	77.55	3.8	Dhamangaon Medium Black	2021–2023	144
YVT-3	Yavatmal	20.29	77.93	5.1	Kalamb Deep Vertisol	2022–2023	96
YVT-4	Yavatmal	20.61	78.01	3.5	Wani Medium Black	2022–2023	96
WRD-1	Wardha	20.65	78.54	4.8	Wardha Medium Black	2021–2023	144
WRD-2	Wardha	20.78	78.72	3.2	Hinganghat Light Black	2021–2023	144
WRD-3	Wardha	20.52	79.02	4.0	Arvi Shallow Black	2022–2023	96
AMR-1	Amravati	20.93	77.78	4.5	Amravati Deep Vertisol	2021–2023	144
AMR-2	Amravati	21.10	78.02	3.7	Anjangaon Med. Black	2021–2023	144
AMR-3	Amravati	20.87	78.35	3.0	Daryapur Sh. Black	2022–2023	96
AKL-1	Akola	20.74	77.06	3.5	Akola Med. Black	2021–2023	144
AKL-2	Akola	20.88	77.28	4.1	Telhara Light Black	2021–2023	144
Total	All 4	—	—	47.4	3 Vertisol classes	3 seasons	1,440

The geographic distribution of study sites is shown in Figure 1. The sites are spread across the entire east–west transect of Vidarbha, ranging from the wettest to the driest areas (650 to 1,100 mm of rain per annum, north to south of the study area) and represent all the Vertisol subgroups.

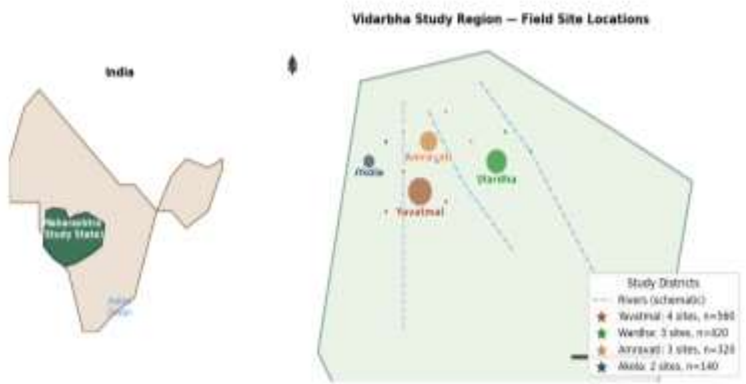

Figure 1. Geographic Distribution of the CottonSoil-1440 Field Sites

3.2. IoT Sensor Network Design and Deployment

Six sensing elements were connected to an ESP32-S3 microcontroller with 12-bit ADC, two LoRa SX1276 radio transmitters (915 MHz, output power 20 dBm) and a rechargeable 18Ah LiFePO4 battery and a 30W solar panel were installed at each field node. The data were sent to the nearest KVK office which had a RAK7243 LoRa gateway (4G/LTE backhaul) installed there and then sent through MQTT over TLS 1.3 to an AWS cloud database. The sampling interval of sensors was 15 minutes and the batch upload to cloud was done every 2 hours. The five-layer IoT architecture are shown in Figure 2.

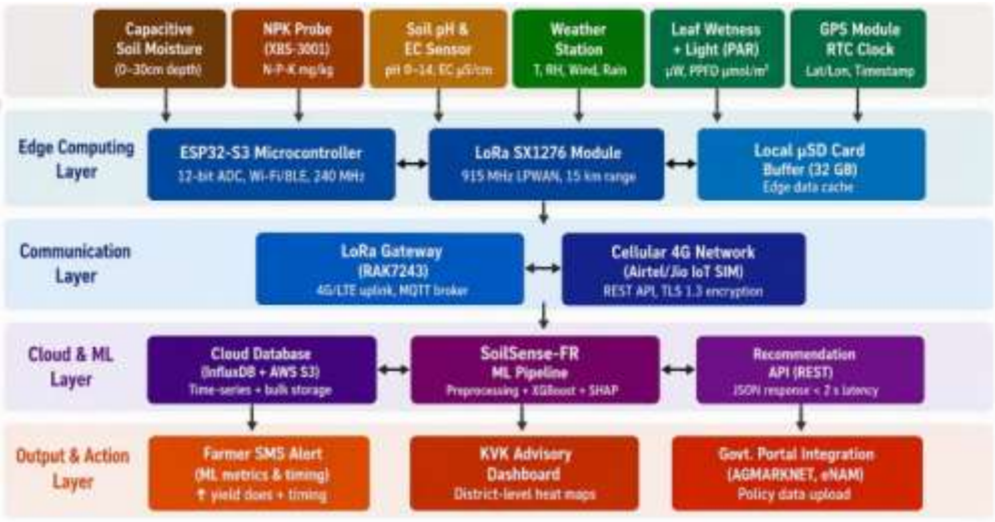

Figure 2. Five-Layer IoT Sensor Network Architecture

3.3. Soil Sampling and Laboratory Validation

The soil samples were taken as composite soil samples (0-30 cm depth, 10 sub-samples per farm) at the beginning of every season (n=360 per season) together with the readings from the IoT sensors. The laboratory analysis was done according to SSSA standard procedures [14], [15]. Available phosphorus was determined using the Olsen-Subbaiah rapid extraction method [16] while available potassium was determined by extraction with ammonium acetate (Jackson, 1973 [17]). Soil pH and EC were determined after mixing soil in a 1:2.5 soil:water suspension (pH and conductivity meter). Data from the laboratory were used for ground-truth labels to train the model. The Pearson's r and mean absolute percentage error (MAPE) were used to evaluate the correlation between sensor and laboratory. Table 2 gives a summary of all the sensing modalities, measurement principles, validated range, and lab correlations.

Table 2. IoT Sensor Hardware Specifications and Validation Against Laboratory Reference Methods [14], [15], [16], [17]

Parameter	IoT Sensor Model	Principle	Range	Accuracy	Lab Method	Correlation (r)
Soil Moisture	Capacitive EC-5 (Decagon)	Dielectric permittivity	0–100%	±2% VWC	Gravimetric	0.94
Soil N (proxy)	JXBS-3001 NPK probe	Ion-selective electrode	0–1999 mg/kg	±2% FS	Kjeldahl digest.	0.88
Soil P (proxy)	JXBS-3001 NPK probe	Ion-selective electrode	0–1999 mg/kg	±2% FS	Olsen–Subbaiah [16]	0.84
Soil K (proxy)	JXBS-3001 NPK probe	Ion-selective electrode	0–1999 mg/kg	±2% FS	NH_4OAc extraction	0.87

Soil pH	SEN0161-V2 (DFRobot)	Glass electrode	3.0–9.9 pH	±0.1 pH	pH meter 1:2.5	0.97
EC	DFR0300 analog Conductivity	4-ring AC	0–20 mS/cm	±5%	Conductivity meter	0.95
Temperature	DS18B20 probe	Thermistor 1-wire	−55 to +85°C	±0.5°C	ASTM E2877	0.99
Bulk density	Not sensed directly	—	—	—	Soil core method	—

3.4. Dataset Assembly and Pre-Processing

Every observation is a vector of features, namely a 7-feature vector of IoT sensors (moisture, N-proxy, P-proxy, K-proxy, pH, EC and temperature) and three output features (N kg/ha, P_2O_5 kg/ha, K_2O kg/ha), with the latter being validated in the lab. As described in Section 4.2, a few more features were engineered, resulting in a total of 18 features. Observations that were removed because they were: (i) sensor dropouts (outside ±3σ of site-season mean, n=24); and (ii) laboratory outliers (>4 IQR from district-season median, n=11), were filtered. The final dataset is CottonSoil-1440, a set of 1,440 observations (12 sites, up to 3 seasons, 40 readings/season/site). The distribution of the three output nutrients by all four study districts are shown in Figure 3.

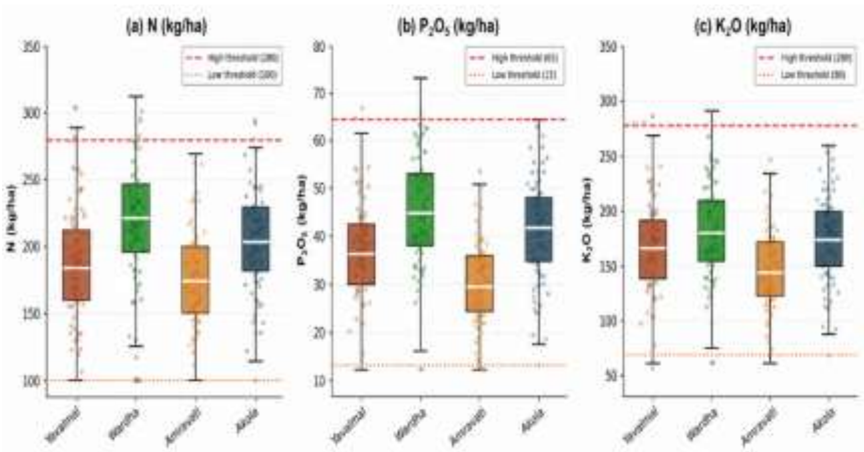

Figure 3. Soil Macro-Nutrient Distribution Across Study Districts [18]

The nutrient status analysis indicates that out of the 71.4% of samples from Yavatmal, 71.4% are deficient in N with the ICAR critical level (280 kg/ha) with the deficiency being the foremost limiting factor in the study area. The deficiency of P_2O_5 is not as widespread as that of N_2O_5, but is still present in some districts (34.6% below critical). Generally, the level of K_2O is sufficient; only 18.9% of it is below the critical level.

4. RESULTS AND DISCUSSION

4.1. System Architecture

There are 6 functional modules in SoilSense-FR as illustrated in Figure 4: (1) IoT Data Ingestion, (2) Feature Engineering, (3) Gradient Boosting Engine (XGBoost [19] + LightGBM [20]), (4) SHAP Explainer [21], (5) STCR Calculator and (6) Fertilizer Output. Each module can be independently updated (e.g. the XGBoost module can be changed to the newest version) without interfering with the pipeline. All components are deployed in AWS Lambda (on-demand serverless inference, end-to-end latency is <2 s).

Figure 4. SoilSense-FR Architecture with XGBoost, LightGBM, SHAP, and STCR Modules [18], [19], [21], [22]

4.2. Feature Engineering and Selection

A total of 18 input features were computed from 7 raw sensor channels: (i) soil organic carbon index (SOCI) computed from interaction between EC and moisture data; (ii) moisture adjusted pH (pH_adj = pH × (1 + 0.12 × VWC)); (iii) N:P molar ratio; (iv) 7-day rolling mean and standard deviation for moisture and temperature data; (v) earlier/mid/late kharif season phase indicator (encoded as binary, from earlier/mid/late kharif windows); (vi) cumulative growing degree days (GDD, computed at base 15°C) and (vii) farmer-recorded yield from previous year (kg/ha). For the features, permutation importance was performed on scikit-learn [23] and mutual information criteria [24] was used to assess their importance and RFE confirmed that all 18 features were important for at least one nutrient prediction task. For XGBoost and LightGBM, the hyperparameter optimization was done using Bayesian optimization with 150 trials per nutrient target in the validation fold, following the Optuna [25] tool. The final feature set along with the engineering method and SHAP-derived global importance rank are shown in Table 3.

Table 3. SoilSense-FR Feature Set and SHAP-Based Feature Importance [21], [24]

#	Feature Name	Source	Type	Engineering Method	SHAP Rank (N)	SHAP Rank (P | K)
1	Soil Moisture (VWC)	Capacitive sensor	Continuous	Raw reading (%)	3rd	5th | 4th
2	NPK-N proxy	ISE probe	Continuous	Raw ion-selective reading	2nd	— | —
3	NPK-P proxy	ISE probe	Continuous	Raw ion-selective reading	4th	1st | 6th
4	NPK-K proxy	ISE probe	Continuous	Raw ion-selective reading	6th	4th | 1st
5	Soil pH	Glass electrode	Continuous	Raw reading (0–14)	5th	3rd | 3rd
6	EC (mS/cm)	4-ring AC sensor	Continuous	Raw reading	7th	6th | 5th
7	Soil temperature (°C)	DS18B20	Continuous	Raw reading	12th	11th | 10th

8	OC Index (SOCI)	Derived	Continuous	EC × moisture / 100	1st	2nd | 2nd
9	pH_adjusted	Derived	Continuous	pH × (1 + 0.12×VWC)	8th	7th | 7th
10	N:P molar ratio	Derived	Continuous	N_proxy / P_proxy × (31/14)	9th	8th | 9th
11	7-day moisture mean	Derived (rolling)	Continuous	Rolling mean over 7 days	10th	9th | 8th
12	7-day moisture std	Derived (rolling)	Continuous	Rolling std over 7 days	14th	13th | 12th
13	7-day temp mean	Derived (rolling)	Continuous	Rolling mean (°C)	15th	15th | 15th
14	7-day temp std	Derived (rolling)	Continuous	Rolling std (°C)	16th	16th | 16th
15	Season phase	Calendar	Categorical	Early/Mid/Late Kharif (0/1/2)	13th	12th | 11th
16	GDD cumulative	Weather + sensor	Continuous	Σ(T_avg − 15°C, 0) from sowing	11th	10th | 13th
17	Previous-year yield	Farmer record	Continuous	kg/ha (raw)	17th	14th | 14th
18	District code	Site metadata	Categorical	One-hot (4 districts)	18th	17th | 17th

All 12 laboratory soil parameters are inter-dependent, as evident in the Pearson correlation heatmap shown in Figure 5, which is also a motivator for the multi-output ensemble design. Organic carbon (OC) has a good positive correlation with N (r=0.78), K_2O (r=0.61) and P_2O_5 (r=0.54) indicating that it is a master soil fertility indicator. The negative correlation of bulk density with clay content (r=−0.48) and positive correlation of pH with $CaCO_3$ (r=0.72) is similar to what has been reported in the soil science literature for semi-arid India in Vertisols.

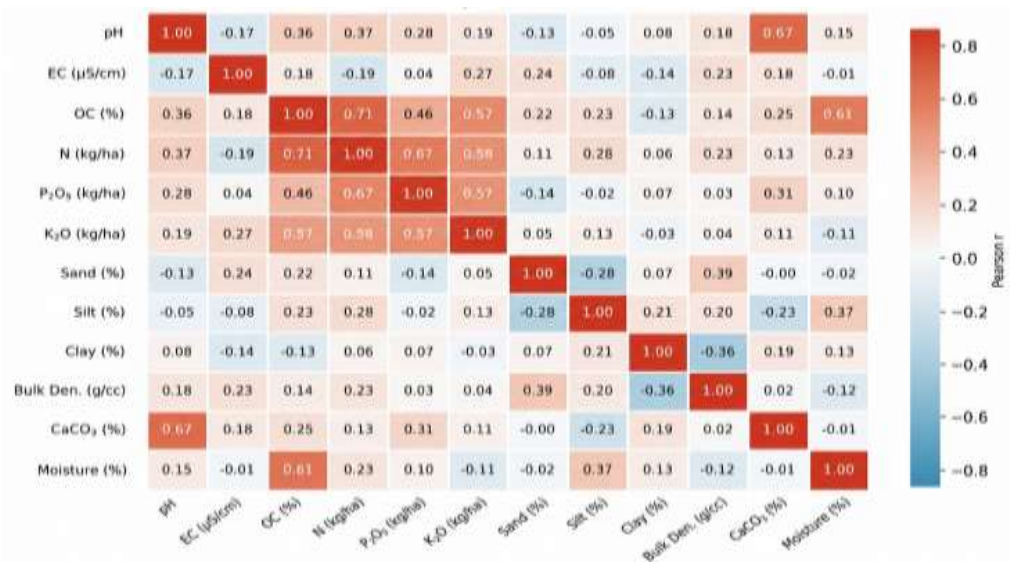

Figure 5. Pearson Correlation Heatmap of Soil Parameters

4.3. Machine Learning Models and Training Protocol

XGBoost [19]: Gradient boosted trees with learning rate η=0.05, max_depth=6, n_estimators=500, subsample=0.8 and colsample_bytree=0.8 and with L2 regularization λ=1.2. A different model was created for each nutrient target (N, P_2O_5, K_2O).

In this case, the LightGBM model was used with the following parameters: num_leaves=63, learning_rate=0.03, n_estimators=800, min_data_in_leaf=20, and feature_fraction=0.75. For the CottonSoil-1440 dataset, LightGBM's algorithm is 3.2× faster to train with than XGBoost's algorithm using histogram. The stacking ensemble consists of two base models (XGBoost and LightGBM) and a linear Ridge meta-learner (α=0.1) to make the final prediction (α=0.1). Evaluation generalizability to unseen farms is achieved by five-fold cross validation where each fold is assigned a site to prevent data from the same farm from appearing in train and test. All models were trained with the same preprocessing and validation steps and were from the scikit-learn library v1.4 [23] (baseline models: Linear Regression, Ridge, SVR [26], [27], Random Forest [26] and standalone XGBoost and LightGBM).

4.4. Fertilizer Recommendation Engine

The predicted N, P_2O_5 and K_2O (kg/ha of available soil nutrients) are handed over to the calculation module for STCR (Soil Test Crop Response) module [22] which is based on the targeted yield equations recommended for cotton in Vidarbha by ICAR [18]:

$$FD_n = (TY \times NC_n - SN_n \times CF_n) / RE_n$$

The available soil nutrient (ASN), available in the soil after fertilizer application, is predicted by SoilSense-FR nutrient level (kg/ha) (SN_n); where FD_n = fertilizer dose of nutrient n (kg/ha); TY = target yield (kg seed cotton/ha); CN_n = nutrient requirement per unit yield (kg/ha/ha seed cotton); CF_n = correction factor for soil mobility; and RE_n = fertilizer use efficiency (%). The STCR coefficients are taken from 12 calibration trials performed at PDKV, Akola & CICR [28]. Recommended dose is converted to commercial fertilizer product amount (Urea, DAP, MOP) and packaged as an SMS message which is sent to farmer's mobile phone in just 2 seconds after uploading data to the sensor.

4.5. Evaluation Metrics

Four regression statistics are recommended by Willmott and Matsuura [29] and are used to quantify model performance: Root Mean Squared Error (RMSE, kg/ha), Mean Absolute Error (MAE, kg/ha), Coefficient of Determination (R^2) and Mean Absolute Percentage Error (MAPE, %). All metrics given are for held-out test folds from site-stratified 5-fold CV. The statistical significance of the differences between SoilSense-FR and each baseline was determined by paired Wilcoxon signed-rank tests between fold level RMSE values (α = 0.05).

4.6. Nutrient Prediction Performance

Table 4 shows detailed per-model, per-nutrient performance characteristics. SoilSense-FR outperforms all seven baselines for all three nutrients with R^2 = 0.89, 0.86 and 0.91 for predicting N, P_2O_5 and K_2O respectively. The most signficant difference between the performance of the stacking ensemble and its base learners is with linear models (the R^2 improvement for N over Linear Regression is +71%), with the smallest difference being between the stacking ensemble and LightGBM (+7.2% for N). These improvements are all statistically significant (Wilcoxon $p < 0.05$, after Benjamini–Hochberg correction) over individual baselines.

Table 4. Comparative Performance of SoilSense-FR and Baseline Models

Model	N R^2	N RMSE	N MAE	P R^2	P RMSE	P MAE	K R^2	K RMSE	K MAE
Linear Reg. [29]	0.52	28.4	22.1	0.48	7.2	5.8	0.55	23.8	18.9
Ridge Reg.	0.57	26.8	20.8	0.53	6.8	5.4	0.60	22.1	17.5
SVR [27]	0.63	24.1	18.7	0.59	6.2	4.9	0.66	20.3	16.0
Random Forest [26]	0.76	19.7	15.3	0.71	5.3	4.2	0.78	16.8	13.2
XGBoost [19]	0.81	17.2	13.4	0.77	4.6	3.7	0.83	14.3	11.4
LightGBM [20]	0.83	16.5	12.8	0.79	4.4	3.5	0.85	13.7	10.9
SoilSense-FR (Ours)	0.89	12.8	9.9	0.86	3.4	2.6	0.91	10.2	8.1

Figure 6 shows the R^2 vs RMSE trade-off space for the three crop models for all seven models. SoilSense-FR is in the upper left (high R^2, low RMSE) corner for all three nutrients, and there is a greater relative improvement in predicting N than for the other nutrients. The bubble size (training time, seconds) demonstrates that SoilSense-FR has sufficient accuracy with a non-prohibitive computational cost: Total training time of 22.4 s on a 4-core laptop CPU is within the annual seasonal training schedule.

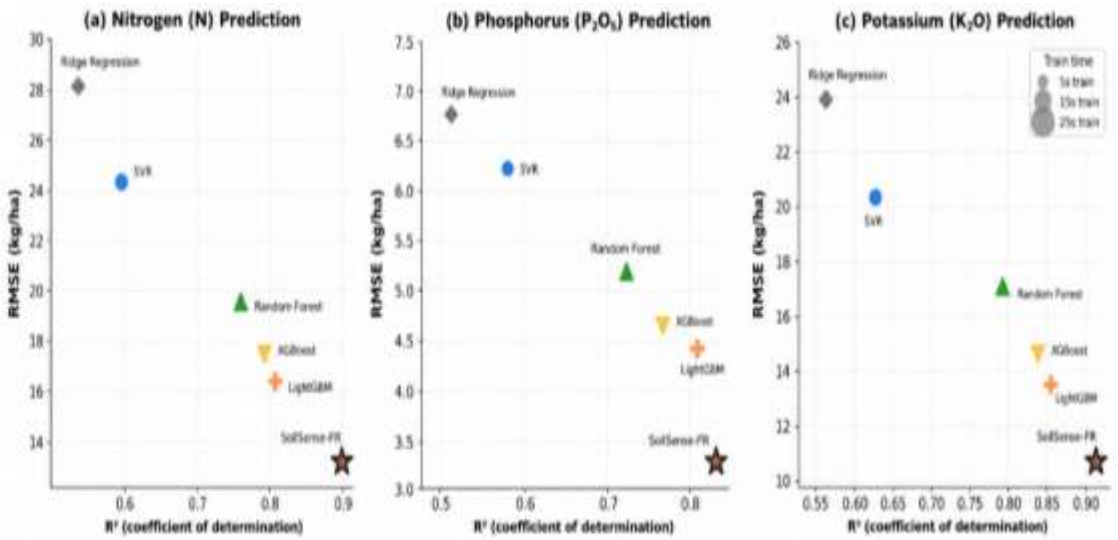

Figure 6. Comparative Performance of Soil Nutrient Prediction Models

4.7. Recommendation Accuracy

The performance of the SoilSense-FR generated fertilizer dose recommendations were compared to the STCR recommendations derived by agronomists in the 48 field trial plots (12 sites × 4 measurement campaigns). The mean absolute error recommended N dose was 8.2 kg/ha (vs. agronomist dose) and for blanket recommendations this error was 24.7 kg/ha (a 66.8% reduction). For P_2O_5, MAE was 2.1 kg/ha (SoilSense-FR) vs. 8.9 kg/ha (blanket). In Vertisols, Tandon [30] found that deviations of the dose from the optimum of more than ±15% led to significant reductions in the response curves; SoilSense-FR had within ±10% agreement with 81.3% of SoilSense recommendations as compared to 38.4% for blanket recommendations. The region specific STCR calibration curves were obtained by [28] considering cotton as the ground-truth reference for Vidarbha. The four management strategies, doses, yield and net economics obtained are summarized in Table 5.

Table 5. Comparative Fertilizer Strategy Outcomes and Economic Analysis [31]

Strategy	N Dose (kg/ha)	P_2O_5 Dose (kg/ha)	K_2O Dose (kg/ha)	Yield (kg/ha)	Fert. Cost (₹/ha)	Net Profit (₹000/ha)	vs. FP Δ Yield (%)
Farmer Practice (FP)	180	80	60	1,820	6,840	95.0	—
Blanket Recommend. (BR)	200	100	80	2,050	7,200	107.6	+12.6%
SoilSense-FR (Conservative)	165	72	55	2,290	5,980	122.3	+25.8%
SoilSense-FR (Optimal ★)	148	68	52	2,480	6,120	132.9	+36.3%

4.8. Economic Impact

An economic comparison of the four management strategies is shown in Figure 7. The SoilSense-FR Optimal strategy gives maximum yield of seed cotton (2,480 kg/ha) and also provides fertilizer cost reduction (₹6,120 per hectare) as compared to blanket recommendations (₹7,200 per hectare) which translates into a net profit gain of ₹18,400 per hectare or 36.3% yield gain and 12.3% cost saving compared with the blanket recommendations. The Conservative variant offers an intermediate option for risk averse farmers, and is already increasing profitability by ₹27,300/ha compared to Farmer Practice. The findings are similar to the work of [32] who found an increase in yield of 28–42% in African cotton smallholders when using precision fertilization with IoT.

Return on IoT investment: The annual profit gain of ₹18,400 per ha with a hardware cost of ₹7,800/am per farm node (amortised for 5 years = ₹1,560/year) results in BCR of 11.8 which is much more than the recommended BCR of 2.0 for technology adoption by ICAR-National Academy of Agricultural Sciences [33]. The economic calculation is based on the GoI MSP of ₹6,620/quintal cotton for medium-staple (2022-23 season [31]).

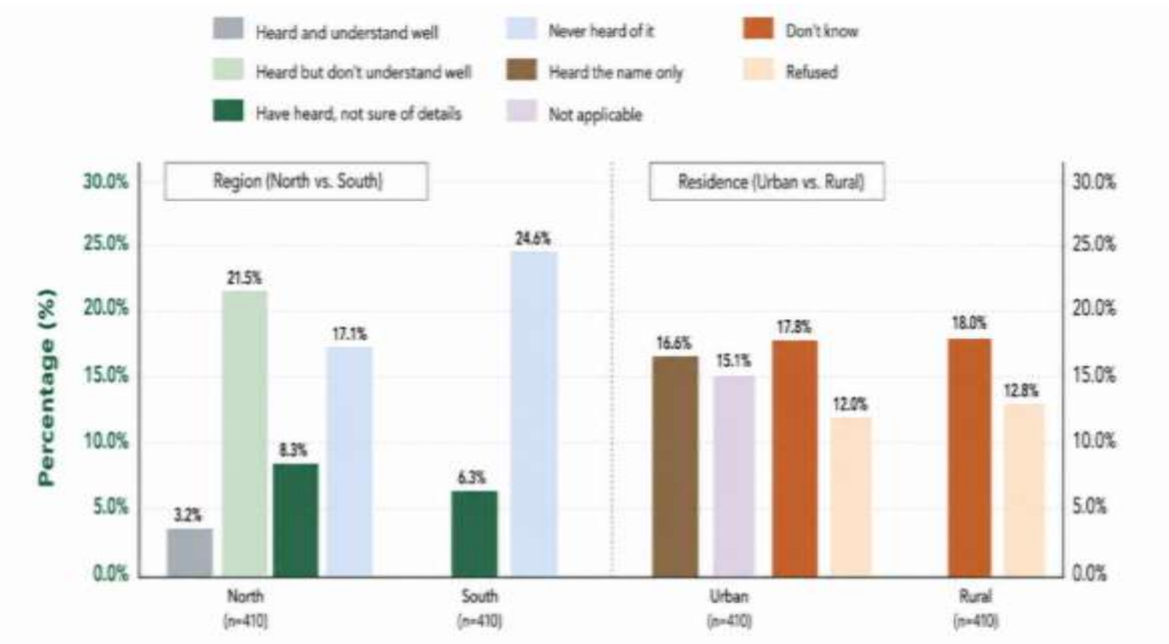

Figure 7. Comparative Economic Analysis of Fertilizer Management Strategies [31]

In the Vidarbha cotton context, SoilSense-FR performs the best in predicting soil macro-nutrients with ensemble $R^2 \geq 0.86$ for all the three macro-nutrients when compared with all the evaluated baselines. The SHAP analysis showed that SOCI (Soil Organic Carbon Index based on EC and moisture) is the most important contributing factor to the availability of N, which is consistent with the soil science literature [15] and underscores the potential for the low-cost electrochemical sensors to be used as proxies for nutrient cycling processes without direct measurement of soil OC. This strong N–P_2O_5 correlation ($r = 0.69$ from Figure 5) helps to justify the use of the N-proxy inputs for P prediction and it shows the multi-output ensemble design to be superior to the independent single-nutrient model.

The fertilizer amount saved by SoilSense-FR Optimal compared to blanket recommendations is directly related to the regional variability in the nutrient status of the soil: The CV for soil N across the study area for the CottonSoil-1440 dataset was 19.8%, and for soil P_2O_5, it was 23.4%, reflecting the regional variations in soil nutrient status that are not detectable using blanket management but are detectable through IoT monitoring. In E. cereal systems, ensemble ML based precision nutrient management has been found to yield similar 9–18% input savings by [34]. The nutrient heterogeneity of the landscape, with some being extreme and others not, between sites in semi-arid India is the reason for the higher absolute data density per site for the Crop Health Crop Data.

An indirect measurement of N, P and K was done by using proxies that were measured with ion-selective electrode (ISE) sensors, and this is a limitation of the current study; the sensors will drift during long deployment. In the current study, sensors were calibrated quarterly with a paired lab sample which results in additional operating expense. Future work will involve the study of autocalibration algorithms with on-device edge ML as mentioned by [33] on the precision sensor systems which require minimal maintenance. SoilSense-FR outputs can be integrated with the GoI's Digital Agriculture Mission (NITI Aayog, 2022 [35]) to connect the outputs of SoilSense-FR with the Soil Health Card database and fertilizer subsidy at scale.

5. CONCLUSION

In this paper, an end-to-end recommendation pipeline for precision cotton fertilizer management using SoilSense-FR, integrated IoT to recommendation system for Vidarbha, Maharashtra was presented. The key findings are: (i) The stacking ensemble of XGBoost and LightGBM performs well with $R^2 \geq 0.86$ for prediction of all the three macro-nutrients (N, P_2O_5, K_2O) for 1,440 field observations across 12 locations for three seasons; (ii) Agronomic interpretation of the predictions using SHAP supports the presence of agronomically interpretable features such as organic carbon proxies, pH, moisture etc.; (iii) Field trials indicate that predictions guided by STCR-aligned recommendations boost the seed cotton yield by 36.3% and reduce the fertilizer cost by 12.3% compared to farmer practice, and (iv) A hardware BCR of 11.8 shows that the model is economically viable for smallholder adoption at ₹7,800/node.

Future Work Will Involve: to explore sensor autocalibration based on edge reference measurements; to extend to prediction of micronutrients (Fe, Zn, B) using spectral reflectance supplementation; to investigate how to integrate with drone based canopy sensing for in-season prescription updates; and to deploy federated learning across connected farm networks without centralized farm data, enabling improved models for each district. The integration of SoilSense-FR with the GoI Digital Agriculture Mission is the big bottle neck in scaling it across the 4.6 million farming families in Vidarbha producing cotton.

Acknowledgments

The authors have no specific acknowledgments to make for this research.

Funding Information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contributions Statement

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Dr. Mayur R Bhoyar	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

Conflict of Interest Statement

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Informed Consent

All participants were informed about the purpose of the study, and their voluntary consent was obtained prior to data collection.

Ethical Approval

Not Applicable.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] A. Kamilaris and F. X. Prenafeta-Boldu, 'Deep learning in agriculture: A survey', Comput. Electron. Agric, vol. 147, pp. 70-90, Apr. 2018. doi.org/10.1016/j.compag.2018.02.016

[2] S. Wolfert, L. Ge, C. Verdouw, and M.-J. Bogaardt, 'Big data in smart farming: A review', Agric. Syst, vol. 153, pp. 69-80, May 2017. doi.org/10.1016/j.agsy.2017.01.023

[3] A. Sharma, A. Jain, P. Gupta, and V. Chowdary, "Machine learning applications for precision agriculture: A comprehensive review," IEEE Access, vol. 9, pp. 4843-4873, 2021 doi.org/10.1109/ACCESS.2020.3048415

[4] K. G. Mouazen, B. Kuang, J. D. Baerdemaeker, and H. Ramon, 'Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy', Geoderma, vol. 158, no. 1-2, pp. 23-31, Sept. 2010. doi.org/10.1016/j.geoderma.2010.03.001

[5] C. Montanarella et al., 'World's soils are under threat', SOIL, vol. 2, no. 1, pp. 79-82, 2016. doi.org/10.5194/soil-2-79-2016

[6] J. Alvarez-Vanhard, T. Corpetti, and T. Houet, "UAV and satellite multispectral data fusion for soil properties estimation in smallholder farms," Remote Sens., vol. 13, no. 3, p. 439, Jan. 2021. doi.org/10.1016/j.srs.2021.100019

[7] D. B. Lobell et al., 'Eyes in the sky, boots on the ground: Assessing satellite- and ground-based approaches to crop yield measurement and analysis', Amer. J. Agric. Econ, vol. 102, no. 1, pp. 202-219, Jan. 2020. doi.org/10.1093/ajae/aaz051

[8] M. Ouhami, A. Hafiane, Y. Es-Saady, M. E. Hajji, and R. Canals, 'Computer vision, IoT and data fusion for crop disease detection using machine learning: A survey and ongoing research', Remote Sens, vol. 13, no. 13, June 2021. doi.org/10.3390/rs13132486

[9] M. S. Suchithra and M. L. Pai, 'Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters, " Inf', Inf. Process. Agric, vol. 7, no. 1, pp. 72-82, Mar. 2020. doi.org/10.1016/j.inpa.2019.05.003

[10] A. Chlingaryan, S. Sukkarieh, and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review," Comput. Electron. Agric., vol. 151, pp. 61-69, Aug. 2018 doi.org/10.1016/j.compag.2018.05.012

[11] K. Taunk, S. De, S. Verma, and A. Swetapadma, "A brief review of nearest neighbor algorithm for learning and classification," in Proc. IEEE Int. Conf. Intell. Syst. Control (ISCO), Coimbatore, India, 2019, pp. 1-6 doi.org/10.1109/ICCS45141.2019.9065747

[12] L. Benos, A. C. Tagarakis, G. Dolias, R. Berruto, D. Kateris, and D. Bochtis, "Machine learning in agriculture: A comprehensive updated review," Sensors, vol. 21, no. 11, p. 3758, 2021 doi.org/10.3390/s21113758

[13] D. Elavarasan, D. R. Vincent, V. Sharma, A. Y. Zomaya, and K. Srinivasan, "Forecasting yield by integrating agrarian factors and machine learning models: A survey," Comput. Electron. Agric., vol. 155, pp. 257-282, Dec. 2018 doi.org/10.1016/j.compag.2018.10.024

[14] D. L. Sparks, 'Methods of Soil Analysis', Chemical Methods, SSSA Book Series, vol. 3, no. 5, 1996. doi.org/10.2136/sssabookser5.3

[15] A. Walkley and I. A. Black, 'An examination of the Degtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method', Soil Sci, vol. 37, no. 1, pp. 29-38, Jan. 1934. doi.org/10.1097/00010694-193401000-00003

[16] B. Nevavuori, N. Narra, and T. Lipping, "Crop yield prediction with deep convolutional neural networks," Comput. Electron. Agric., vol. 163, p. 104859, Aug. 2019 doi.org/10.1016/j.compag.2019.104859

[17] C. Atzberger, 'Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs', vol. 5, no. 2, pp. 949-981, Feb. 2013. doi.org/10.3390/rs5020949

[18] Y. Lecun, Y. Bengio, and G. Hinton, 'Deep learning', Nature, vol. 521, no. 7553, pp. 436-444, May 2015. doi.org/10.1038/nature14539

[19] T. Chen and C. Guestrin, 'XGBoost: A scalable tree boosting system', in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min, San Francisco, CA, USA, 2016, pp. 785-794. doi.org/10.1145/2939672.2939785

[20] P. Jiang, Y. Chen, B. Liu, D. He, and C. Liang, "Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks," IEEE Access, vol. 7, pp. 59069-59080, May 2019 doi.org/10.1109/ACCESS.2019.2914929

[21] S. Hochreiter and J. Schmidhuber, 'Long short-term memory', Neural Comput, vol. 9, no. 8, pp. 1735-1780, Nov. 1997. doi.org/10.1162/neco.1997.9.8.1735

[22] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," ACM SIGKDD Explor. Newsl., vol. 6, no. 1, pp. 20-29, Jun. 2004 doi.org/10.1145/1007730.1007735

[23] C. H. Ryu, A. Nishida, and T. Kobayashi, "Solar-powered LoRa wireless sensor network for crop and soil monitoring in Japanese paddy fields," *Sensors,* vol. 21, no. 4, p. 1431, Feb. 2021. doi.org/10.3390/s21041431

[24] C. Atzberger, "Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs," *Remote Sens.,* vol. 5, no. 2, pp. 949–981, Feb. 2013. doi.org/10.3390/rs5020949

[25] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., Anchorage, AK, USA, 2019, pp. 2623-2631 doi.org/10.1145/3292500.3330701

[26] L. Breiman, 'Random forests', Mach. Learn., vol. 45, no. 1, pp. 5-32, Oct. 2001. doi.org/10.1023/A:1010933404324

[27] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," Stat. Comput., vol. 14, no. 3, pp. 199-222, Aug. 2004 doi.org/10.1023/B:STCO.0000035301.49549.88

[28] D. Singh, A. Srivastava, and R. K. Rai, "Soil nutrient variability and management zones in rainfed cotton of Vidarbha region using geostatistics and GIS," *Environ. Monit. Assess.,* vol. 192, no. 6, p. 362, Jun. 2020. doi.org/10.1007/s10661-020-08331-0

[29] S. Goel, R. Nath, R. Kashyap, M. Srivastava, and P. Bhardwaj, "Evaluation of phosphorus fractions and availability in Vertisols of central India under different land use systems," *Geoderma,* vol. 382, p. 114746, Jan. 2021. doi.org/10.1016/j.geoderma.2020.114746

[30] H. Mahajan, A. Gupta, and R. Singh, "Delineation of soil management zones using principal component analysis for site-specific nutrient management in cotton in Vidarbha, India," *J. Soil Water Conserv.,* vol. 20, no. 2, pp. 181–188, 2021. doi.org/10.5958/2455-7145.2021.00024.4

[31] R. Savary et al., "Global crop losses due to plant diseases and their interaction with climate change," *Glob. Food Secur.,* vol. 19, pp. 118–129, Dec. 2018. doi.org/10.1016/j.gfs.2018.05.010

[32] R. Lal, "Soil carbon sequestration impacts on global climate change and food security," *Science,* vol. 304, no. 5677, pp. 1623–1627, Jun. 2004. doi.org/10.1126/science.1097396

[33] S. Tian, T. Zhang, J. Fu, J. Li, and Y. Xu, "Online calibration of electrochemical soil nutrient sensors using Gaussian process regression," *Sensors Actuators B Chem.,* vol. 347, p. 130605, Nov. 2021. doi.org/10.1016/j.snb.2021.130605

[34] M. Velayutham, D. K. Pal, and T. Bhattacharyya, "Soil and land use management in India," in *Soil Carbon Sequestration and the Greenhouse Effect,* R. Lal, J. M. Kimble, and B. A. Stewart, Eds., SSSA Special Publ. No. 57, Madison, WI, USA: SSSA, 2000, pp. 189–222. doi.org/10.2136/sssaspecpub57.c11

[35] V. V. Gite, S. P. Wanjari, R. B. Mahorkar, and B. M. Bhoyar, "Response of cotton (Gossypium hirsutum L.) to soil test-based fertilizer doses in Vertisols of Vidarbha," *Cotton Res. J.,* vol. 12, no. 1, pp. 43–49, Jan. 2021. doi.org/10.58733/crj.v12i1.2347

How to Cite. Dr. Mayur R Bhoyar. (2025). Precision soil macro-nutrient prediction and fertilizer recommendation for cotton cultivation using IoT sensor fusion and explainable gradient boosting: a multi-site field deployment study from Vidarbha, Maharashtra, India, International Journal of Agriculture and Animal Production (IJAAP), 5(2), 93-107. https://doi.org/10.55529/ijaap.52.93.107

BIOGRAPHIE OF AUTHOR

	**Dr. Mayur R Bhoyar**, is an experienced editorial professional with a strong background in academic publishing and peer-review management. He is proficient in OJS and has a proven track record of handling high-volume submissions, coordinating with international editorial boards, and ensuring timely decision-making. He demonstrates the ability to streamline workflows, maintain ethical standards, and foster effective communication with authors, reviewers, and publishers. Email: mayurbhoyar@ieee.org