The Chitransh Academic & Research
An International Journal of Multidisciplinary Research

An Online Peer Reviewed / Refereed Journal
Volume 2 | Issue 2 |June 2026
ISSN: 3049-303X (Online)
Website: www.thechitranshacadmic.in

From Soil Prediction to Fertilizer Prescription: Machine Learning, Uncertainty, and Agronomic Validation for Sustainable Nutrient Management

Chandrakanta
Department of Computer Science Engineering RVS College of Engineering, Jamshedpur, Jharkhand
Nimmi A. Ekka
Lecturer Department of Computer Science Engineering Government Women Polytechnic, Ranchi, Jharkhand
Dr. Jaidev Kumbhakar
Lecturer Department of Mechanical Engineering Cambridge Institute of Polytechnic, Ranchi, Jharkhand

ARTICLE DETAILS

Research Paper
Received: **06/06/2026**
Accepted: **15/06/2026**
Published: **30/06/2026**

Keywords: soil fertility; machine learning; fertilizer recommendation; precision agriculture; nutrient-use efficiency

ABSTRACT

Machine learning is increasingly used for soil-property prediction, fertility classification, nutrient assessment, and fertilizer recommendation by integrating laboratory data, sensors, spectroscopy, remote sensing, terrain, weather, crop observations, and management records. However, accurate prediction of soil properties or yield does not automatically ensure agronomically valid fertilizer recommendations. Nutrient prescription also requires soil-test interpretation, crop-response calibration, fertilizer recovery, seasonal weather, input costs, environmental risks, spatial transferability, and uncertainty assessment. This review examines the transition from machine-learning-based soil assessment to operational nutrient management and proposes a soil-science-informed framework for precision agriculture. Existing studies often emphasize prediction accuracy while giving limited attention to external validation, uncertainty, nutrient-specific processes, and field testing. The proposed framework integrates observation, data harmonization, soil and crop prediction, agronomic diagnosis, response estimation, optimization, and field feedback. Sustainable recommendations should balance yield, profitability, nutrient-use efficiency, environmental risk, feasibility, and uncertainty through multi-location validation, causal modelling, and farmer-centered evaluation.

Introduction

Soil fertility is a fundamental determinant of crop productivity, food security, farm profitability, and environmental quality. It refers to the capacity of soil to support plant growth by supplying essential nutrients while maintaining suitable chemical, physical, and biological conditions. Although nitrogen, phosphorus, and potassium receive the greatest attention in fertilizer management, soil fertility is also influenced by soil reaction, organic carbon, cation-exchange capacity, salinity, texture, structure, water availability, biological activity, rooting depth, and interactions among nutrients. The same nutrient concentration may support adequate crop growth in one soil and produce deficiency symptoms in another because nutrient availability depends on soil mineralogy, pH, moisture, temperature, crop species, management history, and the analytical method used to measure the nutrient. Sustainable soil-fertility management must therefore consider the broader soil–crop–environment system rather than treat nutrient concentrations as isolated variables (FAO, 2017; FAO, 2019).

Fertilizer use has played a major role in increasing agricultural production, but inefficient nutrient management continues to create both agronomic and environmental problems. Underapplication can reduce yield, accelerate nutrient mining, and weaken long-term soil productivity. Excessive or poorly timed application can reduce fertilizer-use efficiency and increase nitrate leaching, ammonia volatilization, nitrous-oxide emissions, phosphorus loss, soil acidification, salinization, and production costs. Blanket fertilizer recommendations are particularly problematic in heterogeneous fields because they assume that soil nutrient supply, crop response, and loss risk are spatially uniform. In reality, erosion, deposition, irrigation, drainage, manure history, former field boundaries, crop rotation, and variable fertilizer application can produce substantial within-field differences.

Conventional soil testing remains the principal method for evaluating soil nutrient status. Laboratory measurements provide analytically defined indicators of soil pH, salinity, organic matter, macronutrients, micronutrients, and related properties. However, conventional sampling and analysis are limited by cost, turnaround time, laboratory capacity, and sampling density. A small number of composite samples may fail to represent local variation, particularly when the field contains contrasting soil types or management histories. Soil testing also provides only a partial temporal representation because nutrient transformations, moisture conditions, crop uptake, and weather can change rapidly during a growing season.

Precision agriculture offers tools for managing this variability more effectively. Proximal soil sensors, visible and near-infrared spectroscopy, mid-infrared spectroscopy, electrical-conductivity mapping, portable X-ray fluorescence, ion-selective sensors, unmanned aerial vehicles, satellite imagery, weather stations, digital elevation models, crop sensors, and yield monitors can generate large amounts of spatially referenced information. Farm-management records provide additional information about fertilizer rates, organic amendments, irrigation, tillage, crop rotation, and previous yields. Machine-learning algorithms can combine these heterogeneous data sources and identify nonlinear relationships that may be difficult to represent using conventional statistical models.

Machine learning has consequently become an important research area in digital soil mapping, soil spectroscopy, fertility classification, nutrient-deficiency detection, crop-yield prediction, and fertilizer recommendation. Frequently used methods include regularized regression, partial least-squares regression, random forests, gradient-boosting algorithms, support-vector machines, Gaussian processes, artificial neural networks, convolutional neural networks, and model ensembles. Reviews by Padarian et al. (2020), Wadoux et al. (2020), Ennaji et al. (2023), and Sujatha and Jaidhar (2024) document the rapid expansion of these methods in soil and nutrient research.

Despite this progress, the literature contains an important conceptual weakness. Soil-property prediction, nutrient diagnosis, crop-response estimation, and fertilizer prescription are often treated as though they were a single modelling task. In practice, they require different types of data and different forms of validation. A model may predict extractable phosphorus accurately but still fail to determine whether a crop will respond economically to additional phosphorus. A model may predict yield accurately under historically observed fertilizer rates without accurately estimating what would happen if the rate were increased or decreased. A system may correctly classify a soil as low in potassium but still recommend an unsuitable fertilizer rate if soil buffering, residual fertilizer effects, crop removal, fertilizer price, and expected recovery are ignored.

This distinction has been demonstrated in recent on-farm research. De Lara et al. (2023) compared machine-learning approaches for predicting site-specific economically optimal nitrogen rates using on-farm wheat and barley experiments. Although models could produce similar yield-prediction accuracy, their estimated optimal nitrogen rates differed substantially. Tanaka et al. (2024) similarly showed that economically optimal fertilizer-rate estimates were sensitive to algorithm and covariate selection. These findings indicate that prediction accuracy alone is insufficient for evaluating a fertilizer-recommendation system. The quality of the resulting decision must also be assessed.

The present paper develops a critical integrative review and soil-science-informed framework for moving from soil prediction to fertilizer prescription. Its central argument is that machine learning should not be evaluated only as a predictive tool. It should be evaluated as one component of a larger decision system that includes soil measurement, agronomic calibration, crop-response estimation, economic analysis, environmental safeguards, uncertainty assessment, field validation, and continuous updating. The review therefore focuses on the entire prediction-to-prescription pathway rather than on algorithm comparison alone.

Review approach

This paper adopts a critical integrative review approach because the relevant literature is highly heterogeneous. Studies differ in soil properties, crops, sensors, analytical methods, sample sizes, spatial scales, climates, algorithms, validation procedures, fertilizer treatments, and decision outcomes. A pooled statistical meta-analysis would be difficult to interpret without extensive harmonization of these differences. The purpose of the review is therefore to integrate methodological and empirical evidence, identify recurring weaknesses, and develop a decision-oriented framework.

The review draws on peer-reviewed literature concerning digital soil mapping, soil spectroscopy, soil-fertility assessment, machine learning, fertilizer recommendation, soil-test calibration, on-farm experimentation, nutrient-use efficiency, uncertainty quantification, spatial validation, and soil-science-informed modelling. Foundational studies were retained where they established widely used concepts or methods, while recent studies were prioritized when they addressed external validation, field testing, uncertainty, or fertilizer decision quality.

Studies were interpreted according to the quality of their reference data, sampling design, analytical documentation, validation strategy, uncertainty assessment, environmental transferability, agronomic calibration, field validation, economic analysis, environmental evaluation, and reproducibility. Particular attention was given to whether the validation method reflected the intended deployment environment. A random split within one spatially clustered dataset was not considered equivalent to validation in a new field, region, year, or laboratory.

The review also distinguishes between prediction evidence and decision evidence. Prediction evidence concerns the ability of a model to estimate a soil property, crop status, or yield. Decision evidence concerns the ability of the resulting recommendation to improve yield, profit, nutrient-use efficiency, soil condition, or environmental performance. This distinction is essential because a high coefficient of determination or low root-mean-square error does not establish that the recommendation will improve farm outcomes.

Table 1. Positioning of the present review relative to major previous reviews

Publication	**Primary emphasis**	**Main contribution**	**Gap addressed in the present paper**
Padarian et al. (2020)	Machine learning across soil science	Broad overview of machine-learning applications	Limited focus on fertilizer decision quality
Wadoux et al. (2020)	Digital soil mapping	Sampling, validation, uncertainty, interpretation, and soil knowledge	Limited treatment of crop response and fertilizer prescription
Ennaji et al. (2023)	Machine learning in nutrient management	Nutrient prediction, classification, sensing, and management applications	Limited formal distinction between prediction and prescription
Sujatha and Jaidhar	Soil-fertility prediction and	Review of machine-	Limited field-validated

(2024)	classification	learning approaches to fertility assessment	recommendation evidence
Minasny et al. (2024)	Soil-science-informed machine learning	Integration of soil knowledge into machine-learning systems	Focused mainly on reliable prediction rather than complete nutrient decisions
Siatwiinda et al. (2024)	Critical soil nutrient levels	Demonstrated methodological and regional variation in critical values	Did not connect soil-test thresholds with machine-learning decision pipelines
Present paper	Prediction-to-prescription transition	Integrates prediction, diagnosis, response estimation, uncertainty, optimization, and field validation	Provides a complete nutrient-decision architecture

Evidence synthesis

The reliability of any machine-learning model depends first on the quality of its reference data. Laboratory soil measurements remain the principal reference for supervised fertility modelling because they provide standardized observations of pH, electrical conductivity, soil organic carbon, total nitrogen, mineral nitrogen, extractable phosphorus, exchangeable potassium, sulfur, calcium, magnesium, and micronutrients. However, analytical results are method-dependent. Phosphorus values obtained using different extractants cannot be combined without careful harmonization. Organic carbon measured by dry combustion may not be directly interchangeable with estimates obtained using loss on ignition. Sampling depth, sample preparation, laboratory batch effects, detection limits, moisture status, and quality-control procedures can also affect the values used to train a model.

This analytical dependence has important consequences for machine learning. A model may learn differences among laboratories or extraction methods rather than genuine soil variation. It may also perform well within one analytical system but fail when applied to another. Studies intended for regional deployment should therefore record laboratory identity, analytical method, sampling depth, units, detection limits, and quality-control procedures. These variables may need to be used for harmonization, calibration transfer, or hierarchical modelling.

Soil-test values must also be interpreted carefully. A soil-test concentration is not automatically equivalent to the amount of nutrient available for crop uptake during a growing season. Soil tests are generally diagnostic indices whose agronomic meaning depends on correlation and calibration experiments. These experiments determine whether crop response becomes more or less likely as the soil-test value changes. Pearce et al. (2022) and Slaton et al. (2022) emphasize the importance of standardized methods and metadata for soil-test correlation and calibration. Siatwiinda et al. (2024)

further demonstrate that critical nutrient levels vary with analytical method, crop, soil condition, experimental design, site characteristics, and statistical procedure. Soil spectroscopy provides a promising means of increasing observation density. Visible and near-infrared and mid-infrared spectra contain information related to organic matter, clay minerals, carbonates, moisture, and other soil constituents. Viscarra Rossel et al. (2016) demonstrated the potential of large spectral libraries for characterizing soil properties across diverse regions. Spectroscopic models can reduce the marginal cost and time required to estimate multiple properties, particularly when supported by reliable laboratory calibration.

Nevertheless, spectroscopic prediction of individual nutrients requires caution. Some nutrients do not have strong direct spectral features and may be predicted indirectly through their correlations with organic matter, clay, mineralogy, salinity, parent material, or geographic location. A model may therefore appear accurate within the calibration region but fail where those correlations change. Local updating, calibration transfer, and external testing are essential when models are moved between instruments, laboratories, moisture conditions, soil groups, or regions. Padarian et al. (2022), Omondiagbe et al. (2024), and Huang et al. (2025) show that uncertainty assessment is especially important when spectral models encounter observations outside their calibration domain.

Remote sensing adds repeated spatial information on bare-soil reflectance, vegetation vigor, chlorophyll status, biomass, canopy temperature, and crop development. Satellite and unmanned-aerial-vehicle imagery can support soil and crop assessment at scales difficult to achieve through conventional sampling. However, vegetation indices are not unique indicators of nutrient deficiency. Water stress, disease, weeds, compaction, cultivar differences, crop stage, temperature, and management can create similar spectral responses. Crop-sensing data should therefore be combined with soil, weather, management, and phenological information rather than interpreted as direct nutrient measurements. Terrain and climate variables provide additional information about soil-forming processes and nutrient dynamics. Elevation, slope, curvature, flow accumulation, wetness indices, rainfall, temperature, evapotranspiration, parent material, and drainage influence erosion, deposition, organic-matter accumulation, mineral weathering, water availability, and nutrient loss. Management history is equally important. Fertilizer use, manure application, crop rotation, irrigation, tillage, residue management, liming, and previous yields can explain substantial fertility variation.

Management variables can also introduce confounding. A positive relationship between fertilizer rate and yield does not necessarily represent a causal fertilizer effect. Farmers may apply more fertilizer to productive fields, irrigated areas, or crops expected to have higher yield potential. Historical fertilizer rate may therefore serve as a proxy for field quality or farmer expectations. Predictive models should not automatically interpret such associations as evidence of treatment response.

Table 2. Data sources for machine-learning-based soil-fertility assessment

Data source	**Representative variables**	**Principal contribution**	**Major limitations**
Laboratory soil analysis	pH, salinity, organic carbon, nitrogen, phosphorus, potassium, sulfur, micronutrients	Reference measurements and agronomic diagnosis	Cost, sparse sampling, analytical incompatibility, depth

			mismatch
Soil spectroscopy	Visible, near-infrared, shortwave-infrared, and mid-infrared spectra	Rapid estimation of multiple properties	Moisture effects, instrument differences, indirect nutrient prediction
Proximal sensors	Apparent electrical conductivity, portable X-ray fluorescence, pH and ion-selective sensors	Dense within-field observations	Drift, interference, calibration dependence, maintenance
Satellite and UAV data	Bare-soil reflectance, vegetation indices, red-edge and thermal information	Spatial and temporal crop–soil information	Clouds, mixed pixels, confounding stress

			factors
Terrain data	Elevation, slope, curvature, wetness, flow accumulation	Soil redistribution, drainage, and landscape context	Scale dependence and resolution mismatch
Weather data	Rainfall, temperature, evapotranspiration, drought indices	Nutrient transformation and seasonal crop demand	Forecast uncertainty and sparse observations
Crop observations	Crop type, stage, tissue nutrients, biomass, yield	Crop-response and demand assessment	Crop specificity and measurement error
Management records	Fertilizer, manure, rotation, tillage, irrigation, liming	Explanation of management-induced variation	Missing data, reporting bias, confounding
Economic information	Crop price, fertilizer price,	Profitability and risk	Price volatilit

	labor, machinery cost	evaluation	y and regional differences

Machine-learning algorithms differ in their ability to represent nonlinear relationships, interactions, high-dimensional predictors, and uncertainty. Regularized linear models and partial least-squares regression remain useful baselines because they are transparent, computationally efficient, and suitable for correlated predictors. Random forests and gradient-boosting algorithms are widely used because they handle nonlinear relationships, mixed data types, and complex interactions with relatively limited preprocessing. Support-vector machines can perform well with moderate sample sizes, while Gaussian-process models provide probabilistic outputs but may be computationally demanding. Deep-learning approaches are valuable for images, spectra, and time series, particularly when large training datasets are available.

Model complexity should match data quantity and structure. A deep neural network may not provide a meaningful advantage over a tree ensemble when the dataset contains only a few hundred tabular observations. Conversely, convolutional or transformer-based architectures may be justified when the inputs include hyperspectral sequences, imagery, or large temporal datasets. Every study should include simple baselines so that the benefit of additional complexity can be demonstrated.

The literature does not support the existence of one universally superior algorithm. Performance depends on target property, sample size, spatial scale, sensor type, analytical method, predictor availability, and validation design. A model that performs well for organic carbon may perform poorly for extractable phosphorus. A model optimized for interpolation within one region may fail during geographic transfer. Algorithm rankings may also change when evaluation moves from random validation to spatially independent validation.

Spatial validation is one of the most important methodological issues in soil machine learning. Agricultural samples are frequently spatially autocorrelated. Nearby observations share climate, terrain, parent material, soil type, and management history. When neighboring samples are randomly divided between training and testing sets, the test data remain similar to the training data. This can create overly optimistic estimates of performance.

Roberts et al. (2017) show that validation strategies should reflect spatial, temporal, and hierarchical dependence. Wadoux et al. (2020) similarly identify spatial validation as a major challenge in digital soil mapping. For prediction in unsampled locations, spatial-block cross-validation is more appropriate than random splitting. Leave-one-field-out validation is suitable when the intended use is prediction in new fields. Leave-one-farm-out validation assesses transfer across enterprises, while leave-one-region-out validation evaluates broader geographic generalization. Temporal validation is required when recommendations will be applied in future seasons.

External validation is particularly important for operational systems. A model should be tested using samples collected independently from the original training dataset. Ideally, the external dataset should differ in field, year, soil group, laboratory, sensor, or management system. Internal cross-validation can support model development, but it cannot fully establish deployment reliability.

Environmental extrapolation must also be evaluated. Geographic distance is not the only relevant consideration. A nearby field may contain a combination of soil, climate, and management characteristics absent from the training data. Conversely, a distant site may be environmentally similar. Hateffard et al. (2024) demonstrate the importance of assessing whether prediction locations are represented in the environmental feature space of the calibration data.

Applicability should therefore be treated as a continuum rather than a simple geographical boundary. Useful indicators include environmental dissimilarity, distance to similar training observations, local training density, ensemble disagreement, and prediction-interval width. A model should not automatically generate a fertilizer prescription where the environmental combination is poorly represented.

Uncertainty is another major limitation of current systems. A nutrient-management decision contains multiple sources of uncertainty. These include sampling error, laboratory measurement error, sensor uncertainty, model uncertainty, spatial extrapolation, crop-response variability, weather uncertainty, fertilizer recovery, price variability, and environmental-loss uncertainty. Many studies report only a point prediction and ignore the other components.

Prediction intervals provide a more informative output than point estimates alone. Quantile regression forests, Gaussian processes, bootstrap ensembles, Bayesian methods, Monte Carlo dropout, and conformal prediction can be used to estimate uncertainty. Kasraei et al. (2021) evaluated quantile regression as an uncertainty method for digital soil mapping. Padarian et al. (2022) demonstrated that Monte Carlo dropout can produce wider and more responsive intervals for out-of-domain soil spectral observations. Huang et al. (2025) further showed that conformal approaches can improve uncertainty calibration in deep-learning soil spectral models.

Uncertainty estimates must themselves be validated. A nominal prediction interval is only useful when its actual coverage is close to the intended coverage. Studies should report interval coverage, average interval width, coverage across soil groups and concentration ranges, and performance in external datasets. Narrow intervals are not desirable when they exclude the true value too frequently.

The greatest weakness in many nutrient-management studies is the failure to propagate uncertainty from soil prediction into fertilizer decisions. A narrow interval around a predicted soil property does not necessarily produce a narrow interval around an optimal fertilizer rate. Crop-response parameters, fertilizer recovery, yield potential, prices, and nutrient-loss processes can amplify uncertainty. The final output should therefore include a plausible range of fertilizer rates, yields, profits, and environmental outcomes.

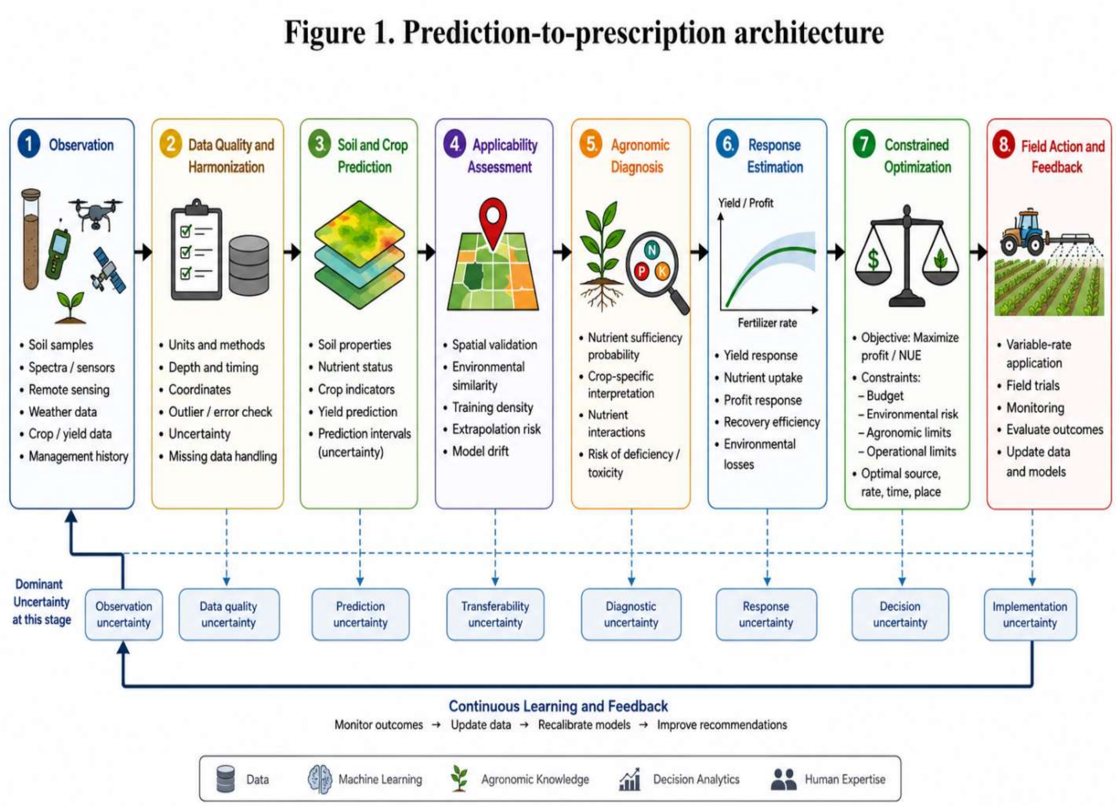

Figure 1. Prediction-to-prescription architecture

The distinction between soil prediction and agronomic diagnosis is critical. A predicted soil-test value becomes useful for fertilizer management only after it is interpreted using crop- and region-specific calibration. Critical soil-test values are derived from experiments that relate relative yield or crop response to measured soil-test levels. The resulting threshold depends on analytical method, crop, soil, response model, experimental design, and the chosen definition of adequate yield.

A universal threshold cannot be assumed to apply across regions or extraction methods. Siatwiinda et al. (2024) document substantial variation in critical nutrient values and show that site and methodological factors contribute to this variation. A model that predicts a phosphorus concentration accurately may still generate an incorrect deficiency classification if the threshold was transferred from a different soil or extraction system.

Probabilistic diagnosis is often preferable to rigid classification. Instead of classifying every field as low, medium, or high, a decision-support system can estimate the probability that a crop will respond profitably to additional nutrient application. Such probabilities better represent uncertainty near a critical soil-test value. They also allow farmers and advisers to adjust decisions according to economic risk and environmental sensitivity.

Crop-response estimation requires different evidence from soil-property prediction. Fertilizer recommendation concerns what will happen when a management action changes. Historical farm data may contain strong associations between fertilizer rate and yield, but these associations may be confounded. Randomized multi-rate experiments provide the strongest basis for estimating fertilizer response.

On-farm precision experiments are particularly valuable because they generate response information under commercial conditions. Variable-rate equipment can impose several fertilizer rates across a field, while yield monitors provide spatially dense outcome data. These experiments can reveal how fertilizer response changes with soil properties, terrain, weather, and management.

De Lara et al. (2023) analyzed site-specific economically optimal nitrogen rates using 20 on-farm wheat and barley experiments. Their findings showed that similar yield-prediction accuracy could produce different optimal nitrogen recommendations. Tanaka et al. (2024) reached a similar conclusion by evaluating several machine-learning algorithms and covariate combinations. These studies demonstrate that fertilizer recommendations should be evaluated directly against experimentally derived response and economic optima.

Colaço et al. (2024) compared multiple digital nitrogen-management methods using large-scale on-farm trials. The study illustrates the value of benchmarking decision tools against local response-derived optima rather than only against farmer practice or regional recommendations. Liben et al. (2024) likewise emphasizes the need to evaluate data-driven fertilizer recommendations through field outcomes such as productivity and resource-use efficiency.

Different nutrients require different decision pathways. Nitrogen is highly dynamic and is influenced by mineralization, immobilization, leaching, volatilization, denitrification, crop uptake, previous crops, irrigation, and weather. Nitrogen recommendations should therefore be updated during the season when possible. Crop sensing, weather forecasts, soil mineral nitrogen, and yield potential can support split-application decisions.

Phosphorus and potassium decisions are commonly based on soil-test calibration, crop-response probability, soil buffering, residual effects, and crop removal. Phosphorus recommendations must consider runoff and erosion risk, while potassium recommendations may require information about cation-exchange capacity, clay mineralogy, fixation, and luxury uptake. Soil-test values for these nutrients should not be treated as direct estimates of kilograms available to the crop.

Lime recommendation requires more than soil pH. Buffer pH, exchangeable acidity, soil organic matter, clay content, crop sensitivity, target pH, amendment neutralizing value, fineness, incorporation depth, and expected reaction time are relevant. Micronutrient recommendations similarly require crop-specific sensitivity, method-specific critical levels, pH interactions, toxicity risk, and confirmation through tissue analysis or field response.

Table 3 summarizes nutrient-specific considerations.

Nutrient or amendment	**Essential information**	**Appropriate decision basis**	**Main risk of oversimplification**
Nitrogen	Mineral nitrogen, organic carbon, crop stage, weather, previous crop, irrigation	Seasonal response and economically optimal rate	Large seasonal variation and environmental loss
Phosphorus	Method-specific soil test, pH, buffering,	Calibrated response probability and critical level	Treating concentration as direct nutrient supply

	erosion and runoff risk		
Potassium	Soil-test potassium, cation-exchange capacity, clay mineralogy, crop removal	Crop response, residual effects, and maintenance strategy	Ignoring fixation, buffering, and luxury uptake
Sulfur	Sulfate sulfur, organic matter, rainfall, irrigation inputs	Seasonal supply and crop-response evidence	High temporal variability and leaching
Lime	Soil pH, buffer pH, acidity, crop sensitivity, material quality	Calibrated lime-requirement procedure	Recommending lime from pH alone
Micronutrients	Extractant, soil pH, organic matter, crop sensitivity, tissue status	Deficiency probability and field confirmation	Toxicity or unnecessary application
Organic	Nutrient	Multi-season	Treating total

amendments	analysis, carbon-to-nitrogen ratio, mineralization, contaminants	nutrient availability	nutrient content as immediately available

A complete fertilizer recommendation includes source, rate, timing, and placement. Rate-only systems provide an incomplete decision. Source affects nutrient availability, salt effects, acidification, compatibility, and cost. Timing affects synchronization with crop demand and environmental-loss risk. Placement affects root access, fixation, runoff, volatilization, and machinery requirements.

Economic and environmental objectives must also be integrated. A recommendation that maximizes predicted yield may not maximize profit because the final increment of fertilizer may cost more than the additional crop value. Similarly, an economically optimal rate may exceed an environmentally acceptable rate in a sensitive watershed. A sustainable system should evaluate crop value, fertilizer cost, application cost, expected yield response, probability of financial loss, nutrient-use efficiency, nitrate leaching, phosphorus loss, greenhouse-gas emissions, soil pH, and salinity.

Machine-learning recommendations should therefore be risk-aware. A farmer with limited cash reserves may prefer a lower but more stable return rather than a high expected return with a large probability of loss. Environmental regulators may require stricter constraints in areas vulnerable to groundwater contamination or eutrophication. The system should make these assumptions explicit rather than hide them within an algorithm.

Human oversight remains essential because models can fail silently. Three operational levels are appropriate. Low-risk recommendations may be automated when the site is well represented, uncertainty is low, the recommendation is consistent with established local guidance, and the consequences of moderate error are limited. Agronomist review is required when uncertainty is moderate, indicators conflict, the recommendation differs substantially from standard practice, or environmental risk is high. Additional sampling or field experimentation is necessary when the site falls outside the applicability domain, prediction intervals are wide, analytical quality is uncertain, or local crop-response calibration is absent.

Explainable artificial intelligence can support this review process. Variable importance, SHAP values, partial-dependence plots, accumulated-local-effects plots, and local explanations can indicate which variables influenced a prediction. However, explanation methods do not establish causality. Correlated variables may divide importance unpredictably, and coordinates may act as proxies for unobserved environmental gradients. A model may appear interpretable while relying on relationships that are unstable outside the training region.

Agronomic auditing should therefore accompany algorithmic explanation. Model relationships should be checked across spatial-validation folds and evaluated by soil scientists and agronomists. Predictions should be examined for plausible nutrient ranges, depth patterns, soil-property interactions, nutrient ratios, and responses to pH, organic matter, texture, weather, and management. Excessive reliance on coordinates, laboratory identifiers, or unavailable deployment variables should be treated as a warning sign.

Proposed decision framework

The proposed framework begins with a clearly defined decision problem. Researchers should specify the crop, nutrient, soil depth, geographic scale, intended user, timing of the decision, and acceptable level of risk. A model designed to map regional soil organic carbon is not automatically suitable for within-field fertilizer prescription. Similarly, a model developed for one crop or analytical method should not be transferred to another without validation.

Sampling should represent the environmental and management variation relevant to deployment. Stratification may be based on soil group, terrain, management zone, crop history, irrigation, and expected fertility gradients. Sampling designs should avoid overrepresenting accessible or highly productive fields. When recommendations will be made at subfield scale, the sampling density should be sufficient to characterize variation at the scale of management.

Data harmonization should occur before modelling. Units, analytical methods, depth intervals, coordinate systems, sampling dates, sensor conditions, spatial resolutions, and temporal alignment should be standardized. Outliers should not be removed simply because they reduce model performance. Extreme values may represent genuine salinity, acidity, nutrient depletion, or localized amendment history. Removal should require analytical or field-based justification.

Feature engineering should reflect soil and crop processes. Useful predictors may include cumulative rainfall, growing-degree days, previous fertilizer inputs, crop nutrient removal, terrain position, moisture status, crop stage, and interactions among pH, organic matter, texture, and nutrient availability. Feature selection, imputation, scaling, and transformation must be conducted within the resampling procedure to prevent information leakage.

Model development should compare simple and complex approaches using identical validation partitions. Performance should be reported using multiple metrics rather than one accuracy value. Continuous predictions should include absolute error, squared error, bias, concordance, residual spatial structure, and interval coverage. Classification studies should report class-specific metrics and probability calibration.

Validation should reflect the intended deployment. Spatially blocked or leave-one-field-out validation should be the primary approach for new-field prediction. Temporal validation should be used for future-season decisions. Independent external testing should be included wherever possible. Applicability-domain and environmental-dissimilarity indicators should accompany every map or recommendation.

Agronomic diagnosis should be based on locally relevant soil-test calibration. Predictions should be converted into probabilities of crop response rather than rigid classes when sufficient evidence is available. Nutrient interactions, soil acidity, salinity, rooting limitations, moisture, crop stage, and management history should be considered before a deficiency diagnosis is translated into a fertilizer action.

Crop-response estimation should rely on randomized multi-rate field experiments, on-farm precision experiments, valid quasi-experimental designs, or carefully constructed response models. Historical observations alone are often insufficient because treatment selection may be confounded. Machine-learning models should be evaluated according to how well they estimate response and optimal decisions, not only observed yield.

The prescription stage should combine crop response with fertilizer cost, crop price, recovery efficiency, environmental-loss risk, operational constraints, and farmer preferences. Source, rate, timing, and placement should be specified together. The output should include uncertainty ranges and decision alternatives. High uncertainty may justify additional sampling, split application, tissue testing, or a small field trial.

Field validation should compare the machine-learning recommendation with farmer practice, conventional regional recommendations, laboratory soil-test recommendations, and other decision tools. Outcomes should include yield, profit, nutrient-use efficiency, soil-test change, nitrate loss, phosphorus risk, greenhouse-gas emissions, and farmer acceptance. Multi-field and multi-season evaluation is essential because one season cannot represent the full range of weather and price conditions.

Monitoring should continue after deployment. Changes in analytical methods, sensors, crop cultivars, management practices, weather patterns, and prices can reduce model performance. Prediction errors, interval coverage, out-of-domain cases, recommendation stability, and farmer feedback should be monitored. Models should be recalibrated, restricted, or withdrawn when performance thresholds are no longer met.

Table 4 provides a minimum reporting checklist for studies intended to support operational nutrient management.

Domain	Minimum information
Objective	Clearly distinguish soil prediction, nutrient diagnosis, response estimation, and prescription
Study area	Region, climate, soil groups, crops, fields, farms, seasons, and management systems
Sampling	Design, number of samples, depth, date, density, compositing, and coordinate accuracy
Laboratory methods	Extractant, instrument, units, detection limits, and quality control
Predictors	Source, resolution, timing, preprocessing, and availability during deployment
Model development	Algorithms, simple baselines, feature selection, tuning, and software
Leakage prevention	Imputation, scaling, selection, and tuning conducted within resampling

Validation	Random benchmark plus spatial, field, regional, temporal, or external validation
Performance	Error, bias, calibration, interval coverage, and agronomic threshold error
Applicability	Environmental similarity, training-data density, and extrapolation indicators
Uncertainty	Method, calibration, interval width, and external-domain behavior
Agronomic calibration	Crop, soil-test method, critical values, and response evidence
Response estimation	Experimental design, fertilizer rates, confounding control, and response model
Recommendation	Source, rate, timing, placement, prices, constraints, and risk assumptions
Field testing	Comparator treatments, replication, seasons, yield, profit, and environmental outcomes
Explainability	Stability across validation folds and agronomic interpretation
Reproducibility	Data, code, software versions, random seeds, and model metadata
Deployment	Intended users, monitoring, updating, human review, and withdrawal criteria

Discussion

The principal contribution of this review is the separation and reconnection of soil prediction, agronomic diagnosis, crop-response estimation, and fertilizer prescription. These stages are often combined in machine-learning studies, but they depend on different scientific assumptions and validation requirements. Soil prediction describes current or expected conditions. Agronomic diagnosis estimates whether those conditions are likely to limit crop performance. Crop-response estimation evaluates how yield or nutrient uptake will change under an intervention. Prescription selects an action according to economic, environmental, operational, and risk considerations.

This distinction changes how machine-learning systems should be evaluated. A soil-property model should be assessed for analytical accuracy, spatial transfer, and uncertainty. A diagnostic model should be assessed for its ability to identify true nutrient-responsive conditions. A response model should be assessed against experimental fertilizer treatments. A prescription system should be assessed using decision quality, profit, nutrient-use efficiency, environmental impact, and field performance.

Soil-science-informed machine learning provides an important foundation for this approach. Minasny et al. (2024) argue that soil knowledge can be integrated through predictor design, model structure, constraints, training objectives, and interpretation. Such integration can reduce implausible outputs, improve transferability, and increase data efficiency. Soil-informed modelling is especially important where training data are limited or environmentally restricted.

Domain knowledge can be incorporated by preserving analytical-method information, modelling soil depth explicitly, constraining impossible values, accounting for nutrient interactions, including soil-forming factors, and checking predictions against known soil processes. Hybrid models can combine process-based crop or nutrient models with machine learning. Process models provide mechanistic structure, while machine learning can correct systematic errors or represent relationships not captured by the process model.

The paper also highlights the need for decision-centered evaluation. Conventional predictive metrics do not directly measure the cost of a wrong recommendation. Two models with similar prediction errors may select different fertilizer rates. The more relevant question is how much yield, profit, or environmental performance is lost because the model selected an imperfect action. Future research should therefore report decision regret, rate error, profit loss, and environmental consequences.

The literature remains much stronger for soil-property prediction than for field-validated nutrient recommendation. Numerous studies compare random forests, support-vector machines, neural networks, and boosting algorithms using soil datasets. Far fewer studies test whether the resulting recommendations improve field outcomes across multiple environments. The central research need is not another small algorithm comparison. It is the development of multi-location, multi-season datasets that link soil measurements, weather, management treatments, crop response, economic outcomes, and nutrient losses.

On-farm experimentation offers a practical way to generate such evidence. Commercial fields can be used to test multiple fertilizer rates under real management conditions. Repeated experiments can provide data for response-oriented machine learning, local calibration, and model updating. Active-learning methods can further improve efficiency by identifying the locations or treatments that would reduce decision uncertainty most effectively.

Transfer learning and hierarchical modelling may help address data scarcity. Large regional or global models can provide an initial prediction, while a smaller number of local samples are used for calibration. However, local updating should not be treated as optional. Global models may capture broad relationships but remain biased in specific soils, climates, or analytical systems.

Smallholder agriculture requires particular attention. Many precision-agriculture systems assume reliable connectivity, expensive sensors, cloud computing, yield monitors, and variable-rate machinery. These assumptions do not apply universally. Lower-cost systems may combine targeted laboratory testing, satellite imagery, weather information, simple management zones, mobile advisory tools, and extension support. Cooperative soil sampling and shared sensing services may provide greater value than highly complex individual-farm technologies.

Data governance is also important. Soil, yield, and management data can reveal land productivity, input use, financial decisions, and production risk. Farmers should understand how their data will be stored, shared, used for model training, and commercialized. Governance should address consent, ownership, access, anonymization, retention, correction, and benefit sharing.

The sustainability claims associated with machine learning must remain evidence-based. A high-accuracy model does not prove reduced nutrient loss, improved soil health, or higher profit. Sustainability outcomes should be measured directly. Studies should distinguish between claimed benefits, simulated benefits, experimentally measured benefits, and long-term observed benefits. These levels of evidence should not be treated as equivalent.

The proposed framework has limitations. It is a critical integrative review rather than an exhaustive systematic review and does not provide formal screening counts or pooled performance estimates. The literature is too heterogeneous to support a single conclusion about the best algorithm or universal fertilizer strategy. The proposed framework is conceptual and requires crop-, nutrient-, soil-, and region-specific calibration. Institutional capacity, farmer trust, market conditions, advisory systems, and policy can also determine whether a technically valid recommendation is adopted.

Nevertheless, the framework provides a coherent basis for future research and journal reporting. It encourages researchers to move beyond isolated prediction exercises and evaluate the full chain from data collection to field outcome. It also provides reviewers and editors with criteria for distinguishing exploratory modelling from operational nutrient-management evidence.

Conclusion

Machine learning can improve soil-fertility assessment by integrating laboratory measurements, sensors, spectroscopy, remote sensing, terrain, weather, crop observations, and management data. However, the value of these technologies depends on whether predictions can be translated into reliable agronomic decisions.

Accurate soil-property or yield prediction is only the first step. Fertilizer prescription also requires crop-specific soil-test calibration, response estimation, economic analysis, environmental safeguards, uncertainty propagation, field validation, and human oversight. Models must be evaluated in new fields and seasons, and recommendations should not be automated where the deployment environment is poorly represented or uncertainty is high.

The most important future research priority is the development of multi-location, multi-season experimental datasets connecting soil conditions, fertilizer treatments, crop response, profit, and environmental outcomes. Machine learning should be positioned as part of an integrated soil–crop decision system rather than as a replacement for soil science, field experimentation, or agronomic judgment.

When prediction, calibration, response estimation, uncertainty, optimization, and field feedback are combined transparently, machine learning can support more efficient fertilizer use, higher profitability, lower environmental risk, and more sustainable agricultural production.

References

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32. https://doi.org/10.1023/A:1010933404324

Chen, T., and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Colaço, A. F., Whelan, B. M., Bramley, R. G. V., Richetti, J., Fajardo, M., McCarthy, A. C., Perry, E. M., Bender, A., Leo, S., Fitzgerald, G. J., and Lawes, R. A. (2024). Digital strategies for nitrogen management in grain production systems: Lessons from multi-method assessment using on-farm experimentation. *Precision Agriculture, 25*, 983–1013. https://doi.org/10.1007/s11119-023-10102-z

De Lara, A., Mieno, T., Luck, J. D., and Puntel, L. A. (2023). Predicting site-specific economic optimal nitrogen rate using machine learning methods and on-farm precision experimentation. *Precision Agriculture, 24*, 1792–1812. https://doi.org/10.1007/s11119-023-10018-8

Ennaji, O., Vergütz, L., and El Allali, A. (2023). Machine learning in nutrient management: A review. *Artificial Intelligence in Agriculture, 7*, 100–118. https://doi.org/10.1016/j.aiia.2023.06.001

Food and Agriculture Organization of the United Nations. (2017). *Voluntary guidelines for sustainable soil management*. FAO.

Food and Agriculture Organization of the United Nations. (2019). *International code of conduct for the sustainable use and management of fertilizers*. FAO.

Hateffard, F., Steinbuch, L., and Heuvelink, G. B. M. (2024). Evaluating the extrapolation potential of random forest digital soil mapping. *Geoderma, 441*, 116740. https://doi.org/10.1016/j.geoderma.2023.116740

Huang, Y.-C., Padarian, J., Minasny, B., and McBratney, A. B. (2025). Using Monte Carlo conformal prediction to evaluate the uncertainty of deep-learning soil spectral models. *SOIL, 11*, 553–563. https://doi.org/10.5194/soil-11-553-2025

Kasraei, B., Heung, B., Saurette, D. D., Schmidt, M. G., Bulmer, C. E., and Bethel, W. (2021). Quantile regression as a generic approach for estimating uncertainty of digital soil maps produced from machine learning. *Environmental Modelling & Software, 144*, 105139. https://doi.org/10.1016/j.envsoft.2021.105139

Liben, F. M., Abera, W., Chernet, M. T., Ebrahim, M., Tilaye, A., Erkossa, T., et al. (2024). Site-specific fertilizer recommendation using data-driven machine learning enhanced wheat productivity and resource-use efficiency. *Field Crops Research, 313*, 109413. https://doi.org/10.1016/j.fcr.2024.109413

Lundberg, S. M., and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems, 30*.

McBratney, A. B., Mendonça Santos, M. L., and Minasny, B. (2003). On digital soil mapping. *Geoderma, 117*, 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4

Minasny, B., Bandai, T., Ghezzehei, T. A., Huang, Y.-C., Ma, Y., McBratney, A. B., Ng, W., Norouzi, S., Padarian, J., Rudiyanto, Sharififar, A., Styc, Q., and Widyastuti, M. (2024). Soil science-informed machine learning. *Geoderma, 452*, 117094. https://doi.org/10.1016/j.geoderma.2024.117094

Omondiagbe, O. P., Roudier, P., Lilburne, L., Ma, Y., and McNeill, S. (2024). Quantifying uncertainty in the prediction of soil properties using mid-infrared spectra.

Geoderma, 448, 116954. https://doi.org/10.1016/j.geoderma.2024.116954

Padarian, J., Minasny, B., and McBratney, A. B. (2020). Machine learning and soil sciences: A review aided by machine learning tools. *SOIL, 6*, 35–52. https://doi.org/10.5194/soil-6-35-2020

Padarian, J., Minasny, B., and McBratney, A. B. (2022). Assessing the uncertainty of deep learning soil spectral models using Monte Carlo dropout. *Geoderma, 425*, 116063. https://doi.org/10.1016/j.geoderma.2022.116063

Pearce, A. W., Slaton, N. A., Lyons, S. E., Bolster, C. H., Bruulsema, T. W., Grove, J. H., et al. (2022). Defining relative yield for soil-test correlation and calibration trials in the Fertilizer Recommendation Support Tool. *Soil Science Society of America Journal, 86*, 293–310. https://doi.org/10.1002/saj2.20450

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography, 40*, 913–929. https://doi.org/10.1111/ecog.02881

Siatwiinda, S. M., Ros, G. H., Yerokun, O. A., et al. (2024). Options to reduce ranges in critical soil nutrient levels used in fertilizer recommendations by accounting for site conditions and methodology: A review. *Agronomy for Sustainable Development, 44*, 9. https://doi.org/10.1007/s13593-023-00943-3

Slaton, N. A., Lyons, S. E., Osmond, D. L., Brouder, S. M., Culman, S. W., Drescher, G., et al. (2022). Minimum dataset and metadata guidelines for soil-test correlation and calibration research. *Soil Science Society of America Journal, 86*, 19–33. https://doi.org/10.1002/saj2.20338

Sujatha, M., and Jaidhar, C. D. (2024). Machine learning-based approaches to enhance soil fertility—A review. *Expert Systems with Applications, 240*, 122557. https://doi.org/10.1016/j.eswa.2023.122557

Tanaka, T. S. T., Heuvelink, G. B. M., Mieno, T., and Bullock, D. S. (2024). Can machine learning models provide accurate fertilizer recommendations? *Precision Agriculture, 25*, 1839–1856. https://doi.org/10.1007/s11119-024-10136-x

Viscarra Rossel, R. A., Behrens, T., Ben-Dor, E., Brown, D. J., Demattê, J. A. M., Shepherd, K. D., et al. (2016). A global spectral library to characterize the world's soil. *Earth-Science Reviews, 155*, 198–230. https://doi.org/10.1016/j.earscirev.2016.01.012

Wadoux, A. M. J.-C., Minasny, B., and McBratney, A. B. (2020). Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews, 210*, 103359. https://doi.org/10.1016/j.earscirev.2020.103359