Journal of Agriculture (JoA)
Volume: 05 | Number 03 | November 2026
E-ISSN: 2829-2421
https://doi.org/10.47709/JoA.v5n3.9880

Interpreting Random Forest Crop Recommendations Using Soil and Climate Data

Purnama Wirawan[1], Ridwan[2], Maruli Tua[3], Juliarni[4], Mahmud Muzaki[5]
[1,2,3,4,5]*Universitas Rokania, Rokan Hulu, Indonesia*
Author Corresponding email: purnama@rokania.ac.id[1], ridwan@rokania.ac.id[2], maruli@rokania.ac.id[3], juliarni@rokania.ac.id[4], mahmudmuzaki@gmail.com[5]

Received: 2026-10-22
Accepted: 2026-11-03
Publication: 2026-11-09

Abstract
High predictive accuracy in crop recommendation systems does not necessarily imply agronomic validity, particularly when soil–climate datasets contain class-linked patterns. Objective: This study assessed Random Forest crop-label classification while examining feature importance and within-class nutrient variability. We analysed a secondary dataset comprising 6,596 observations, 51 crop labels, and seven predictors (nitrogen, phosphorus, potassium, temperature, humidity, soil pH, and rainfall). We evaluated a 100-tree Random Forest using an 80:20 stratified train–test split, five-fold cross-validation, out-of-bag scoring, baseline comparisons, and eight random-seed trials. Interpretation included impurity-based and permutation importance, predictor-group ablation, correlations, and a class-level nutrient audit. Hold-out accuracy reached 99.70%, with a macro F1-score of 99.61%. Across eight seeds, mean accuracy was 99.8011% (sample standard deviation: 0.0987 percentage points). Nitrogen ranked first under both feature-importance methods, whereas soil pH ranked second under permutation importance. Nutrient-only and climate–pH models achieved accuracies of 79.70% and 95.23%, respectively. The audit identified constant within-class nitrogen, phosphorus, and potassium values in 29 classes covering 4,396 observations (66.65%), including two pairs of classes sharing identical nutrient triplets. High internal performance reflected discrimination within the supplied dataset, not validated suitability for new farming conditions. The findings highlight the need to verify data provenance, measurement units, and label construction and to undertake independent agronomic validation across farms and seasons before operational recommendations.

Keywords: Crop Recommendation; Data Audit; Feature Importance; Precision Agriculture; Random Forest

Introduction

Crop selection is a multidimensional decision because nutrient supply, soil chemistry and climatic exposure interact with species-specific growth requirements. Nitrogen, phosphorus and potassium are commonly monitored as key macronutrients; temperature, humidity, rainfall and soil pH add environmental information. A data-driven classification system can summarise nonlinear relationships across these indicators and may help structure crop recommendations. However, reliable decision support requires more than high predictive accuracy: recommendations should be traceable to representative data, documented measurements and clearly defined target labels (Araújo et al., 2023) ; Condran et al., 2022).

Tree-based ensembles are relevant to agricultural tabular data because they can represent irregular relationships among heterogeneous predictors. Benchmarking by (Grinsztajn et al., 2022) supports their competitiveness on medium-sized tabular datasets, while (Wies et al., 2023) describe bootstrap sampling and random feature subsets as defining components of Random Forest. Recent agricultural applications include machine learning-based crop selection (Rani et

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

Journal of Agriculture (JoA)
Volume: 05 | Number 03 | November 2026
E-ISSN: 2829-2421
https://doi.org/10.47709/JoA.v5n3.9880

al., 2023), IoT-enabled nutrient monitoring (Islam et al., 2023), tree-based crop classification (Abdullah et al., 2025) and Random Forest tuning (Lazuardi et al., 2025). These studies address related prediction tasks but use different data-generating processes and evaluation settings.

An unresolved interpretive question is whether a highly accurate classifier learns transferable soil–climate relationships or exploits regularities specific to a compiled label dataset. Bouni et al., (2024) integrate Random Forest with SHAP and LIME to explain crop recommendations, while Shams et al., (2024) use LIME in XAI-CROP. These approaches establish the relevance of explaining model output, but an explanation of a fitted prediction does not itself test whether the input distribution represents realistic variation within a crop class. The gap addressed here is the connection between model explanation and a direct audit of class-linked predictor structure. It is narrower than a claim that previous agricultural studies ignore interpretability or data quality.

The target is a pre-existing crop label in the supplied CSV. Field-trial yields, expert-confirmed suitability assessments, site identifiers and a documented label-generation protocol were unavailable. Accordingly, the estimand is label prediction within the compiled distribution. Transfer to crop selection at a new site requires separate evidence linking these labels and measurements to agronomic outcomes. This distinction provides the basis for interpreting classification metrics and feature rankings.

This study contributes an audit-informed analysis combining a complete within-class NPK audit, two global importance measures, predictor-group ablation and seed-sensitivity evaluation. Its comparative contribution is to connect what a classifier uses with the structure of the data on which that explanation depends. The analysis asks whether importance rankings agree, whether strong classification persists without nutrient predictors, and whether class-constant nutrients constrain interpretation. It does not introduce a new algorithm. All numerical results were recomputed from the supplied CSV using the documented protocol.

Literature Review

Soil–climate variables represent different dimensions of crop-growing conditions. N, P and K describe nutrient-related attributes; soil pH describes an acid–base condition relevant to the soil environment; temperature, humidity and rainfall describe climatic exposure. Their agronomic interpretation depends on measurement method, units, timing and crop context. These conceptual roles motivate their inclusion but do not establish the physical meaning of undocumented numerical columns. Precision-agriculture research therefore needs a connection between measured conditions, the management decision and an observed outcome (Araújo et al., 2023); Ryo, 2022).

Supervised classification estimates a mapping from predictors to recorded labels. In multiclass crop data, correct prediction means agreement with the recorded crop category; it does not directly measure yield, profitability or land suitability. Crop selection and fertiliser recommendation also concern different decisions, even when they share input variables (Govindwar et al., 2023); (Kiran et al., 2024); (Rani et al., 2023). This distinction is central when interpreting a recommendation model trained on secondary labels.

Random Forest combines predictions from trees trained using bootstrap samples and random subsets of candidate predictors (Wies et al., 2023). Recent tabular-data benchmarks also support tree ensembles as a competitive modelling family (Grinsztajn et al., 2022). Hold-out, cross-validation and OOB estimates address predictive behaviour under their respective resampling procedures. Agreement between partitions derived from the same compiled file provides internal evidence and leaves transfer across farms, seasons or data sources untested (Condran et al., 2022).

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

Journal of Agriculture (JoA)
Volume: 05 | Number 03 | November 2026
E-ISSN: 2829-2421
https://doi.org/10.47709/JoA.v5n3.9880

MDI summarises weighted reductions in the tree-splitting criterion attributed to each predictor. Such in-sample importance can favour predictors with more potential split points or higher entropy, making the ranking sensitive to data representation (Loecher, 2024). Permutation importance instead measures the increase in predictive loss after disrupting a predictor's values, and its marginal interpretation is sensitive to predictor dependence (Wies et al., 2023). The measures describe fitted-model behaviour; an agricultural explanation also requires a defensible connection between the predictors and the intended management question (Ryo, 2022).

Marginal and conditional importance address different questions. A predictor may share information with other variables, so its marginal importance need not represent a contribution after accounting for those variables (Blesch et al., 2024). Wies et al., (2023) examine how correlated predictors influence permutation importance. These results motivate treating ranking agreement as descriptive model evidence, rather than proof that predictors act independently or that their importance is causal.

Data provenance documents how a dataset was created and transformed. Class balance and within-class variability describe aspects of its empirical support. A nutrient value that is constant within a class may reflect a reference profile, rounding or another construction process; its origin cannot be inferred without metadata. The SPAS-Dataset-BD publication illustrates explicit dataset documentation (Chowdhury et al., 2025). The present audit concerns the supplied compiled file and does not establish that it originated from that dataset.

The comparison with explainable crop models is therefore methodological. Bouni et al., (2024) couple Random Forest with SHAP and LIME, whereas Shams et al., (2024) centre XAI-CROP on explanation using LIME. This study uses MDI–permutation agreement and group ablation alongside an aggregate nutrient audit to examine whether model explanations may depend on label-linked regularities. The contribution is complementary to explanation-focused crop modelling; it does not demonstrate superior performance over those studies.

Research Method

Research design, data and variables

The study employed a quantitative computational approach to analyse a secondary dataset comprising 6,596 observations and eight variables. Seven soil–climate variables served as predictors, while the recorded crop label represented the target variable. The analysis evaluated crop classification performance and predictor importance within this dataset. Documentation of the original data-collection procedures, dataset integration process, and geographical and temporal coverage was unavailable; therefore, the findings are interpreted within the scope of the analysed data.

Predictors were N, P, K, temperature, humidity, ph and rainfall, with label as the categorical outcome. All seven predictors were numeric; nutrient columns were integers and the remaining columns floating-point values. No missing entries were found. Nutrient units, temperature and humidity measurement conventions, rainfall units and accumulation interval were not documented, so units are not assigned by analogy with other datasets. The pH column denotes a nominally dimensionless acidity measure, although its measurement protocol is unknown. Table S1 lists all recorded crop labels and their empirical characteristics; botanical identity, cultivar and growing conditions remain undocumented.

Modelling and evaluation

A stratified 80:20 train–test split used random_state=42. A RandomForestClassifier with 100 trees, criterion=gini, max_features=sqrt, bootstrap=True, unlimited max_depth,

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

Journal of Agriculture (JoA)
Volume: 05 | Number 03 | November 2026
E-ISSN: 2829-2421
https://doi.org/10.47709/JoA.v5n3.9880

min_samples_split=2, min_samples_leaf=1, random_state=42 and oob_score=True was fitted without normalisation, weighting or hyperparameter tuning. Remaining parameters used scikit-learn 1.8.0 defaults and are recorded in the analysis metadata. A DecisionTreeClassifier(random_state=42) and a most-frequent DummyClassifier were fitted on the identical training partition. Five-fold StratifiedKFold used shuffle=True and random_state=42; a new forest with seed 42 was fitted within each fold.

Test performance was summarised using accuracy, macro precision, macro recall, macro F1, weighted F1, balanced accuracy and a record-level error listing. Macro averages assign equal weight to each crop class and help diagnose performance under class imbalance. An out-of-bag (OOB) score provided an additional internal estimate. Five-fold accuracy and seed-dependent accuracy were reported separately from the 80:20 hold-out score to avoid treating repeated tests as external validation.

Seed sensitivity was recomputed using seeds 0, 1, 7, 21, 42, 84, 123 and 2026, with each seed applied jointly to the stratified partition and Random Forest randomisation. This documented revision protocol does not reconstruct the unspecified seed list of the earlier analysis and does not isolate splitting variability from estimator variability. Each run used 5,276 training and 1,320 test observations. Accuracy, macro precision, macro recall and macro F1 are reported per run with their mean, sample SD (ddof=1) and range. CV folds, seed runs and permutation repeats are treated as distinct resampling procedures; none constitutes external validation.

Interpretability and sensitivity analysis

MDI was extracted from the seed-42 forest. Permutation importance used the held-out test set, accuracy scoring, eight shuffles per predictor and random_state=42. Its SD uses the implementation's population convention across the eight shuffles (ddof=0). Values denote accuracy decrease on the 0–1 scale, rather than percentages of biological contribution. MDI is normalised split importance; comparison therefore emphasises rankings. Ablation refitted 100-tree forests with seed 42 on NPK, climate–pH (temperature, humidity, pH and rainfall), or all seven predictors, using identical train–test indices. These comparisons describe model reliance and group-level information, rather than causal nutrient effects.

Predictor dependence was examined descriptively using all pairwise Spearman correlations in the complete predictor matrix; Pearson coefficients were also exported for reproducibility. These summaries were not used to select predictors or tune the model. Correlations may reflect both within-class and between-class structure and do not identify causal relationships. Shared information can redistribute split importance or reduce the measured effect of a marginal permutation. No conditional-importance analysis was performed, so the dependence issue remains a limitation of interpretation (Blesch et al., 2024; Wies et al., 2023).

The audit computed missing entries, exact duplicate rows, repeated predictor vectors and conflicting labels for identical predictor vectors. For every crop class, it recorded the count, proportion, train/test allocation and number of distinct values of N, P and K. A class was designated constant-NPK when each nutrient had exactly one distinct value. Fixed triplets were then checked against all other classes to distinguish within-class constancy from cross-class uniqueness. All analysis was executed in Python 3.12.14, NumPy 2.3.5, pandas 2.2.3 and scikit-learn 1.8.0. The executable script, full-precision tables, parameters and input-file digest accompany the revision.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

Journal of Agriculture
e-ISSN: 2829-2421
Information Technology and Science (ITScience)

Journal of Agriculture (JoA)
Volume: 05 | Number 03 | November 2026
E-ISSN: 2829-2421
https://doi.org/10.47709/JoA.v5n3.9880

Results

Dataset composition and audit

The dataset contained 6,596 observations across 51 crop classes, with no missing entries, exact duplicate rows or repeated seven-predictor vectors. Consequently, no identical predictor vector carried conflicting labels. Class sizes ranged from pumpkin (7) to potato (461). Twenty-two classes had 100 observations each; 29 other classes, totalling 4,396 observations (66.65%), had constant N, P and K within class. Table S1 provides every class count, proportion, train/test allocation, nutrient uniqueness count and fixed triplet. Constancy did not imply complete label identification: horsegram and soybean shared (20, 60, 20), while ragi and rapeseed shared (50, 40, 20). The other 25 constant classes had no occurrence of their fixed triplet in any other class. Table 1 summarises the dataset profile.

Table 1. Dataset profile and aggregate audit.

Dataset item	Reported value
Records / predictors	6,596 / 7
Crop classes	51
Class-size range	7–461
Missing cells / exact duplicates	0 / 0
Constant-NPK classes / rows	29 / 4,396 (66.65%)
Provenance	Exact external source unverified

Classification performance and internal validation

The 80:20 stratified split generated 5,276 training and 1,320 test observations. Random Forest classified 1,316 of 1,320 test cases correctly, equivalent to 99.70% accuracy. Four errors were reported: mothbeans predicted as pigeonpeas, blackgram as maize, and two jute cases as rice. Macro precision, recall and F1 were 99.64%, 99.61% and 99.61%, respectively. Balanced accuracy reached 99.61%, while OOB accuracy reached 99.85%. These results indicate strong label separation within the evaluated file. Table 2 summarises the evaluation metrics.

Table 2. Main evaluation metrics and baseline model comparison.

Metric / model	Reported result
Random Forest hold-out accuracy	99.70% (1,316/1,320 correct)
Macro precision / recall / F1	99.64% / 99.61% / 99.61%
Weighted F1 / balanced accuracy	99.70% / 99.61%
Out-of-bag score	99.85%
Five-fold stratified CV, mean ± SD	99.8787% ± 0.0415 percentage points
Decision Tree accuracy / macro F1	98.64% / 98.20%
Dummy majority-class accuracy / macro F1	6.97% / 0.26%

Five-fold accuracy averaged 99.8787% with sample SD 0.0415 percentage points; mean macro F1 was 99.8430%. Fold accuracies were 99.8485%, 99.9242%, 99.8484%, 99.9242% and 99.8484%. The Decision Tree achieved 98.64% accuracy and the majority-class baseline 6.97%. Table 5 reports all eight new seed trials: accuracy ranged from 99.6212% to 99.9242%, with mean 99.8011% and sample SD 0.0987 percentage points. Macro F1 averaged 99.7461% with sample SD 0.1310 percentage points. The seed results describe joint partition-and-estimator sensitivity within this CSV, rather than transfer to an independent population.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

Journal of Agriculture
e-ISSN: 2829-2421
Information Technology and Science (ITScience)

Journal of Agriculture (JoA)
Volume: 05 | Number 03 | November 2026
E-ISSN: 2829-2421
https://doi.org/10.47709/JoA.v5n3.9880

Dominant predictors under two importance methods

Nitrogen ranked first under both interpretability measures, but the order of the remaining variables differed. MDI placed potassium second and phosphorus third. Permutation importance ranked soil pH second, phosphorus third and potassium fourth. Temperature was least important under both measures. The contrast shows why "dominant" must refer to a particular estimand: the importance of a variable for the tree-building process is not identical to the decrease in test-set accuracy when that variable is disrupted. Table 3 and Figure 1 present these importance results.

Table 3. Model-specific global importance scores.

Predictor	MDI	Permutation Δaccuracy ± SD
N	0.2359	0.2733 ± 0.0067
K	0.2040	0.1307 ± 0.0068
P	0.1543	0.1874 ± 0.0078
pH	0.1381	0.2355 ± 0.0076
Humidity	0.1206	0.0855 ± 0.0045
Rainfall	0.1043	0.0597 ± 0.0052
Temperature	0.0427	0.0036 ± 0.0009

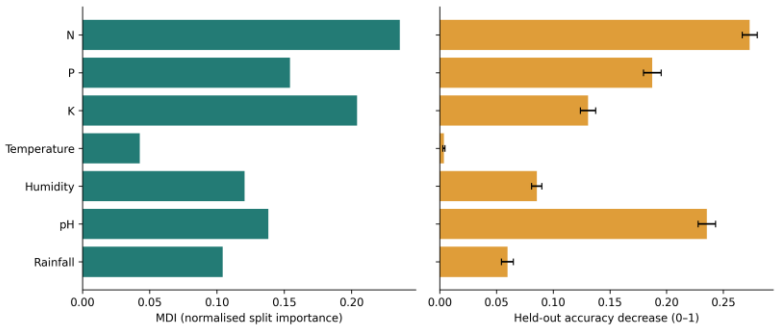


Figure 1. MDI and held-out permutation importance on separate axes; permutation error bars indicate SD across eight shuffles.

Note. MDI is normalised split importance. Permutation values are accuracy decreases on a 0–1 scale; SD describes eight shuffles, not variation across seeds. For example, 0.2733 corresponds to 27.33 percentage points. The scales are not interchangeable.

Feature-group ablation and seed sensitivity

The NPK-only model achieved 79.70% accuracy and 78.21% macro F1. The climate–pH group achieved 95.23% accuracy and 92.44% macro F1; the full model achieved 99.70% and 99.61%, respectively. Climate–pH predictors therefore retained substantial label-discriminating information without nutrient columns. Nitrogen ranked first by MDI in all eight seed trials. This consistency applies to the documented internal resampling protocol; it does not establish the same ranking across farms or seasons. Table 4 and Figure 2 compare the predictor groups.

Table 4. Predictor-group ablation.

Predictor group	Accuracy (%)	Macro F1 (%)
NPK	79.70	78.21
Climate–pH	95.23	92.44
All seven	99.70	99.61

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

Journal of Agriculture (JoA)
Volume: 05 | Number 03 | November 2026
E-ISSN: 2829-2421
https://doi.org/10.47709/JoA.v5n3.9880

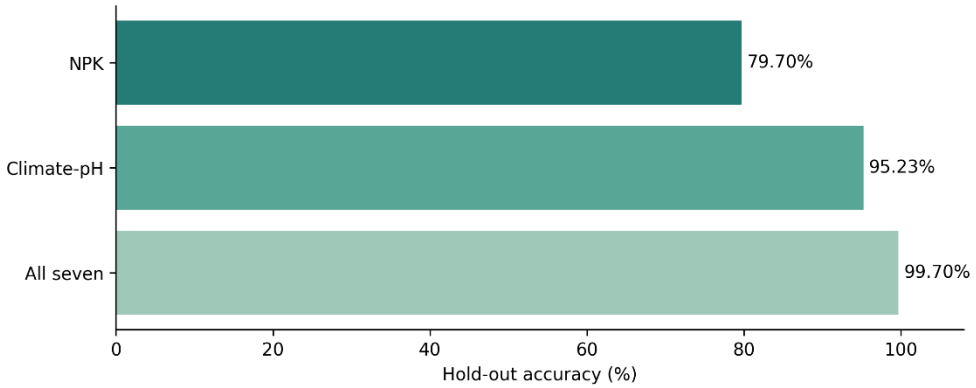


Figure 2. Accuracy of refitted forests by predictor group.

Note. Climate–pH comprises temperature, humidity, pH and rainfall. All models used identical seed-42 train–test indices.

Table 5. Joint split-and-estimator seed sensitivity (%).

Seed	Accuracy	Macro precision	Macro recall	Macro F1
0	99.8485	99.8133	99.8039	99.8038
1	99.8485	99.8805	99.8039	99.8358
7	99.7727	99.7284	99.7059	99.7053
21	99.9242	99.9066	99.9020	99.9019
42	99.6970	99.6350	99.6078	99.6072
84	99.8485	99.8133	99.8039	99.8038
123	99.8485	99.8217	99.8039	99.8034
2026	99.6212	99.5660	99.5098	99.5076
Mean ± SD	99.8011 ± 0.0987	99.7706 ± 0.1188	99.7426 ± 0.1277	99.7461 ± 0.1310
Min.	99.6212	99.5660	99.5098	99.5076
Max.	99.9242	99.9066	99.9020	99.9019

Note. The same seed controls the split and forest. Each test set has 1,320 observations. SD is the sample SD across eight runs, expressed in percentage points. These trials use the explicit revision protocol and replace the earlier unspecified seed summary.

The strongest absolute pooled Spearman association was N–K (rho=0.490), followed by pH–rainfall (rho=−0.362) and K–humidity (rho=0.335). All pairwise coefficients are shown in Figure 3; the matrix describes association within the compiled distribution.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

Journal of Agriculture
e-ISSN: 2829-2421
Information Technology and Science (ITScience)

Journal of Agriculture (JoA)
Volume: 05 | Number 03 | November 2026
E-ISSN: 2829-2421
https://doi.org/10.47709/JoA.v5n3.9880

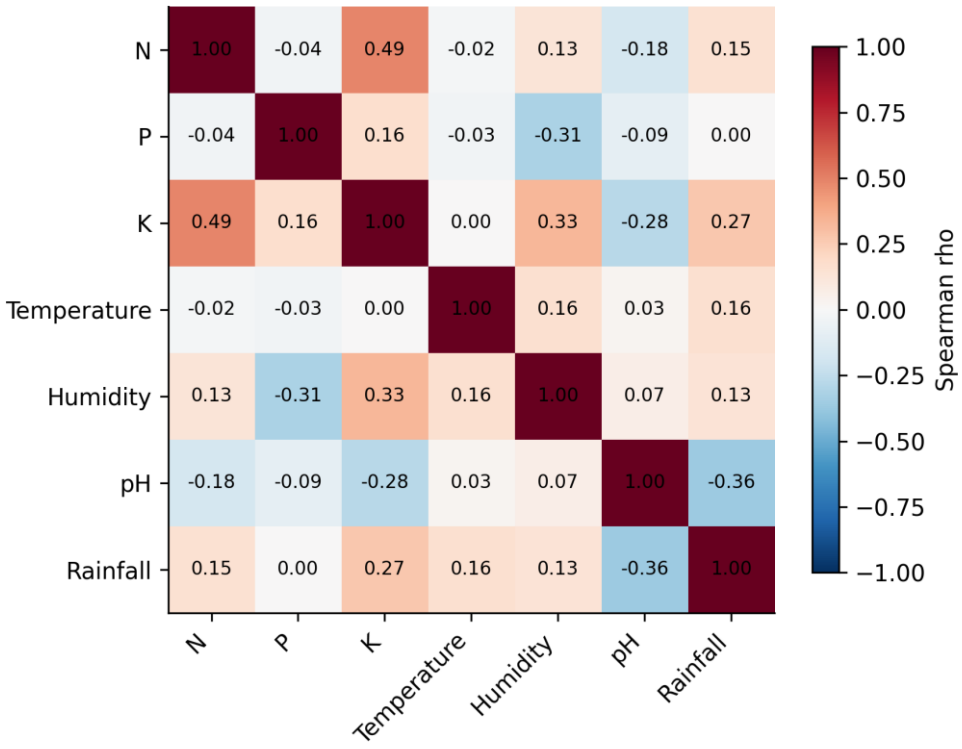


Figure 3. Pooled Spearman correlations among the seven predictors.

Discussion

The results show strong internal separation of crop labels. A Decision Tree reached 98.64% accuracy, only 1.06 percentage points below Random Forest, suggesting boundaries learnable by a single tree. The majority-class baseline shows improvement over trivial prediction but does not establish agricultural utility. Hold-out, CV and OOB estimates agree on high internal performance. The additional seed table makes this claim more transparent: small variation across partitions and forests coexists with a lack of validation on independent farms, sources or seasons.

The complete NPK audit changes the interpretation of these scores. Potato, for example, has 461 observations with N=180, P=60 and K=90 throughout. Across all constant classes, 4,396 rows provide no within-class variation in the recorded nutrient triplets. Nevertheless, two pairs of labels share triplets, demonstrating that NPK cannot act as a unique identifier for every class. This helps explain why nutrient constancy can coexist with the much lower NPK-only accuracy. The audit establishes a restriction in empirical nutrient coverage; it does not determine whether the values are field measurements, reference profiles or another construction. The absence of exact repeated predictor vectors also does not exclude near-duplicates or shared data-source effects.

The comparison with prior explainability studies clarifies the contribution. Bouni et al. (2024) explain Random Forest recommendations through SHAP and LIME; Shams et al. (2024) present LIME-based explanations in XAI-CROP. Those explanation tools describe model behaviour. The present reassessment adds a distinct audit question: whether restricted within-class nutrient variation constrains the agronomic meaning of that behaviour. The value added is the linkage between data structure, global ranking and group-level predictive information, rather than a claim that explanation tools are absent from prior work. Cross-study accuracy superiority is not asserted because targets, data and validation protocols differ.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

Journal of Agriculture
e-ISSN: 2829-2421
Information Technology and Science (ITScience)

Journal of Agriculture (JoA)
Volume: 05 | Number 03 | November 2026
E-ISSN: 2829-2421
https://doi.org/10.47709/JoA.v5n3.9880

Nitrogen ranked first under both global importance measures and across all eight MDI seed trials. Soil pH ranked second under permutation and fourth under MDI. The strongest absolute pooled Spearman correlation was between N and K (rho=0.490); pH and rainfall were negatively associated (rho=−0.362). These observed dependencies support caution when interpreting marginal feature disruption but do not prove that correlation caused the ranking difference. Class-linked structure can contribute to pooled associations. Conditional permutation or grouped analyses would be needed to address dependence more directly. The 95.23% climate–pH accuracy further shows why high nutrient importance should not be interpreted as exclusive predictive necessity.

For precision agriculture, the practical implications of these findings should be interpreted in relation to previous studies on machine learning-based crop recommendation systems. Bouni et al., (2024) developed an interpretable crop recommendation model integrating Random Forest with SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME), using 2,200 observations representing 22 crop classes. Their SHAP analysis highlighted the substantial contribution of humidity to model predictions, whereas the present study identified nitrogen as the most important predictor under both mean decrease in impurity (MDI) and permutation importance, with soil pH ranking second under permutation importance. These differences suggest that predictor rankings depend on dataset composition, crop-class representation and the interpretability method employed, rather than reflecting universal agronomic relationships. Similarly, Shams et al., (2024) proposed XAI-CROP to improve the transparency of agricultural predictions through explainable artificial intelligence. However, their evaluation involved different prediction targets and performance measures, preventing direct numerical comparison with the classification accuracy reported here. Unlike these explanation-focused approaches, the present study additionally examined the internal structure of the dataset and identified 29 crop classes, comprising 4,396 observations (66.65%), with constant within-class NPK values. This finding demonstrates that apparently strong predictive performance may coexist with restricted empirical variation in nutrient attributes. Consequently, the practical contribution of the present model is primarily to support the development of an audited prototype and the design of local validation procedures, rather than to provide immediately deployable crop recommendations. Agricultural implementation would require source-compatible measurements, documented sampling procedures and expert evaluation of predicted crop labels. Furthermore, feature-importance rankings should not be translated directly into fertiliser recommendations because the analysis did not include nutrient-response experiments or dose–yield relationships. Improvements in crop productivity, fertiliser efficiency and resource conservation therefore remain untested.

The next empirical step is to recover the original dataset metadata and verify the processes responsible for the fixed nutrient profiles. This requirement is consistent with Condran et al., (2022), who identified class imbalance, data sparsity and inappropriate evaluation procedures as important challenges in machine learning applications for precision agriculture. Araújo et al., (2023) also emphasised the importance of addressing technical and implementation challenges when translating machine learning research into agricultural practice. In the present study, the high Random Forest hold-out accuracy of 99.70%, together with strong cross-validation and out-of-bag results, demonstrates effective classification within the compiled dataset but does not establish reliability under independent farming conditions. This distinction is particularly important because the climate–pH model achieved 95.23% accuracy without nutrient predictors, indicating that multiple predictor groups contain substantial information for distinguishing the recorded crop classes. Future research should therefore evaluate the model using independently collected observations from previously unseen farms, geographical locations and growing

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

Journal of Agriculture
e-ISSN: 2829-2421
Information Technology and Science (ITScience)

Journal of Agriculture (JoA)
Volume: 05 | Number 03 | November 2026
E-ISSN: 2829-2421
https://doi.org/10.47709/JoA.v5n3.9880

seasons. In Rokan Hulu, this could involve collecting traceable soil nutrient measurements, soil pH, temperature, humidity and rainfall data, followed by expert assessments of crop suitability and observations of actual crop performance. Such an approach would permit evaluation of the relationships between classification outputs and agronomic outcomes. Model calibration, uncertainty assessment and empirically defined recommendation-deferral criteria should also be investigated before implementing the system in agricultural decision-making.

The principal limitations of this study concern data provenance, restricted within-class nutrient variability and the absence of external validation. Although the supplied CSV supports reproducible computational analysis, the original data sources, redistribution licence, measurement units, geographical coverage, sampling conditions and crop-label generation procedures remain unverified. These limitations are consistent with the concerns raised by Condran et al., (2022) regarding the quality and evaluation of agricultural machine learning datasets. Furthermore, Wies et al., (2023) demonstrated that correlations among predictors can complicate the interpretation of permutation-based variable importance, while Blesch et al., (2024) distinguished between marginal and conditional feature importance when evaluating the contributions of correlated predictors. These methodological considerations are relevant to the present findings because nitrogen and potassium showed a moderate positive Spearman correlation ($\rho = 0.490$), whereas soil pH and rainfall showed a negative association ($\rho = -0.362$). Consequently, the importance assigned to individual predictors may partly reflect shared information rather than independent agronomic effects. Although nitrogen consistently ranked first across both global importance methods, this result should not be interpreted as evidence that nitrogen is universally the most influential factor determining crop suitability. In addition, some crop classes contained very few test observations, including only one pumpkin record in the seed-42 test partition, limiting the precision of class-specific performance estimates. No independent field outcomes, near-duplicate similarity analysis or conditional feature-importance estimates were available. Future investigations should therefore prioritise dataset provenance verification, conditional interpretability analysis and external agronomic validation to determine whether the observed classification patterns can support reliable crop recommendation decisions under real farming conditions.

The analysis used the author-supplied CSV; the exact file is identified by a SHA-256 digest in the accompanying metadata. Code, model parameters and numerical outputs accompany this revision. Earlier materials attribute the data to Kaggle, but the exact source URL and redistribution licence remain unverified; the source CSV is therefore not included in the analysis package. The authors' conflict-of-interest declaration requires confirmation before submission.

Conclusion

The study provides an audit-informed interpretation of Random Forest crop-label classification linking complete class-level nutrient checks, global importance, ablation, and repeated-seed evaluation. High internal accuracy is accompanied by restricted nutrient variability in 29 classes, including two pairs that share fixed triplets. Nitrogen's consistent ranking and the strong climate–pH model demonstrate why predictor importance, predictive necessity, and agronomic effects require distinct interpretations. The scientific contribution lies in connecting model behaviour with documented data structure, highlighting the importance of data quality and interpretability in machine learning-based crop recommendation systems.

Further research should focus on verifying dataset provenance, identifying the measurement units and sampling procedures, and evaluating model performance using independently collected soil and climate data from different agricultural locations and growing seasons. Field validation, particularly in Rokan Hulu, Indonesia, is recommended to assess the reliability of crop

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

Journal of Agriculture (JoA)
Volume: 05 | Number 03 | November 2026
E-ISSN: 2829-2421
https://doi.org/10.47709/JoA.v5n3.9880

recommendations under actual farming conditions. Future studies should also investigate conditional feature importance, compare Random Forest with alternative machine learning algorithms, and incorporate agronomic indicators such as crop growth, yield, and soil fertility characteristics. Additionally, uncertainty assessment and model calibration should be explored to improve the reliability of recommendations. These investigations are essential for determining whether the classification patterns identified in this study can be translated into practical, reliable, and sustainable precision agriculture decision-support systems.

Acknowledgements

The authors express their gratitude to Universitas Rokania for its academic support throughout the research and preparation of this manuscript.

References

Abdullah, A., Iwan, M., & Dani, S. R. (2025). Applying tree based model for crop recommendation system based on soil parameters and weather conditions. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, *11*(1), 228–235. https://doi.org/10.33480/jitk.v11i1.6476

Araújo, S. O., Peres, R. S., Ramalho, J. C., Lidon, F., & Barata, J. (2023). Machine learning applications in agriculture: Current trends, challenges, and future perspectives. *Agronomy*, *13*(12), 2976. https://doi.org/10.3390/agronomy13122976

Blesch, K., Watson, D. S., & Wright, M. N. (2024). Conditional feature importance for mixed data. *AStA Advances in Statistical Analysis*, *108*, 259–278. https://doi.org/10.1007/s10182-023-00477-9

Bouni, M., Hssina, B., Douzi, K., & Douzi, S. (2024). Interpretable machine learning techniques for an advanced crop recommendation model. *Journal of Electrical and Computer Engineering*, *2024*, 7405217. https://doi.org/10.1155/2024/7405217

Chowdhury, R., Nur, F. N., Islam, M. N., Islam, M. N., Das, P., & Afridi, A. S. (2025). SPAS-Dataset-BD: Dataset for smart precision agriculture system in Bangladesh. *Data in Brief*, *61*, 111727. https://doi.org/10.1016/j.dib.2025.111727

Condran, S., Bewong, M., Islam, M. Z., Maphosa, L., & Zheng, L. (2022). Machine learning in precision agriculture: A survey on trends, applications and evaluations over two decades. *IEEE Access*, *10*, 73786–73803. https://doi.org/10.1109/ACCESS.2022.3188649

Govindwar, R., Jawale, S., Kalpande, T., Zade, S., Futane, P., & Williams, I. (2023). Crop and fertilizer recommendation system using machine learning. In *AI, IoT, big data and cloud computing for Industry 4.0* (pp. 139–149). Springer. https://doi.org/10.1007/978-3-031-29713-7_7

Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? *Advances in Neural Information Processing Systems*, *35*, 507–520. https://proceedings.neurips.cc/paper_files/paper/2022/file/0378c7692da36807bdec87ab043cdadc-Paper-Datasets_and_Benchmarks.pdf

Islam, M. R., Oliullah, K., Kabir, M. M., Alom, M., & Mridha, M. F. (2023). Machine learning enabled IoT system for soil nutrients monitoring and crop recommendation. *Journal of Agriculture and Food Research*, *14*, 100880. https://doi.org/10.1016/j.jafr.2023.100880

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

Journal of Agriculture (JoA)
Volume: 05 | Number 03 | November 2026
E-ISSN: 2829-2421
https://doi.org/10.47709/JoA.v5n3.9880

Kiran, P. S., Abhinaya, G., Sruti, S., & Padhy, N. (2024). A machine learning-enabled system for crop recommendation. *Engineering Proceedings*, *67*(1), 51. https://doi.org/10.3390/engproc2024067051

Lazuardi, M. R., Hadi, M. Z. S., & Kristalina, P. (2025). Comparative performance analysis of IoT-based crop recommendation systems using Random Forest and GridSearchCV hyperparameter tuning. *International Journal on Advanced Science, Engineering and Information Technology*, *15*(6), 1706–1713. https://doi.org/10.18517/ijaseit.15.6.20997

Loecher, M. (2024). Debiasing SHAP scores in random forests. *AStA Advances in Statistical Analysis*, *108*, 427–440. https://doi.org/10.1007/s10182-023-00479-7

Rani, S., Mishra, A. K., Kataria, A., Mallik, S., & Qin, H. (2023). Machine learning-based optimal crop selection system in smart agriculture. *Scientific Reports*, *13*, 15997. https://doi.org/10.1038/s41598-023-42356-y

Ryo, M. (2022). Explainable artificial intelligence and interpretable machine learning for agricultural data analysis. *Artificial Intelligence in Agriculture*, *6*, 257–265. https://doi.org/10.1016/j.aiia.2022.11.003

Shams, M. Y., Gamel, S. A., & Talaat, F. M. (2024). Enhancing crop recommendation systems with explainable artificial intelligence: A study on agricultural decision-making. *Neural Computing and Applications*, *36*, 5695–5714. https://doi.org/10.1007/s00521-023-09391-2

Wies, C., Miltenberger, R., Grieser, G., & Jahn-Eimermacher, A. (2023). Exploring the variable importance in random forests under correlations: A general concept applied to donor organ quality in post-transplant survival. *BMC Medical Research Methodology*, *23*, 209. https://doi.org/10.1186/s12874-023-02023-2

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License