Artificial Intelligence in Agriculture 9 (2023) 1–11

Contents lists available at ScienceDirect

Artificial Intelligence in Agriculture

journal homepage: http://www.keaipublishing.com/en/journals/artificial-intelligence-in-agriculture/

Machine learning in nutrient management: A review

Oumnia Ennaji [a,b], Leonardus Vergütz [a], Achraf El Allali [b,*]

[a] Chair of Soil Science, University Mohammed VI Polytechnic, Lot 660, Hay Moulay Rachid, Benguerir 43150, Morocco
[b] African Genome Center,University Mohammed VI Polytechnic, Lot 660, Hay Moulay Rachid, Benguerir 43150, Morocco

ARTICLE INFO

Article history:
Received 13 September 2022
Received in revised form 6 April 2023
Accepted 13 June 2023
Available online 20 June 2023

Keywords:
Fertilization
Nutrient management
Machine learning

ABSTRACT

In agriculture, precise fertilization and effective nutrient management are critical. Machine learning (ML) has recently been increasingly used to develop decision support tools for modern agricultural systems, including nutrient management, to improve yields while reducing expenses and environmental impact. ML based systems require huge amounts of data from different platforms to handle non-linear tasks and build predictive models that can improve agricultural productivity. This study reviews machine learning based techniques for estimating fertilizer and nutrient status that have been developed in the last decade. A thorough investigation of detection and classification approaches was conducted, which served as the basis for a detailed assessment of the key challenges that remain to be addressed. The research findings suggest that rapid improvements in machine learning and sensor technology can provide cost-effective and thorough nutrient assessment and decision-making solutions. Future research directions are also recommended to improve the practical application of this technology.

Contents

1. Introduction

Agriculture is the main source of food, income, and employment for developing countries and their rural populations, especially in Africa Preethi et al. (2020); Tapsoba et al. (2020); Collier and Dercon (2014); Rejeb et al. (2022). Given current population growth, pressure on agricultural systems will continue to increase Kamilaris et al. (2017); Bharath et al. (2022). Many countries are agrarian and their economies are based primarily on agricultural productivity Salami et al. (2010); Livingston et al. (2011). African agriculture, for example, is influenced by various factors such as climate, geography, water scarcity, spatial variability of soils, and policies Aworka et al. (2022); Diao et al. (2010); Olanipekun et al. (2019); Jayne et al. (2010). Despite the population explosion and increasing demand in the last century, farmers still suffer from large economic losses due to under-fertilization Chivenge et al. (2022); Jha et al. (2019). Although the amount and quality of experimental data is constantly increasing, researchers are still unable to integrate it, analyze it, and make the best decisions possible. Modern

* Corresponding author.
E-mail address: achraf.elallali@um6p.ma (A. El Allali).

https://doi.org/10.1016/j.aiia.2023.06.001

agriculture proves that the introduction of new technologies has solved many problems that farmers have faced in the last decades Liakos et al. (2018); Patrício and Rieder (2018); Adnan et al. (2019); Elavarasan et al. (2018); Babaie Sarijaloo et al. (2021). Technology transfer centers that promote the adoption of new agricultural management practices and the use of sensors, drones, and low-cost satellite imagery for pest control and better fertilization should help improve the welfare of smallholder farmers Kuijpers and Swinnen (2016); Holzworth et al. (2015). The rise of new data-intensive scientific fields has led to this modern agriculture, which now generates large amounts of data due to the multitude of sensors deployed in experimental fields. Traditional data processing techniques are not up to the ever-increasing demands of smart agriculture, making it difficult to extract useful information from numerous field experiments and soil tests.

Machine learning (ML) is an emerging technology that can help discover patterns in large data sets Sarker (2021). This technology makes predictions directly from given data Ayodele (2010); Singh et al. (2016). ML Algorithms can predict yields using fertilizer rates, genetic data, and environmental and land management variables. Advances in machine learning, a subfield of artificial intelligence, are benefiting agriculture. The digital transformation of agriculture is evolving in parallel with artificial intelligence systems in various aspects, optimizing the ever-growing data coming from numerous sources, not only in crop disease detection, which has received the most attention from data scientists Barbedo (2019). In the area of nutrient management, it is possible to integrate and interpolate various pieces of information that have never been explored for this purpose before. This improves the overall understanding of agricultural systems, including nutrient requirements, and also allows economic aspects to be incorporated into decisions.

Accurate diagnosis of current crop nutritional status and nutrient requirements plays a critical role in overall farm management and impacts not only the environment, but also the economic sustainability of the farm Goulding et al. (2008); Monaghan et al. (2007); Fairhurst et al. (2007); Dhal et al. (2022). Yield loss, under-utilization of natural resources, decline in soil organic carbon content (OC), lower carbon use, and other problems can be caused by either nutrient excess or deficiency. Accurate diagnosis would benefit farmers on many levels, including yields, fertilizer recommendations, and revenue.

In this paper, we review ML based fertilizer estimation and nutrient status algorithms developed in the last decade. Commonalities in this area are identified, weaknesses are discussed, and possible solutions and future perspectives are proposed. A thorough review of detection and classification approaches was conducted, which served as the basis for a detailed assessment of key challenges that remain to be addressed. This review is organized as follows: Section 2 presents the most common machine-learning algorithms, feature selection approaches, and performance metrics used in the reviewed work. In Section 3, the most common features for nutrient management and fertilizer recommendation studies and the methodology for study selection are presented, along with the ML-based algorithms used by each approach. Finally, Section 4 discusses the advantages and disadvantages of using ML in nutrient management.

2. Machine learning

Machine learning is a branch of artificial intelligence in which the computer, referred to as a machine, learns to perform various tasks automatically Venkataraju et al. (2023). ML combines mathematical modeling and complex algorithms to perform tasks by learning from existing data. ML has been successfully applied in many fields that require classification, prediction, and recommendations Abioye et al. (2022). For example, it provides farmers with soil quality monitoring tools and personalized recommendations based on experimental and field data. ML uses features extracted from known experimental input data to develop models capable of predicting the desired outcome from new data. Machine learning can be divided into three main categories supervised learning, unsupervised learning and Reinforcement learning depending on the desired outcome El Allali et al. (2021).

In supervised learning, an input is mapped to an output based on a training dataset Venkataraju et al. (2023), as shown in Fig. 1. In the case of regression, the output is a value, while the output of classification is a class. Unlike supervised learning, which is task-oriented, unsupervised learning is data-oriented, meaning that the goal is to identify patterns in data sets based on a set of features Rainville et al. (2014). One of the most common applications of unsupervised learning models is clustering, where data with similar patterns are grouped into clusters.

Recently, Deep Learning (DL) has gained traction in the machine learning world as computational power is no longer a hurdle Ahmad et al. (2022); Durai and Shamili (2022). DL is used in either supervised, unsupervised or semi-supervised environments. Unlike the algorithms of classical ML, which often require pre-processing of data, Deep Learning can bypass this step and is therefore better suited for unstructured data such as images. The data is often used directly, eliminating the human factor and automating the feature extraction step. A key advantage of DL is its ability to analyze huge unlabeled datasets, making it an invaluable tool for Big Data Analytics Najafabadi et al. (2015).

2.1. Machine learning algorithms

With advances in machine learning algorithms, even the most complicated regression and classification problems can now be solved Tesfaye et al. (2021). In addition, many of the algorithms are available on a number of open source platforms designed specifically for ML. Regression is a part of supervised learning that provides a prediction of an output variable as a function of input variables that are usually known and available in these subcategories. There are many algorithms used in almost all fields, e.g., linear regression (LR), least absolute shrinkage and selection operator (lasso), logistic regression, and stepwise regression. Many complex algorithms have seen the light of day to solve much more complex problems, such as multivariate adaptive regression splines.

Classification is another important type of supervised learning that uses models to predict a discrete label instead of a continuous output. Support Vector Machine (SVM) is an algorithm that classifies data instances by constructing a linear separation hyperplane John et al. (2020); Karimi et al. (2006). To improve classification, SVMs use a kernel function to transform the original feature space into a higher-dimensional feature space. Classification, regression, and grouping have all been performed with SVMs. SVMs are useful in a variety of applications because they overcome the difficulties of overfitting that occur in high-dimensional spaces and are based on global optimization. Support Vector Regression, Least Squares Support Vector Machine and Successive Projection Algorithm Support Vector Machine are the most commonly used SVM algorithms.

Decision tree (DT) based algorithms use trees to gradually group the dataset into smaller homogeneous subsets (subpopulations) while creating an associated tree graph. Each branch of the tree graph displays the result of a particular pairwise comparison for a particular attribute, while each internal node represents the comparison itself. After following the path from root to leaf, the leaf nodes show the final judgment or prediction (expressed as a classification rule). The classification and regression trees, the automatic chi-square interaction detector, and the iterative dichotomizer are the most common learning techniques in this category.

The Random Forest algorithm (RF) is also widely used and consists of a sequence of decision trees. The result predictions are combined for better predictive performance. DT can be either a regression algorithm or a classification algorithm and represents many suboptimal solutions. Based on a set of decision rules that follow a tree-like architecture, it can make recommendations based on classified data.

Artificial neural networks (ANNs) are also supervised models commonly used for regression and classification Zupan (1994); Zou et al.

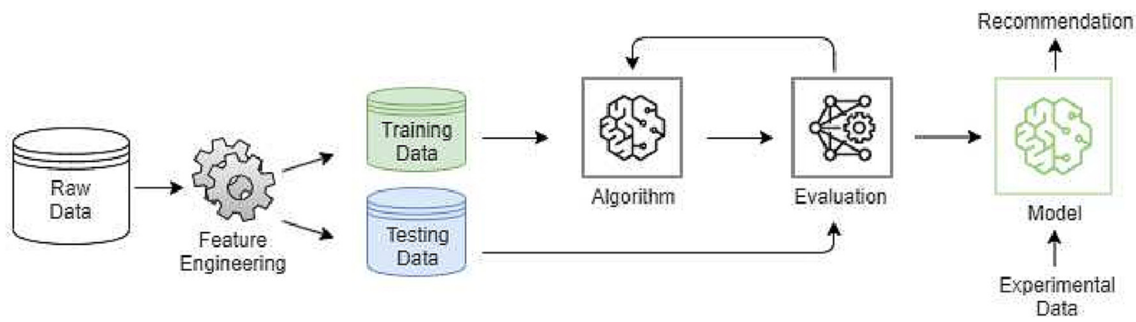

Fig. 1. A general supervised machine learning workflow.

(2008). ANNs use radial basis function networks, perceptron algorithms, and backpropagation to build predictive models Griffel et al. (2022). Deep ANNs, also called deep learning (DL) or deep neural networks (DNNs), are a relatively new branch of machine learning research that enables computer models to represent complex data at multiple levels of abstraction by using numerous processing operations Dharani et al. (2021). The Convolutional Neural Network (CNN) is also a well-known Deep Learning algorithm that was first used to classify images Wu (2017); Kattenborn et al. (2021). CNN creates an artificial neural network that can autonomously learn and make intelligent decisions by automatically extracting the most appropriate features from input sequences and layering techniques Al-Ajlan and El Allali (2018a); Farooque et al. (2023). One of the most attractive features of Deep Learning is that feature extraction is often performed by the model itself. DL Models have greatly improved prediction in several fields, including agriculture Subeesh et al. (2022); Raouhi et al. (2022); Bedi and Gole (2021); Paymode and Malode (2022).

2.2. Feature selection

The first step in any ML experiment is to extract impactful features from the raw data Chandrashekar and Sahin (2014); Dash and Liu (1997). The goal of feature extraction is to find the most informative group of features (unique patterns) to improve the effectiveness of the classifier. Feature extraction, also called feature engineering, extracts features from the original data to achieve accurate classification. Feature extraction is an important step in ML as classification performance can decrease if features are not carefully selected Kebonye et al. (2022).

Feature selection, commonly known as dimensionality reduction (DR), is a technique used in supervised and unsupervised learning to construct a reduced dimensional representation of a dataset while maintaining as much discriminative information as possible. To prevent the implications of excessive dimensionality, it is frequently used before building classification or regression models. There are many feature selection methods in supervised learning, classified into three categories, as shown in Fig. 2. The most commonly used DR techniques are principal component analysis (PCA), Partial Least Squares Regression (PLS), and linear discriminate analysis (LDA). However, algorithms based on metaheuristic approaches such as genetic algorithms (GA) are gaining ground in the field of feature selection.

The goal of feature selection is to extract all useful information from the data. The features are then used in the training phase to build the ML model. Feature selection chooses the smallest possible subset of features from the original set of features to maximize generalizability. Given many independent variables, the feature selection function selects a subset of variables on which to focus a learning algorithm. The model-based approach considers the correlation structure among predictors and assigns a value to each feature indicating how useful or important it is for model construction. The performance of the model would not be affected if features of little or no importance were eliminated. To reduce dimension and select the most acceptable features for classification, dimension reduction can be paired with a feature extraction algorithm Al-Ajlan and El Allali (2018b). On the other hand, Deep Learning uses an alternative architecture that incorporates the process of feature extraction from the data into the learning phase LeCun et al. (2015); Al-Ajlan and El Allali (2018a). This crucial step has a significant impact on the results, as it can improve the performance of the model by removing unnecessary and redundant features. The model becomes less complicated and the likelihood of bias and error is decreased by reducing overfitting. The model gives more realistic recommendations by selecting the most relevant features.

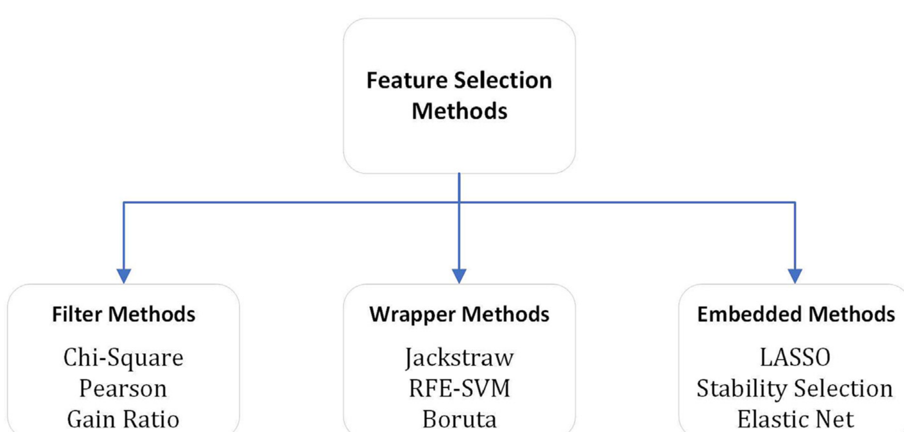

Fig. 2. Different approaches to feature selection with examples for each type.

2.3. Performance metrics

In order to measure the performance of machine learning algorithms, several performance metrics such as Accuracy, Recall, Precision F-Measure, and the area-under-the-curve (AUC) are commonly used as performance metrics. In this section, we describe the metrics that have been used in most studies included in this review such as the coefficient of determination R^2, the mean absolute error (MAE), and the root mean square error.

R^2, or the coefficient of determination, evaluates the proportion of the variance of the target variable that is explained by the model, as shown in Eq. (1):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(yi - \hat{y}i)^2}{\sum_{i=1}^{n}(yi - \overline{y}i)^2} \quad (1)$$

Yi is the observed value of the target variable, $\hat{y}i$ is the predicted value of the target variable, and $\overline{y}i$ represents the mean of the observed target variables. The highest possible value for R^2 is one, which means that the model is 100% accurate. However, this value can also be negative if the model produces slightly worse predictions. In the case of $R^2 = 0$, the model is constant and always predicts the expected value of y, ignoring the input feature. Most often, values above 0.5 are considered good.

MAE or Mean Absolute Error is the average of the absolute difference between the observation and the predicted value, as shown in Eq. (2):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|yi - \hat{y}i| \quad (2)$$

The root-mean-square error (*RMSE*), as shown in Eq. (3), assigns a high weight to large errors due to squaring. When the value is equal to 0, the model is considered to have the best fit. *MAE* is less sensitive to large errors than R^2 and *RMSE* because equal weight is given to each error.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi - \hat{y}i)^2} \quad (3)$$

3. ML based algorithms for nutrient management and fertilizer recommendation

This section describes various machine learning algorithms used in the area of nutrient management and fertilizer recommendations. This review focuses primarily on the current state-of-the-art ML techniques for nutrient management and fertilizer recommendation. It addresses the many factors that influence yield and how machine learning could help predict fertilizer recommendations. Table 1 and Table 2 show a selection of articles from the literature on this topic. An extensive search was conducted to select relevant studies that use machine learning for nutritional management of cropping systems. In the first step, keywords such as "nutrient management", "machine learning in agriculture", "fertilizer forecasting" were used in various combinations including "machine learning in nutrient management", "fertilizer recommendation tools with machine learning" for a broad search through Science Direct, Pubmed, and Scopus. At a first level, publications were classified into two general categories: Nitrogen Management and Nitrogen Phosphorous Potassium (NPK) Management. Only articles published in peer-reviewed journals were selected. Although climate prediction is critical to agricultural productivity, it is not covered in this review because machine learning applications to climate prediction are part of a separate area of study. Finally, all articles covered here were selected for the period from 2010 to July 2022 and found to be all within the scope of this work.

3.1. Most common features

The best nutrient management strategies are predicted based on a set of characteristics Khanal et al. (2018); Jian et al. (2020). For example, 17 nutrients are necessary for optimal crop development, including nitrogen, phosphorus, and potassium. They are all equally important to crops, although their proportions are constantly changing. Because of these variations, the availability of nutrients in the soil affects plant growth, as the soil is the most important source of these nutrients for plants. If any of these nutrients are lacking, crop yield is reduced. In addition to the nutrients already mentioned, the following section discusses some other common features that can be divided into three categories:

Climate is one of the most misunderstood factors in crop development Crane-Droesch (2018); Cai et al. (2019); Newlands et al. (2014). Although it may not appear so at first glance, water availability has a significant impact on agricultural productivity. Even low levels of rainfall can have a detrimental effect on crops, and production can fluctuate significantly due to extreme variations in rainfall amounts and periods. On the other hand, weather conditions are more complicated than just a rainy or dry climate. There are a number of climate features to consider. Total daily and annual precipitation, high-yield and well-distributed precipitation, and the Shannon diversity index (SDI) for precipitation are some of the most common. Other characteristics include daily maximum and minimum temperatures and total solar radiation, pest infestation under various atmospheric conditions and weather patterns.

Algorithms for nutrient management include soil type and texture, organic matter, pH, accessible and exchangeable nutrients, total inorganic and organic carbon, and the total capacity of a soil to hold exchangeable cations (CEC). Qin et al. (2018) created two additional characteristics: a water table-adjusted available water capacity (AWCwt) and a ratio of cumulative seasonal precipitation to AWCwt (RAWCwt). Both features are used to reflect field-level hydrologic conditions and are dependent on conditions that are meant to showcase how soil hydrology impacts nitrogen dynamics. Taking into consideration that the amount of nitrogen lost under wet conditions is considered significantly negative for a soil's ability to store water beyond the saturated zone, which is generally equal to the depth of the water table.

Several soil management variables are considered and used as harvest management features: Planting date and density, date of nitrogen application (both at planting and side-dress application), grain yield, harvest key, fresh and dry biomass, fertilizer rate, and nutrient uptake Ahmed et al. (2021). The information contained in the red and near-infrared (NIR) canopy radiances is used mainly in common vegetation indices: Ratio Vegetation Index (RVI)[4] and Normalized Difference Vegetation Index (NDVI)[5].

$$RVI = \frac{\rho RED}{\rho NIR} \quad (4)$$

$$NDVI = \frac{\rho NIR - \rho RED}{\rho NIR + \rho RED} \quad (5)$$

Where ρRED and ρNIR represent the spectral radiances in the red and NIR regions, respectively. These indices increase the contrast between the ground and vegetation while reducing illumination effects.

3.2. Nitrogen management based systems

In recent years, many applications of nutrient management models have surfaced, especially regarding nitrogen recommendations, as it is considered very important for yield estimation Puntel et al. (2016); Dai et al. (2013); Castaldi et al. (2016); Shi et al. (2021). In the present time, the implementation of ML techniques for N management can be divided into different approaches such as EONR and NNI determination, remote sensing and spectroscopy.

Table 1
Summary of machine learning tools for nitrogen management:

AUTHOR	ALGORITHM	CROP	DATASET	FEATURES	OUTPUT	VALIDATION	RESULTS
Kou et al. (2022)	LR/SVM CNN	Cotton	RGB images	25 visible-band vegetation indexes The HSI color model, Image texture Mean standard deviation Angular second moment, Entropy Contrast and inverse differential moment under gray level co-occurrence matrices Hue, Saturation and Intensity.	Leaf Nitrogen Content	Idependent testing	$R2 = 0.80$ $RMSE = 1.67$ g per Kg
Wang et al. (2021a)	RTM-RF LAI PLSR	Maize	Hyperspectral images	Soil brightness BSM model parameters latitude BSM model parameters longitude Volumetric Soil Moisture Content Leaf chlorophyll content, Leaf carotenoid content Leaf anthocyanin content, Leaf dry matter per area Equivalent Leaf Water Thickness, Senescent Material Mesophyll structure parameter, Canopy leaf area index Averaged leaf slope, Distribution bimodality.	Leaf nitrogen, Chlorophyll, Vmax, Leaf Area Index, Harvested grain yield	Independent testing	Canopy chlorophyll ($R2 = 0.80$) Nitrogen ($R2 = 0.85$) Vmax,27 ($R2 = 0.82$) Canopy chlorophyll ($R2 = 0.80$) Nitrogen ($R2 = 0.85$) Vmax,27 ($R2 = 0.82$)
Qiu et al. (2021)	RF	Rice	RGB images	Fertilizer treatments Vegetation indexes.	NNI AVAILABLE N	–	$R2 = 0.88–0.97$ $RMSE = 0.03–0.07$
Correndo et al. (2021)	XGBoost	Maize	maize observations	Previous crop, tillage system, irrigation, SOM, clay, sand, silt, Precipitations, Mean Temperature, SDI Extreme PP Events, Vapor Pressure Deficit Incident radiation, Photothermal quotient, Extreme T Events.	Feature contribution: Yield without N fertilizer B0 (YEONR), EONR N fertilizer efficiency at the EONR	Folds CV	RMSE medians of 1.90 Mg per ha for B0 ($RRMSE = 24\%$), 1.68 Mg per ha for YEONR ($RRMSE = 14\%$), 52 kg N per ha for EONR ($RRMSE = 34\%$), 10.2 kg yield kg fertilizer N for NFE ($RRMSE = 40\%$). R2 ranged from 0.08 to 0.71
Barbosa et al. (2020b)	CNN	Maize	9 on-farm experiments	N and seed rates prescription maps, Elevation map, Soil's shallow electroconductivity, Soil EC measurements, Texture, Bulk density, Soil organic C, water content, Salinity, and CEC.	Yield responce to crop management	5 Folds Cv	$RMSE = 6,8$
Wang et al. (2021b)	RF	Maize	12 site-year data	Site Soil type, Soil texture, pH, SOM, Total N, Available P, Exchangeable K, Planting density, Base N rate Side-dress N rate.	In-season EONR (NNI) and grain yield	–	$R2 = 0,49$ $RMSE = 17$ kg/ha
Tavakoli and Gebbers (2019)	PLSR RF	Wheat	3 years field digital images and sensor data	Plant nitrogen content Fresh and dry biomasses Water content.	Fresh biomass, Dry biomass, Water content, N content, Crop yield	10 Folds CV	$RMSE = 0.24$ $R2 = 0.76$
Escalante et al. (2019)	Adaboost	Barley	72 aerial RGB Pictures	Barley variety, Fertilizer rate, Vegetation indexes.	N fertilization Crop yield	LOLO	$Acc = 83.3\%$
Ransom et al. (2019)	RF	Maize	49 fields weather and soil data	Texture, Bulk density, pH salt, pH water, CEC, Total N, Total carbon, Inorganic carbon, Organic carbon, Organic Matter, Total precipitation Growing degree days, Corn heat units SDI of precipitation, Abundantly and well-distributed rainfall.	EONR	10 Folds CV * 5	$R2 = 0,84$ $RMSE = 94$ kg/ ha
Qin et al. (2018)	RR	Maize	47 fields weather and soil data	Weather Features Soil Features Management Features (AWCwt) (R AWCwt).	EONR	LOLO LOYO	$R2 = 0,46$ $MAE = 33,6$ kg/ha
Yu et al. (2018)	SAE-FNN	Rapeseed	hyperspectral leaf images	N concentration (%) of leaf The reflectance mean spectra Red-edge of leaf reflectence The overtones of O—H functional groups related to water in leaf The overtones of N—H functional groups related to nitrogen in leaf.	Detection and quantification of leaf (N) concentration	–	$R\ 2 = 0.903$ $RMSEP = 0.307\%$ $RPD = 3.238$
Montes Condori et al. (2017)	CNN	Maize	1152 maize leaf images	Different Growth Stages.	N deficency detection	4 Folds	$Acc = 93.5\%$
Morellos et al. (2016)	CUBIST	–	140 soil samples	Soil spectral results.	MC, OC, TN	LOO	$RMSED = 0.071$ $RPD = 1.96$

Determining the economically optimal amount of nitrogen can be influenced by various factors such as rainfall and management practices. Advances in ML paved the way for better Economic Optimum Nitrogen Rate (EONR) predictions that take these various parameters into account. Qin et al. (2018) looked at the prediction of EONR from field trial data at two different time points: i) seeding, ii) split application time points. Model-derived features such as AWCwt and RAWCwt and weather data were used to build the prediction algorithm. RR provided better results compared to LR, LASSO and Gradient Boosted Regression Trees (GBRT). However, the results of this study cannot be generalized due to the limited data sets. Ransom et al. (2019) evaluated the ability of a set of machine learning and statistical algorithms to improve

Table 2
Summary of machine learning tools for nutrient management.

AUTHOR	ALGORITM	CROP	DATASET	FEATURES	OUTPUT	VALIDATION	RESULTS
Yan et al. (2021)	RF	Maize	7025 field experiment data	Soil type, pH Olsen-P, Organic Matter Ava-N, Ava-K variety planting system	Most important factor affecting YRP and AEP	Independent testing	Maize variety
Timsina et al. (2021)	RF LME	Maize Wheat Rice	600 field experiment data	Cobs m2 (for maize) Straw yield RoI, PFP of N grain N uptake Total N uptake IE and PNB of N PFP of P2O5 Grain P uptake Total P uptake IE and PNB of P PFP of K2O Grain K uptake Total K uptake IE and PNB of K.	Partial nutrient balance Best management practice Yield Prediction	Independent testing	$R2 = 0.99$
Archana and Saranya (2020)	VB Classifier	Many	–	Soil Nutrient level Crop data pH, EC, T.	Yield Prediction Crop recommendation fertilizer recommendation	Independent testing	Acc = 94%
Coulibali et al. (2020)	KNN RFR	Potato	273 field experiments data	Cumulative precipitation Shannon Diversity Index Number of growing degree days Cultivar and Weather data Soil profiles Soil properties and texture.	EONR Crop yield Prediction	70% Training 30% Testing	$R2 = 0{,}490$
Suchithra and Pai (2020)	Gaussian Radial basis	–	Soil testing data	Available (P) Available (K) Organic Carbon (OC) Boron (B), (pH).	pH level Fertilization index	Independent testing	Acc = 80%
Moreno et al. (2018)	ANN	Pasture crops	44 Soil samples	Major soil Nutrients pH CEC OM Salts saturation Amendments.	Fertilizers and Amendments recommmendations	3, 5 k-folds	$R2 = 0.75$
Ghosal et al. (2018)	CNN	Soybean	65,768 RGB Images	Color-based features related to visual stress symptoms.	Potassium deficiency	80% Training 10% Validation 10%Testing	Acc = 94%
Kouadio et al. (2018)	ELM	Robusta coffee	44 field experiments data	Available B,K,P,S Zn Exchangeable Ca, Mg Total Nitrogen OM, pH.	Crop yield	80% Training 20% Testing	RMSE = 0.13 MAE = 7.9%
Culman et al. (2017)	MLPNN	Oil Palm	52 RGB Images Historical data	Visual leaf nutrient deficiency symptoms.	Fertilizer recommendation	60% Training, 10% Model selection 20% Testing	Acc = 87%
Cholissodin et al. (2016)	ANN BISPO	Maize	180 Fertilizer trials	Urea, SP36, KCL Biochar, Dry Weight Weight of 1000 seeds Diameter of the cob Cob hight Harvest Production.	Fertilizer recommmendation	10-Folds 100 iterations	MSE = 8.60
Chen et al. (2014)	SVFS Fisher DA	Rice	RGB Images	Leaf color characteristics Shape characteristics Lightness.	N,P,K nutrient deficiency	–	Acc = 90%
Asraf et al. (2012)	SVM	Oil Palm	420 RGB Images	Color feature extraction Histogram-based texture Mean (R,G,B) Variance(R,G,B) Skewness(R,G,B) Kurtosis(R,G,B) Energy(R,G,B) Entropy (R,G,B) Gray level co-occurrence Matrix(G).	Nutrient deficiency	Independent testing	Acc = 95%
Backhaus et al. (2011)	SVM	Tobacco	60,000 HS images	Vein Epidermis.	Generic nutrient deficiency	5-Folds	Acc = 80%
Yu et al. (2010)	K-means Lagrange multiplier	Maize	Soil Nutrient levels Fertilization rate data	Soil(NPK) concentration NPK fertilizer input.	Crop yield	Independent testing	RMSE = 0.98

nitrogen recommendations for corn using measured soil properties along with weather variables from 49 sites in the Midwestern United States. The authors examined different modeling scenarios for improving three different nitrogen recommendation tools and tested whether adjustments for multicollinearity and accounting for the interaction between soil and weather parameters would result in improvements. Performance was evaluated using the out-of-sample RMSE and included soil and weather data prior to each nitrogen recommendation tool. RF provided better nitrogen recommendations, but compared to the performance of the models in terms of number of variables, the decision tree provided the best fit with the smallest number of variables. The results showed that reducing multicollinearity slightly improved the performance of the different ML algorithms.

The objective of a study by Correndo et al. (2021) was defining the importance of soil, weather conditions, and cropping management in estimating and magnifying uncertainty in key components affecting the nitrogen response of corn yield. When yield without fertilizer, Yield at Economic Optimum N Rate (YEONR), and nitrogen fertilizer efficiency (NFE) are considered in EONR. Bayesian statistics were used for the N response curves fitting along with Extreme Gradient Boosting that evaluates the importance of the traits in the predictability process. EONR was the most difficult attribute to estimate, with an average uncertainty of 50 kg N ha-1. Weather accounted for about two-thirds of the variation in estimated values for YEONR, EONR, and NFE. Uncertainty in all aspects of the N response mechanism was also influenced by weather conditions (72% to 81%). With a constant but moderate contribution, soil factors explained both the predicted N response and the associated uncertainty (10% to 23%). Model uncertainty as a form of risk, potential seasonal weather predictions, and the development of probabilistic frameworks to optimize this data-driven decision process for corn nitrogen application should be considered in improving nitrogen decision support systems.

Wang et al. (2021b) investigated the possibility of improving in-season N-nutrient index (NNI) along with corn grain yield prediction by combining management, soil, and weather data with'GreenSeeker' data, an optical sensor that instantaneously measures crop health and vigor using NDVI. RFR was compared to Lasso linear regression (LLR), which contains similar combined data from multiple sources, and to the simple regression model, which uses only plant data. Studies of corn nitrogen fertilization and crop density were conducted at four sites in northeastern China. Using the RFR model to predict grain yield to simulate the response of corn grain yield response to a range of nitrogen fertilizer applications at different growth stages is an innovative method for seasonal nitrogen fertilizer recommendations that have been developed. The simulated results were as good as the measured results. The RFR model-based recommendation technique, which combines crop sensing data with soil data, is a promising avenue for corn nitrogen management. To improve machine learning-based N recommendation for seasonal seeding, additional data from the year of the site, varying conditions on the farm, and sensors with more spectral ranges are needed. Similarly, data from the RapidSCAN active canopy sensor was also combined with Genetic*Environmental*Management (G*E*M) information to increase the accuracy of NNI estimation in the Midwestern U.S. under a variety of weather, management and soil variables Li et al. (2022);

Qiu et al. (2021) presented that conventional methods for monitoring nitrogen nutrient index (NNI) require manual real field measurement data, which is time-consuming and costly, red-green-blue (RGB) imagery from unmanned aerial vehicles (UAVs) is an alternative. In this study, six machine learning algorithms were used to extract relevant data for predicting the NNI and vegetation index (VI). The results of predicting the NNI of rice using these algorithms showed that the methods of RF performed the best in each growing season, with the best NNI prediction occurring during the filling and early maturity stages. At the early maturity stage, rice NNI was found to be significantly related to both available nitrogen in the soil (AN) and yield. Combining RGB imagery from UAVs with ML algorithms provided a robust and simple solution for instant validation of rice NNI that improved nitrogen use efficiency and provided suggestions for precise fertilization in rice crops despite its limited accuracy.

Modern remote sensing methods for precise nutrient management and fertilization could then be divided into two approaches: Deep learning methods for predictions based on multiple remote sensing data or multivariate regressions Wang et al. (2021a). Barbosa et al. (2020a) used five input variables (nitrogen rate, seed rate, elevation map, soil electrical conductivity, and the NDVI index) to estimate corn yield production using a deep-learning approach, and the results were compared with other machine learning approaches (Fully connected neural network (FNN), multiple linear regression (MLR), SVM, and RF regression models). Although a reduction in RMSE of up to 29% was validated compared to the random forest, it should be noted that the deep learning model, satellite data, required a total of 1800 plots to be defined.This study also shows that the most benefit comes from the higher variability of the spatial structure of the data. In their study, Escalante et al. (2019) investigated the recognition performance of a set of classifiers and deep learning convolutional neural networks for evaluating optimal fertilization using RGB images as input. The average LOO cross-validation performance was presented for each of the studied classifiers and each of the pre-trained models. The percentage of correctly identified images is used to quantify the performance. ANN, SVM and Adaboost performed the best with an accuracy of 81.9%, 80.5% and 83.3%, respectively. Considering that only visual features (Barley variety, Fertilizer rate, Vegetation indexes) extracted from RGB images were used, these are promising results (random guessing leads to an accuracy of 33%). This could be due to the fact that sophisticated models are more specific to the target they were created for, while a simple model can be more general and captures relevant information for the representation of images.

Yu et al. (2018) used a deep learning-based regression model with a fully connected neural network (FNN) and stacked autoencoders (SAE) to quantify nitrogen concentration in canola leaves. SAE was used to infer deep spectral features in the visible and near-infrared regions from a hyperspectral image of a canola leaf, which in turn were used as input data for the FNN to predict N content. With $R^2 = 0.903$, RMSEP = 0.307%, and a residual prediction error RPDp = 3.238 for N concentration, the model SAE-FNN performed quite well. The results show that with a combination of hyperspectral imaging and deep learning, it is possible to quickly and non-destructively detect N concentration in canola leaves to provide better fertilizer recommendations.

In another study, Deep Learning was used to detect nitrogen deficiency in corn. For this purpose, four CNN models were pre-trained by applied transfer learning, considering (V4, V6 and R1) as growth stages. The leaf samples were then digitized to generate 384 images representing the 16 nitrogen treatments considered in the study. The results showed significant differences with respect to each growth stage. However, CNNs built using RGB images provided great results (average 61.8%) compared to traditional text-based methods (average 50–60%) Montes Condori et al. (2017).

Similarly, in the study by Kou et al. (2022), RGB images of the cotton canopy were acquired using a UAV digital camera. The nitrogen content of the cotton canopy was predicted using two cotton cultivars and six nitrogen gradients. Forty-six features were extracted from the image after image preprocessing, and CNN were used to extract deep features. Pearson and partial least squares were used feature selection. For accurate prediction of nitrogen content of cotton crowns, manual features were used as input to build linear regression models, support vector machines, and one-dimensional CNN regression models. Deep learning based features were used as inputs to build a two-dimensional CNN regression model. The results were $R^2 = 0.80$ and RMSE = 1.67 g kg-1 for Xinluzao 45 and $R^2 = 0.42$ and RMSE = 3.13 g kg-1 for Xinluzao 53, indicating that cotton nitrogen content can be predicted on a large scale using UAV RGB images and machine learning. However, the accuracy and stability of the prediction model still need to be improved due to insufficient data samples.

In addition, soil properties such as total nitrogen can be determined by combining ML algorithms and spectroscopy. These algorithms are an excellent alternative to simple regression models when the goal is to increase the accuracy of the regression. Partial least square regression (PLSR) is a widely used approach for processing multivariable inputs, extracting successive linear combinations of the spectra to achieve the coupled goals of effectively explaining response variation and optimally explaining predictor variation. This is the most commonly used method in soil spectroscopy and chemometrics and has proven successful in assessing water and nitrogen status. The availability of large data sets coupled with better training and the provision of highly accurate models is one of the major challenges for this system when applied to nutritional management. Predictive models have been developed using several technologies, including PLSR, which reduces the number of features required, and neural networks, which compensate for non-linearity Wang et al. (2021a); Tavakoli and Gebbers (2019); Morellos et al. (2016). Table 1 summarizes and includes brief descriptions of the above studies.

3.3. NPK management based systems

The development of NPK management based systems follows two general approaches based the type of input, namely field data (soil tests, fertilization trials, etc.) or RGB and hyperspectral images. Yu et al. (2010) developed a fertilizer model with data points from 10 experimental fields with 4 fertilizer rates and 14 treatments. A neural network ensemble was presented to calculate the fertilizer rate more accurately. The authors used K-means clustering to select the best networks individually and then combined the models using a Lagrange multiplier. A fertilizer model was created using the ensemble method for neural networks described above. In this model, soil nutrient content and fertilizer rate are used as inputs to the neural network, while yield is considered as an output. With this approach, the calculation of fertilizer rates becomes a programming problem and can be used to determine the fertilizer rate with the highest yield and profit and to predict the yield. This fertilizer model was also validated using data on the effect of fertilizers. The results show that using an ensemble of neural networks to predict yield is more accurate than using individual neural networks.

According to Cholissodin et al. (2016), the integrated artificial neural network (ANN) can not only optimize fertilizer rates for corn, but also when used in conjunction with Bidirectional Improved Particle Swarm

Optimization (BIPSO), the fusion provides better results. The ANN approach provided good results in predicting recommendations using field trial data as training data, while BIPSO optimized multiple features simultaneously, which accelerated the operation of the system. The smallest mean square error (MSE) value was obtained by 10-fold cross-validation, repeating CV hundreds of times. The goal of the study was to initially use ML in pasture and rotational cropping and then extend it to other crops based on the results, which could lead to new insights when used as input by other researchers. Kouadio et al. (2018) investigated the use of Extreme Learning Machines (ELM) to analyze soil fertility parameters and produce an effective yield estimate. The effectiveness of different ELM-based models based on Soil organic matter (SOM), accessible exchangeable nutrients, and pH was tested with single and multiple combinations of predictor variables. The results of the ELM model were compared with those of existing predictive techniques such as MLR and RF. The authors believe that the ELM model makes a unique contribution to the agricultural sector, particularly in terms of selecting optimal soil parameters for predicting coffee yields. The research demonstrates the potential value of combining machine learning with biophysical crop models. Moreno et al. (2018) found that they could determine the fertilizers and amendments needed for pasture production based on the major nutrients in the soil. They used a multilayer perceptron network trained with the backpropagation algorithm. For fertilizers and additives, a multilayer artificial neural network was trained with an input layer consisting of soil test variables and an output layer with many simultaneous outputs. The MSE of the test data and its standard deviation, and the MSE of the training data and its standard deviation were used to determine the quality of each neural network. The objective of the study was to develop a recommendation system suitable for pasture cropping despite the limited amount of data.

Suchithra and Pai (2020) developed a system based on soil test results to classify and predict soil fertility indices and pH values based on various soil characteristics. The algorithm used in this study was an Extreme Learning Machine (ELM) for classification and prediction. This algorithm provides better-generalized results by feeding forward NN with a single hidden layer (SLFNs). The potassium fertility index had an accuracy of up to 78%, while the pH classification was up to 89%. The activation function in this NN that had the best classification, accuracy, and kappa values was the GRB function. Similarly, Timsina et al. (2021) also compared three different nutrient management strategies for common cereals considering various factors such as nutrient use efficiency (NUE). A developed site-specific decision support system Nutrient Expert (NE) was used to evaluate the nutrient balance of the field by analyzing past yields or costs. RF algorithm showed that NUE for rice and P and K uptake for wheat and maize were the most important factors contributing to grain yield. Random Forest was used in the study by Yan et al. (2021) to quantify the relative importance of different traits (Phosphorous (P) fertilization practices, region, soil properties, variety, and cropping system) on yield response to P and agronomic efficiency of P in maize. Although this algorithm produced very good results, Maize variety was selected as the most important factor affecting yield. The study encountered several limitations that should be considered in future studies, such as the differences in yield and agronomic efficiency between regions, the different release of fertilizer-P in soil dynamics, and the fact that only top soil samples were used rather than the entire soil profile.

Coulibali et al. (2020) investigated and compared machine learning and probabilistic models with site-specific predictive models for fertilization of potato crops in eastern Canada. Using statistical models, potato (*Solanum tuberosum* L.) performance is often associated with fertilizer requirements. Because of the many variables involved, such as soil, land management, genotypes, weather, and pests and diseases, it is difficult to predict appropriate nutrient levels. This study compared different models for evaluating NPK requirements for high quality and yield as a function of soil, land management, and weather. They used data from 273 field trials conducted in Quebec between 1979 and 2017. Predictions from k-nearest neighbors, RF, NN, hierarchical Mitscherlich model, and Gaussian processes were created, tested, and compared. For the prediction of marketable tuber yield, the ML models gave R^2 values of 0.49–0.59, which were higher than the R^2 of the Mitscherlich model (0.37). Some models did not agree in obtaining optimal rates from dose-response surfaces under constant conditions. Gaussian processes proved to be a promising method for site-specific fertilizer recommendations that can reduce economic or agronomic risks because of their ability to make suggestions in the context of probabilistic risk assessment.

RGB or hyperspectral images can be used to detect nutrient deficiencies at early growth stages non-destructively. In their study, Backhaus et al. (2011) investigated whether supervised approaches for predicting plant nutritional status using classification models are robust enough to handle large data sets with significant variance, such as leaf age or pixel position in the leaf. The learning algorithms tested were SVM, Generalized Relevance Learning Vector Quantization (GRLVQ), Supervised Relevance Neural Gas (SRNG), and a Radial Basis Function (RBF) Network. Leaf growth stage had the greatest impact on classification accuracy, with SVM and RBF providing reliable results and SRNG and GRLVQ techniques falling to nearly determinable values. The importance of spectral bands in predicting nutrient content was estimated using three cameras covering the visible and shortwave infrared spectrum (VIS /SWIR). The simple separation of pixels of veins and epidermis proved to be a source of confounding variance for nutrient categorization in this work, but had little effect on actual classification performance. Leaf age had a much stronger effect on GRLVQ, SRNG, and SVM classification performance (simple spectra) for nutrient states, such that it was close to the rate. Automatic detection of nutrient deficiencies in leaves of oil palm using a visual system and pattern recognition was presented by Asraf et al. (2012). In this study, Support Vector Machine (SVM) is also used as a classifier using three different kernels: a linear kernel, hard edge polynomial kernel and a soft edge polynomial kernel. According to the preliminary data, the SVM classifiers were able to detect oil palm leaves. The soft-edge polynomial kernel was able to accurately classify nutritional disorders in oil palm leaves with 95% correctness. Support vector feature selection (SVFS), as a variant of SVM, showed great potential in selecting relevant features for nutritional deficiencies. Images of the top three leaves of a rice plant (*Oryza sativa* L.) and associated leaf sheaths were acquired using static scanning techniques. Thirty-two spectral and shape features were identified from these images by fusion of an RGB mean function and a Matlab region-prop function. NPK deficiencies were effectively detected using hierarchical identification. The overall accuracy of NPK deficiencies for the four growth stages was 86.15, 87.69, 90.00, and 89.23%, respectively. Validation was performed with data from different years, and the accuracies were 83.08, 83.08, 89.23, and 90.77%, respectively Chen et al. (2014).

In the study by Ghosal et al. (2018) individual soybean leaves that exhibited a range of deficiency symptoms such as potassium and iron were manually selected and collected in the field by destructive sampling. The leaves and charts were manually recorded with a digital camera. In this way, 25,000 images were collected and labeled to create a dataset of leaf images. The authors used a CNN classifier and were able to achieve an accuracy of 94%. In this study, a deep machine vision based approach was used to detect early symptoms of stress. The presented method is widely applicable in modern agriculture and provides accurate and not time-consuming immediate stress detection. This method has been shown to be relatively insensitive to illumination variations, making it a simple technique for large-scale stress detection. Another deep-learning approach was developed by Culman et al. (2017), who presented PalmHand, a unique smart-device application that allows farmers to detect instant oil palm deficiencies using leaf images. The developed app works as an IoT device that stores and analyzes historical data collected from numerous users simultaneously in the cloud. A single MLP was used to classify oil palms into one of four possible categories. The MLP classified the image into four categories corresponding to a healthy palm or a specimen deficient in potassium (K), magnesium (Mg), or nitrogen (N). The average macroprecision and accuracy were 0.67 and 0.50, respectively. Considering the limited dataset available for training the classifier, these results were considered promising evidence for the concept. However, no information was provided on the circumstances under which the 52 RGB images were acquired. Similarly no details concerning the data were shared in the work of Archana and Saranya (2020) that proposed a method focusing on macronutrients (NPK), soil pH and electrical conductivity, and temperature to provide the best crop recommendations. Crop rotation, fertilizer recommendations, forecasting, and crop production prediction were part of the collaborative proposed system. The objective of this study was to develop a system that combines an agricultural dataset and uses a voting-based ensemble as a classification algorithm to recommend suitable crops. It was found that yield prediction could be of great help to farmers in increasing their yields. Crop rotation has shown good results in improving soil fertility when applied regularly. This technique helps in making fertilizer decisions that are beneficial to farmers. This system had an accuracy of 92%.

Although machine learning is widely used in agriculture in general and in nutrient management and fertilizer recommendations in particular, the literature shows that research needs to focus on the availability of data, including all the important factors that need to be used to overcome the limitations Ransom et al. (2019); Wang et al. (2021b); Yu et al. (2010); Cholissodin et al. (2016). Table 2 summarizes the above work for the case of nutrient management subcategory.

4. Summary and outlook

The development of a nutrient management tool using machine learning generated a great deal of interest in the scientific community, which recognized the importance of food security and the sustainability of its processes Thompson et al. (2015). With the revolution that our modern agriculture is experiencing today, the use of various sensors, smart irrigation systems, and remote sensing, more and more datasets are becoming available Chlingaryan et al. (2018). These large datasets have emerged simultaneously with major advances in computer science such as machine learning, which has led to incredible breakthroughs in this research area. In this paper, we collected and reviewed various machine learning approaches in the field of nutrient management. Analysis of these articles revealed that in the nitrogen management category, Random Forests was the most frequently used ML algorithm. The popularity of RFs correlated with better results (R^2 was up to 0.97 Qiu et al. (2021)). In the NPK category, many models such as SVMs were used, but an RF-based algorithm gave the best R^2 score Timsina et al. (2021).

In the Nitrogen category, we can see a high presence of image inputs, reflecting the emerging trend and the wide use of multi-source images (HS, RGB, NIR, etc.). This has been enabled by the development and affordability of this technology along with high computational power. By integrating machine learning with IoT data, farm management systems are evolving into true artificial intelligence systems that provide better recommendations and insights for future decisions and actions, with the goal of increasing productivity. To this end, it is envisioned that the use of machine learning models will become even more popular over time, enabling the development of integrated and applicable technologies. Despite the great strides that have been made recently in the field of nutrient management, actual practical/accurate application is currently a challenge. Many limitations have come to light that are related to the methods used and are both intrinsically and extrinsically related to the target problem. The current state of nutrient management using digital imagery and machine learning was also discussed to provide an in-depth analysis of the various challenges faced by many researchers working on this topic. Collecting datasets with such precise factors and for a specific crop is usually not an easy task, considering that not all available datasets are representative enough of the variation found in the field.

To determine suitability for practical application, basic data from studies of crop response to nutrients in fields with cross-validation are needed. Soil properties John et al. (2020), plant density, weather indices, plant variety, plant growth stage and N, P, and K fertilizer rates, are all very important and should be considered in data collection because they can directly affect the model output Trontelj ml and Chambers (2021). Yield response can vary by plant variety in addition to the fertilizer rate proposed by recommendation systems. In addition, field trials would directly impact yield and improve not only biomass, but also solution sustainability and soil health. This practice would shed light on forgotten features such as SOM and tillage systems, leading to a better understanding of soil dynamics and better recommendations. For example, Ransom et al. (2019) and Qin et al. (2018) tested several models and algorithms to improve nitrogen requirement for corn. In the case of Ransom et al. (2019), the R^2 was up to 0.94, with data sets collected from 49 experimental sites in the Midwestern United States. While R^2 in Qin et al. (2018) was only 0.46, using data from a limited number of sites. Both studies tested the effect of ML algorithms on nutrient management but reached different conclusions due to the quality and size of the data they used.

The most critical point to mention is that many people see machine learning as a solution to future needs in agriculture. On the other hand, machine learning only provides the best recommendations based on the inputs. In practice, various soil parameters need to be considered to achieve greater or higher yields and better environmental performance than traditional fertilizer management systems. However, to fully understand this system, a standardized research technique based on site-specific parameters is required. We believe that a comprehensive knowledge of what accuracy means is insufficient. As a result, different technologies, procedures, and criteria for accuracy and good results have been used, leading to a scenario of conflicting results. We also find that all approaches look at single procedures and solutions and are not sufficiently linked to the decision process as is the case in other application areas. The real problem is understanding the right path to the best nutrient management recommendations and making them accessible and understandable to farmers. Researchers should also focus on the transition between great modeling results and tangible practices to achieve them. From most of the articles reviewed, we can conclude that there were strong correlations between model accuracy and dataset size and quality. Compiling comprehensive data sets that account for all the different scenarios and variabilities is the most difficult task Barbedo (2019), but is considered key to accurate results. The first step would be to identify the data source, which could include field studies and research. Second, we need to collect data on soil type and climate from research studies and APIs. We also need data on crop type, crop yield, and management data that can be obtained from farmers, extension services, and government agencies. Finally, the collected data needs to be cleaned and preprocessed. This includes removing outliers and normalizing the data to combine them into a single data set that can be used by the ML algorithm.

However, we believe that data availability is the bottleneck when it comes to taking full advantage of what ML can offer to the field of fertilizer recommendations. We encourage researchers to share their data with the community to enhance the capabilities and impact of the research. We also encourage researchers developing ML-based recommendation systems to suggest the best suitable features and develop other features that can be collected at a large scale. In addition, it is crucial to standardize the methodological approach to nutrient management systems based on a detailed description of the entire system so that results from different experiments and parts of the world can be logically compared.

The accessibility of powerful machine learning tools and techniques on many open source platforms is certainly beneficial to the research community. However, this also leads to many misleading publications by researchers who either do not know exactly what ML expertise is required or do not have the proper soil and plant nutrient knowledge. This phenomenon has led to a great pipe dream Schut and Giller (2020); Lischeid et al. (2022); Paudel et al. (2021) in which scientists confirm the accuracy of their models while considering the wrong parameters.

Given the rapid progress ML has made in recent decades, it is very likely that the world will see an increase in the applications of ML models. The integration and applications are becoming more and more remarkable. ML algorithms show great potential when properly used for support or decision-making. The fusion of spectral and spatial features and unique hybrid processing systems that complement each other's limitations would definitely improve the understanding of soil dynamics. This success will likely be paired with better yields leading to a more food-secure world.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abioye, E.A., Hensel, O., Esau, T.J., Elijah, O., Abidin, M.S.Z., Ayobami, A.S., Yerima, O., Nasirahmadi, A., 2022. Precision irrigation management using machine learning and digital farming solutions. AgriEngineering 4, 70–103.

Adnan, A.A., Diels, J., Jibrin, J.M., Kamara, A.Y., Craufurd, P., Shaibu, A.S., Mohammed, I.B., Tonnang, Z.E., 2019. Options for calibrating CERES-maize genotype specific parameters under data-scarce environments. PLoS One 14, 1–20.

Ahmad, A., Saraswat, D., El Gamal, A., 2022. A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools. Smart Agricult. Technol. 100083.

Ahmed, U., Lin, J.C.W., Srivastava, G., Djenouri, Y., 2021. A nutrient recommendation system for soil fertilization based on evolutionary computation. Comput. Electron. Agric. 189, 106407.

Al-Ajlan, A., El Allali, A., 2018a. CNN-MGP: convolutional neural networks for metagenomics gene prediction. Interdisciplin. Sci.: Comput. Life Sci. 11, 628–635.

Al-Ajlan, A., El Allali, A., 2018b. Feature selection for gene prediction in metagenomic fragments. BioData Minin. 11.

Archana, K., Saranya, K., 2020. Crop yield prediction, forecasting and fertilizer recommendation using voting based ensemble classifier. SSRG Int. J. Comput. Sci. Eng 7, 1–4.

Asraf, H.M., Nooritawati, M., Rizam, M.S., 2012. A comparative study in kernel-based support vector machine of oil palm leaves nutrient disease. Proced. Eng. 41, 1353–1359.

Aworka, R., Cedric, L.S., Adoni, W.Y.H., Zoueu, J.T., Mutombo, F.K., Kimpolo, C.L.M., Nahhal, T., Krichen, M., 2022. Agricultural decision system based on advanced machine learning models for yield prediction: case of east african countries. Smart Agricult. Technol. 2, 100048.

Ayodele, T.O., 2010. Types of machine learning algorithms. New Adv. Mach. Learn. 3, 19–48.

Babaie Sarijaloo, F., Porta, M., Taslimi, B., Pardalos, P.M., 2021. Yield performance estimation of corn hybrids using machine learning algorithms. Artif. Intell. Agricult. 5, 82–89.

Backhaus, A., Bollenbeck, F., Seiffert, U., 2011. Robust classification of the nutrition state in crop plants by hyperspectral imaging and artificial neural networks. 2011 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS). IEEE, pp. 1–4.

Barbedo, J.G.A., 2019. Detection of nutrition deficiencies in plants using proximal images and machine learning: a review. Comput. Electron. Agric. 162, 482–492.

Barbosa, A., Trevisan, R., Hovakimyan, N., Martin, N.F., 2020a. Modeling yield response to crop management using convolutional neural networks. Comput. Electron. Agric. 170, 105197.

Barbosa, A., Trevisan, R., Hovakimyan, N., Martin, N.F., 2020b. Modeling yield response to crop management using convolutional neural networks. Comput. Electron. Agric. 170, 105197.

Bedi, P., Gole, P., 2021. Plant disease detection using hybrid model based on convolutional autoencoder and convolutional neural network. Artif. Intell. Agricult. 5, 90–101.

Bharath, S.M., Manoj, S., Adhappa, P., Patagar, P.L., Bhaskar, R., 2022. Crop yield prediction with efficient use of fertilizers. Lectu. Notes Elect. Eng. 783, 937–943.

Cai, Y., Guan, K., Lobell, D., Potgieter, A.B., Wang, S., Peng, J., Xu, T., Asseng, S., Zhang, Y., You, L., Peng, B., 2019. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. Agric. For. Meteorol. 274, 144–159.

Castaldi, F., Castrignanò, A., Casa, R., 2016. A data fusion and spatial data analysis approach for the estimation of wheat grain nitrogen uptake from satellite data. Int. J. Remote Sens. 37, 4317–4336.

Chandrashekar, G., Sahin, F., 2014. A survey on feature selection methods. Comput. Electr. Eng. 40, 16–28.

Chen, L., Lin, L., Cai, G., Sun, Y., Huang, T., Wang, K., Deng, J., 2014. Identification of nitrogen, phosphorus, and potassium deficiencies in Rice based on static scanning technology and hierarchical identification method. PLoS One 9, e113200.

Chivenge, P., Zingore, S., Ezui, K., Njoroge, S., Bunquin, M., Dobermann, A., Saito, K., 2022. Progress in research on site-specific nutrient management for smallholder farmers in sub-saharan africa. Field Crop Res. 281, 108503.

Chlingaryan, A., Sukkarieh, S., Whelan, B., 2018. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review. Comput. Electron. Agric. 151, 61–69.

Cholissodin, I., Dewi, C., Surbakti, Eunike Endariahna, 2016. Integrated ANN and Bidirectional improved PSO for optimization of fertilizer dose on Palawija plants. 2016 2nd International Conference on Science in Information Technology (ICSITech). IEEE, pp. 193–197.

Collier, P., Dercon, S., 2014. African Agriculture in 50years: Smallholders in a Rapidly Changing World? World Development. 63. Economic Transformation in Africa, pp. 92–101.

Correndo, A.A., Tremblay, N., Coulter, J.A., Ruiz-Diaz, D., Franzen, D., Nafziger, E., Prasad, V., Rosso, L.H.M., Steinke, K., Du, J., et al., 2021. Unraveling uncertainty drivers of the maize yield response to nitrogen: a bayesian and machine learning approach. Agric. For. Meteorol. 311, 108668.

Coulibali, Z., Cambouris, A.N., Parent, S.É., 2020. Site-specific machine learning predictive fertilization models for potato crops in eastern Canada. PLoS One 15, e0230888.

Crane-Droesch, A., 2018. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. Environ. Res. Lett. 13.

Culman, M.A., Gomez, J.A., Talavera, J., Quiroz, L.A., Tobon, L.E., Aranda, J.M., Garreta, L.E., Bayona, C.J., 2017. A novel application for identification of nutrient deficiencies in oil palm using the internet of things. 2017 5th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud). IEEE, pp. 169–172.

Dai, X., Zhou, X., Jia, D., Xiao, L., Kong, H., He, M., 2013. Managing the seeding rate to improve nitrogen-use efficiency of winter wheat. Field Crop Res. 154, 100–109.

Dash, M., Liu, H., 1997. Feature selection for classification. Intell. Data Anal. 1, 131–156.

Dhal, S.B., Bagavathiannan, M., Braga-Neto, U., Kalafatis, S., 2022. Nutrient optimization for plant growth in aquaponic irrigation using machine learning for small training datasets. Artif. Intell. Agricult. 6, 68–76.

Dharani, M., Thamilselvan, R., Natesan, P., Kalaivaani, P., Santhoshkumar, S., 2021. Review on crop prediction using deep learning techniques. Journal of Physics: Conference Series. IOP Publishing, p. 012026.

Diao, X., Hazell, P., Thurlow, J., 2010. The Role of Agriculture in African Development. World Development 38, 1375–1383. The Future of Small Farms Including Special Section: Impact Assessment of Policy-Oriented International Agricultural Research. pp. 1453–1526.

Durai, S.K.S., Shamili, M.D., 2022. Smart farming using machine learning and deep learning techniques. Decis. Analyt. J. 3, 100041.

El Allali, A., Elhamraoui, Z., Daoud, R., 2021. Machine learning applications in RNA modification sites prediction. Comput. Struct. Biotechnol. J. 19, 5510–5524.

Elavarasan, D., Vincent, D.R., Sharma, V., Zomaya, A.Y., Srinivasan, K., 2018. Forecasting yield by integrating agrarian factors and machine learning models: a survey. Comput. Electron. Agric. 155, 257–282.

Escalante, H.J., Rodríguez-Sánchez, S., Jiménez-Lizárraga, M., Morales-Reyes, A., De La Calleja, J., Vazquez, R., 2019. Barley yield and fertilization analysis from UAV imagery: a deep learning approach. Int. J. Remote Sens. 40, 2493–2516.

Fairhurst, T., Witt, C., Buresh, R., Dobermann, A., Fairhurst, T., 2007. Rice: A practical guide to nutrient management. Int. Rice Res. 1, 2–4.

Farooque, A.A., Hussain, N., Schumann, A.W., Abbas, F., Afzaal, H., McKenzie-Gopsill, A., Esau, T., Zaman, Q., Wang, X., 2023. Field evaluation of a deep learning-based smart variable-rate sprayer for targeted application of agrochemicals. Smart Agricult. Technol. 3, 100073.

Ghosal, S., Blystone, D., Singh, A.K., Ganapathysubramanian, B., Singh, A., Sarkar, S., 2018. An explainable deep machine vision framework for plant stress phenotyping. Proc. Natl. Acad. Sci. 115, 4613–4618.

Goulding, K., Jarvis, S., Whitmore, A., 2008. Optimizing nutrient management for farm systems. Philosoph. Trans. Royal Soc. B: Biol. Sci. 363, 667–680.

Griffel, L., Delparte, D., Whitworth, J., Bodily, P., Hartley, D., 2022. Evaluation of artificial neural network performance for classification of potato plants infected with potato virus y using spectral data on multiple varieties and genotypes. Smart Agricult. Technol. 100101.

Holzworth, D.P., Snow, V., Janssen, S., Athanasiadis, I.N., Donatelli, M., Hoogenboom, G., White, J.W., Thorburn, P., 2015. Agricultural production systems modelling and software: current status and future prospects. Environ. Model. Softw. 72, 276–286.

Jayne, T.S., Mather, D., Mghenyi, E., 2010. Principal challenges confronting smallholder agriculture in sub-saharan africa. World Dev. 38, 1384–1398.

Jha, K., Doshi, A., Patel, P., Shah, M., 2019. A comprehensive review on automation in agriculture using artificial intelligence. Artif. Intell. Agricult. 2, 1–12.

Jian, J., Du, X., Stewart, R.D., 2020. A database for global soil health assessment. Sci. Data 7, 3–10.

John, K., Abraham Isong, I., Michael Kebonye, N., Okon Ayito, E., Chapman Agyeman, P., Marcus Afu, S., 2020. Using machine learning algorithms to estimate soil organic carbon variability with environmental variables and soil nutrient indicators in an alluvial soil. Land 9, 487.

Kamilaris, A., Kartakoullis, A., Prenafeta-Boldú, F.X., 2017. A review on the practice of big data analysis in agriculture. Comput. Electron. Agric. 143, 23–37.

Karimi, Y., Prasher, S., Patel, R., Kim, S., 2006. Application of support vector machine technology for weed and nitrogen stress detection in corn. Comput. Electron. Agric. 51, 99–109.

Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. Review on convolutional neural networks (cnn) in vegetation remote sensing. ISPRS J. Photogramm. Remote Sens. 173, 24–49.
Kebonye, N.M., Agyeman, P.C., Biney, J.K., 2022. Optimized modelling of countrywide soil organic carbon levels via an interpretable decision tree. Smart Agricult. Technol. 100106.
Khanal, S., Fulton, J., Klopfenstein, A., Douridas, N., Shearer, S., 2018. Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield. Comput. Electron. Agric. 153, 213–225.
Kou, J., Duan, L., Yin, C., Ma, L., Chen, X., Gao, P., Lv, X., 2022. Predicting leaf nitrogen content in cotton with uav rgb images. Sustainability 14, 9259.
Kouadio, L., Deo, R.C., Byrareddy, V., Adamowski, J.F., Mushtaq, S., Phuong Nguyen, V., 2018. Artificial intelligence approach for the prediction of Robusta coffee yield using soil fertility properties. Comput. Electron. Agric. 155, 324–338.
Kuijpers, R., Swinnen, J., 2016. Value chains and technology transfer to agriculture in developing and emerging economies. Am. J. Agric. Econ. 98, 1403–1418.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.
Li, D., Miao, Y., Ransom, C.J., Bean, G.M., Kitchen, N.R., Fernández, F.G., Sawyer, J.E., Camberato, J.J., Carter, P.R., Ferguson, R.B., et al., 2022. Corn nitrogen nutrition index prediction improved by integrating genetic, environmental, and management factors with active canopy sensing using machine learning. Remote Sens. 14, 394.
Liakos, K.G., Busato, P., Moshou, D., Pearson, S., Bochtis, D., 2018. Machine learning in agriculture: a review. Sensors (Switzerland) 18, 1–29.
Lischeid, G., Webber, H., Sommer, M., Nendel, C., Ewert, F., 2022. Machine learning in crop yield modelling: a powerful tool, but no surrogate for science. Agric. For. Meteorol. 312, 108698.
Livingston, G., Schonberger, S., Delaney, S., 2011. Sub-saharan africa: the state of smallholders in agriculture, in: paper presented at the IFAD conference on new directions for smallholder agriculture. Citeseer. p. 25.
Monaghan, R., Hedley, M., Di, H., McDowell, R., Cameron, K., Ledgard, S., 2007. Nutrient management in New Zealand pastures—recent developments and future issues. N. Z. J. Agric. Res. 50, 181–201.
Montes Condori, R.H., Romualdo, L.M., Martinez Bruno, O., de Cerqueira Luz, P.H., 2017. Comparison between traditional texture methods and deep learning descriptors for detection of nitrogen deficiency in maize crops. 2017 Workshop of Computer Vision (WVC). IEEE, pp. 7–12.
Morellos, A., Pantazi, X.E., Moshou, D., Alexandridis, T., Whetton, R., Tziotzios, G., Wiebensohn, J., Bill, R., Mouazen, A.M., 2016. Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using Vis-nir spectroscopy. Biosyst. Eng. 152, 104–116.
Moreno, R.H., Garcia, O., Arias, R.L.A., 2018. Model of neural networks for fertilizer recommendation and amendments in pasture crops. 2018 ICAI Workshops (ICAIW). IEEE, pp. 1–5.
Najafabadi, M.M., Villanustre, F., Khoshgoftaar, T.M., Seliya, N., Wald, R., Muharemagic, E., 2015. Deep learning applications and challenges in big data analytics. J. Big Data 2, 1.
Newlands, N.K., Zamar, D.S., Kouadio, L.A., Zhang, Y., Chipanshi, A., Potgieter, A., Toure, S., Hill, H.S., 2014. An integrated, probabilistic model for improved seasonal forecasting of agricultural crop yield under environmental uncertainty. Front. Environ. Sci. 2, 1–21.
Olanipekun, I.O., Olasehinde-Williams, G.O., Alao, R.O., 2019. Agriculture and environmental degradation in africa: the role of income. Sci. Total Environ. 692, 60–67.
Patrício, D.I., Rieder, R., 2018. Computer vision and artificial intelligence in precision agriculture for grain crops: a systematic review. Comput. Electron. Agric. 153, 69–81.
Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylianidis, C., Athanasiadis, I.N., 2021. Machine learning for large-scale crop yield forecasting. Agric. Syst. 187, 103016.
Paymode, A.S., Malode, V.B., 2022. Transfer learning for multi-crop leaf disease image classification using convolutional neural network vgg. Artif. Intell. Agricult. 6, 23–33.
Preethi, G., Rathi Priya, V., Sanjula, S.M., Lalitha, S.D., Vijaya Bindhu, B., 2020. Agro based crop and fertilizer recommendation system using machine learning. Eur. J. Mol. Clin. Med. 7, 2043–2051.
Puntel, L.A., Sawyer, J.E., Barker, D.W., Dietzel, R., Poffenbarger, H., Castellano, M.J., Moore, K.J., Thorburn, P., Archontoulis, S.V., 2016. Modeling long-term corn yield response to nitrogen rate and crop rotation. Front. Plant Sci. 7, 1–18.
Qin, Z., Myers, D.B., Ransom, C.J., Kitchen, N.R., Liang, S., Camberato, J.J., Carter, P.R., Ferguson, R.B., Fernandez, F.G., Franzen, D.W., Laboski, C.A., Malone, B.D., Nafziger, E.D., Sawyer, J.E., Shanahan, J.F., 2018. Application of machine learning methodologies for predicting corn economic optimal nitrogen rate. Agron. J. 110, 2596–2607.
Qiu, Z., Ma, F., Li, Z., Xu, X., Ge, H., Du, C., 2021. Estimation of nitrogen nutrition index in rice from uav rgb images coupled with machine learning algorithms. Comput. Electron. Agric. 189, 106421.
Rainville, D., Durand, A., Fortin, F.A., Tanguy, K., Maldague, X., Panneton, B., Simard, M.J., et al., 2014. Bayesian classification and unsupervised learning for isolating weeds in row crops. Pattern. Anal. Applic. 17, 401–414.
Ransom, C.J., Kitchen, N.R., Camberato, J.J., Carter, P.R., Ferguson, R.B., Fernández, F.G., Franzen, D.W., Laboski, C.A., Myers, D.B., Nafziger, E.D., Sawyer, J.E., Shanahan, J.F., 2019. Statistical and machine learning methods evaluated for incorporating soil and weather into corn nitrogen recommendations. Comput. Electron. Agric. 164, 104872.
Raouhi, E.M., Lachgar, M., Hrimech, H., Kartit, A., 2022. Optimization techniques in deep convolutional neuronal networks applied to olive diseases classification. Artif. Intell. Agricult. 6, 77–89.
Rejeb, A., Rejeb, K., Zailani, S., Keogh, J.G., Appolloni, A., 2022. Examining the interplay between artificial intelligence and the agrifood industry. Artif. Intell. Agricult. 6, 111–128.
Salami, A., Kamara, A.B., Brixiova, Z., 2010. Smallholder Agriculture in East Africa: Trends, Constraints and Opportunities. African Development Bank Tunis, Tunisia.
Sarker, I.H., 2021. Machine learning: algorithms, real-world applications and research directions. SN Comput. Sci. 2, 1–21.
Schut, A.G., Giller, K.E., 2020. Soil-based, field-specific fertilizer recommendations are a pipe-dream. Geoderma 380, 114680.
Shi, P., Wang, Y., Xu, J., Zhao, Y., Yang, B., Yuan, Z., Sun, Q., 2021. Rice nitrogen nutrition estimation with rgb images and machine learning methods. Comput. Electron. Agric. 180, 105860.
Singh, A., Thakur, N., Sharma, A., 2016. A review of supervised machine learning algorithms. 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom). IEEE, pp. 1310–1315.
Subeesh, A., Bhole, S., Singh, K., Chandel, N., Rajwade, Y., Rao, K., Kumar, S., Jat, D., 2022. Deep convolutional neural network models for weed detection in polyhouse grown bell peppers. Artif. Intell. Agricult. 6, 47–54.
Suchithra, M., Pai, M.L., 2020. Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters. Informat. Process. Agricult. 7, 72–82.
Tapsoba, P.K., Aoudji, A.K.N., Kabore, M., Kestemont, M.P., Legay, C., Achigan-Dako, E.G., 2020. Sociotechnical context and Agroecological transition for smallholder farms in Benin and Burkina Faso. Agronomy 10, 1447.
Tavakoli, H., Gebbers, R., 2019. Assessing nitrogen and water status of winter wheat using a digital camera. Comput. Electron. Agric. 157, 558–567.
Tesfaye, A.A., Osgood, D., Aweke, B.G., 2021. Combining machine learning, space-time cloud restoration and phenology for farm-level wheat yield prediction. Artif. Intell. Agricult. 5, 208–222.
Thompson, L.J., Ferguson, R.B., Kitchen, N., Frazen, D.W., Mamo, M., Yang, H., Schepers, J.S., 2015. Model and sensor-based recommendation approaches for in-season nitrogen Management in Corn. Agron. J. 107, 2020–2030.
Timsina, J., Dutta, S., Devkota, K.P., Chakraborty, S., Neupane, R.K., Bishta, S., Amgain, L.P., Singh, V.K., Islam, S., Majumdar, K., 2021. Improved nutrient management in cereals using nutrient expert and machine learning tools: productivity, profitability and nutrient use efficiency. Agric. Syst. 192, 103181.
Trontelj ml, J., Chambers, O., 2021. Machine learning strategy for soil nutrients prediction using spectroscopic method. Sensors 21, 4208.
Venkataraju, A., Arumugam, D., Stepan, C., Kiran, R., Peters, T., 2023. A review of machine learning techniques for identifying weeds in corn. Smart Agricult. Technol. 3, 100102.
Wang, S., Guan, K., Wang, Z., Ainsworth, E.A., Zheng, T., Townsend, P.A., Liu, N., Nafziger, E., Masters, M.D., Li, K., et al., 2021a. Airborne hyperspectral imaging of nitrogen deficiency on crop traits and yield of maize by machine learning and radiative transfer modeling. Int. J. Appl. Earth Obs. Geoinf. 105, 102617.
Wang, X., Miao, Y., Dong, R., Zha, H., Xia, T., Chen, Z., Kusnierek, K., Mi, G., Sun, H., Li, M., 2021b. Machine learning-based in-season nitrogen status diagnosis and side-dress nitrogen recommendation for corn. Eur. J. Agron. 123, 126193.
Wu, J., 2017. Introduction to Convolutional Neural Networks. National Key Lab for Novel Software Technology. vol. 5. Nanjing University, China, p. 495.
Yan, X., Chen, X., Ma, C., Cai, Y., Cui, Z., Chen, X., Wu, L., Zhang, F., 2021. What are the key factors affecting maize yield response to and agronomic efficiency of phosphorus fertilizer in China? Field Crop Res. 270, 108221.
Yu, H., Liu, D., Chen, G., Wan, B., Wang, S., Yang, B., 2010. A neural network ensemble method for precision fertilization modeling. Math. Comput. Model. 51, 1375–1382.
Yu, X., Lu, H., Liu, Q., 2018. Deep-learning-based regression model and hyperspectral imaging for rapid detection of nitrogen concentration in oilseed rape (Brassica napus L.) leaf. Chemom. Intell. Lab. Syst. 172, 188–193.
Zou, J., Han, Y., So, S.S., 2008. Overview of artificial neural networks. Artif. Neural Net. 14–22.
Zupan, J., 1994. Introduction to artificial neural network (ann) methods: what they are and how to use them. Acta Chim. Slov. 41, 327–352.