HortScience 60(6):878–888. 2025. https://doi.org/10.21273/HORTSCI18508-25

A Data-driven Approach to Generating Foliar Nutrient Interpretation Ranges and Machine Learning-based Interpretation for Petunia

Patrick Veazie and Hsuan Chen
Department of Horticultural Science, North Carolina State University, 2721 Founders Drive, Raleigh, NC 27607, USA

Kristin Hicks
Agronomic Division, North Carolina Department of Agriculture and Consumer Services, Raleigh, NC 27607, USA

Jennifer Boldt
US Department of Agriculture, Agricultural Research Service, Application Technology Research, Unit, 2801 West Bancroft Street, Mail Stop 604, Toledo, OH 43606, USA

Brian Whipker
Department of Horticultural Science, North Carolina State University, 2721 Founders Drive, Raleigh, NC 27695, USA

Keywords. foliar tissue ranges, machine learning, nutrient distribution, plant nutrition, petunia

Abstract. **Historically, leaf tissue standards have been developed and used to interpret foliar tissue analyses for the majority of horticultural crops to diagnose nutrient disorders. However, leaf tissue standards for petunia (*Petunia ×hybrida*) are based on survey concentrations from small datasets. This study presents a novel method to create data-driven nutrient interpretation ranges by fitting models to provide more refined ranges of deficient, low, sufficient, high, and excessive for 11 essential elements based on 1420 data points. Data distributions were analyzed by fitting normal, Gamma, and Weibull distributions. Additionally, four machine learning algorithms J48 (a decision tree classifier), random forest (RF), which is a learning method that uses multiple decision trees, sequential minimal optimization (SMO), which is an optimization technique for support vector machines, and multilayer perceptron (MLP), which is a type of artificial neural network, were examined to determine if machine learning models could accurately classify foliar tissue analysis samples into the correct interpretation range. For all examined essential nutrients, J48 or RF yielded the highest classification accuracy compared with MLP or SMO. This study established the novel use of machine learning for interpreting petunia foliar nutrient analysis results with a higher accuracy rate than that of traditional statistical methods.**

The economic goal of growers is to produce high-quality plants while minimizing inputs such as fertilizers; however, nutrient deficiencies can occur if the nutrients are not supplied to the plant in available forms, at the required concentration, or at the appropriate time (Alem et al. 2015; van Iersel et al. 1998). Nutrient deficiencies can stunt plant growth, increase production time, and induce visual symptoms (Henry 2017). Nutrient toxicities, as a result of surplus fertilizer, can result in excess salinity, visual toxicity symptoms, and stunted plant growth (Alem et al. 2015). Qualitative and quantitative approaches for fertilization management exist (van Iersel et al. 1998). Although qualitative approaches, such as visual nutrient toxicity or deficiency symptoms, rely on physical changes already occurring within the plant, quantitative approaches, such as foliar tissue analyses, can detect variations before visual differences occurring. However, optimal foliar tissue nutrient analysis concentrations vary depending on the plant species and growth stage (Bryson and Mills 2015; Reuter and Robinson 1997).

Currently, foliar tissue analysis standards for horticultural crops are based on the survey approach (SA), which consists of sampling healthy plants to set a baseline standard for foliar nutrient concentrations for an actively growing healthy plant (Bryson and Mills 2015). Although this approach is limited because of the small sample used to establish the baseline, many analytical laboratories rely on standards set by the SA to evaluate and diagnose foliar samples submitted by growers and technical specialists for many specialty crops.

More robust evaluation standards that account for varying growing conditions and plant development stages are needed. Expansion from the SA has led to several refined evaluation methods, including the critical value approach (CVA) (Sumner 1990), compositional nutrient diagnosis (CND) (Parent and Dafir 1992), diagnosis and recommendation integrated system (DRIS) (Beaufils 1973), and sufficiency range approach (SRA) (Soltanpour et al. 1995). All four approaches have advantages and limitations when used to evaluate and diagnose plant nutrient status. The SRA provides an assessment of individual nutrient concentrations (deficient or sufficient) but does not explicitly account for interactions between nutrients that the DRIS provides. Although these methods provide a baseline for creating reference values for specialty crops, the limited sample numbers used with these methods can negatively impact the accuracy of the values identified by these methods.

To develop an interpretation model using the SRA that includes deficient, low, sufficient, high, and excessive ranges, an optimal distribution curve must be identified or established. However, most data tend to be skewed, thus making the normal distribution curve less suitable. Two distribution curves that account for possible skewness are Gamma and Weibull (Cera et al. 2022; Mhango et al. 2021; Slaton et al. 2021; Weibull 1951). Individual datasets should be evaluated using multiple distributions to determine the one that most accurately depicts the data.

Plant diagnostics can be challenging, even with well-defined leaf tissue concentration ranges, because of potential errors that can occur when interpreting laboratory analysis results. Machine learning (ML) provides the ability to use large datasets to understand and interpret data-intensive processes in the agricultural field (Liakos et al. 2018). Machine learning has already been used in crop management for yield prediction (Amatya et al. 2016; Ramos et al. 2017), disease detection (Chung et al. 2016; Ebrahimi et al. 2017), weed detection (Pantazi et al. 2016, 2017), and nutrient deficiency detection through imaging (Li et al. 2022; Shi et al. 2021). Using industry-wide standardized published laboratory analysis methods for foliar nutrient analyses allows ML to be used for diagnostics regardless of onsite equipment. This research aimed to use the foliar interpretation ranges

Received for publication 6 Feb 2025. Accepted for publication 7 Mar 2025.
Published online 8 May 2025.
We thank Fred C. Gloeckner Foundation and American Floral Endowment for funding associated with this project. We thank Douglas Sturtz and Mona-Lisa Banks at the USDA-ARS for technical assistance and elemental analysis of the USDA-ARS petunia leaf tissue samples. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture. USDA is an equal opportunity provider and employer.
P.V. is the corresponding author. E-mail: phveazie@ncsu.edu.

developed to create ML algorithms that can aid interpretation using readily available foliar tissue concentration testing methods.

A variety of ML algorithms with various architectures have been developed for different purposes. Decision trees, such as J48, use branches to group data into subpopulations while creating associated tree graphs (Ennaji et al. 2023). Each branch of a tree uses a pairwise comparison for a particular attribute (Mingers 1989). Similarly, random forest (RF) is a ML decision tree-based algorithm that combines a sequence of trees for better predictive performance (Ennaji et al. 2023). In contrast, artificial neural networks (ANNs) such as multilayer perceptron (MLP) use radial basis function networks, backpropagation, and perceptron algorithms to build predictive models for regression or classification (Griffel et al. 2023). Support vector machines, such as sequential minimal optimization (SMO), were originally designed for binary classification by creating a linear separation hyperplane (Keerthi et al. 2001). To improve foliar tissue nutrient interpretation standards for petunia (*Petunia ×hybrida*), refined evaluation ranges were first established for the following 11 essential elements commonly analyzed using leaf tissue analysis: nitrogen (N), phosphorus (P), potassium (K), calcium (Ca), magnesium (Mg), sulfur (S), boron (B), copper (Cu), iron (Fe), manganese (Mn), and zinc (Zn). In addition, creating an automated system to evaluate leaf tissue analysis results would increase the accuracy of diagnosing nutrient disorders. Therefore, the study objectives were to develop more robust leaf tissue classification ranges and create an automated ML-based classification system for petunia tissue nutrient interpretation.

Materials and Methods

Sample collection. Foliar tissue analysis samples were obtained from controlled federal or university research studies conducted in North Carolina or Ohio and supplemented with samples from public and commercial analytical laboratories. Leaf tissue samples (n = 1420) included only petunias grown in controlled environments, such as greenhouses and growth chambers (Table 1), and were analyzed for each study based on the cited procedures. Because of the short production time used with bedding plant production, only one set of foliar nutrient standards for the entire approximately 45-d to 60-d production cycle was developed.

Nutrient distribution statistical analysis. Distribution analyses were conducted using R studio (version 4.1.1; R Foundation for Statistical Computing, Vienna, Austria). Each element was modeled independently, and outliers that were extremely excessive (greater than biologically feasible or a significant break in the population) were removed before further analyses were performed. Data were fit to normal, Gamma, and Weibull distributions, and the three statistical distributions were compared (Cera et al. 2022; Mhango et al. 2021; Slaton et al. 2021; Weibull 1951). Corresponding *P* values that described the fitness of the data in the statistical distributions were calculated based on the Shapiro-Wilk test for normality (normal and Gamma distributions) or the Kolmogorov-Smirnov test (Weibull distribution). The optimal distribution was selected based on the lowest Bayesian information criterion (BIC) value and visual fitness. Results were illustrated using ggplot2 (Wickham 2011) in R. For macronutrients (N, P, K, Ca, Mg, and S), the deficiency range was established based on the left tail of a 95% distribution (lowest 2.5% of the samples that contained >40 observations), the low range corresponded to the region between the lowest 2.5% of the observations and the 0.25 quantile, the sufficiency range was the area between the 0.25 and 0.75 quantiles, the high range corresponded to the region between the 0.75 quantile and the highest 2.5% of the observations, and the excessive range was based on the right tail of a 95% distribution (highest 2.5% of the samples that contained >40 observations). For micronutrients (B, Cu, Fe, Mn, and Zn), the deficiency range was established based on the left tail of a 90% distribution (lowest 5% of the samples), the low range corresponded to the region between the lowest 5% of the observations and the 0.25 quantile, the high range corresponded to the region between the 0.75 quantile and the highest 5% of the observations, and the excessive range was based on the right tail of a 90% distribution (highest 5% of the distribution).

Machine learning algorithm development. Foliar tissue concentrations were classified using the Waikato Environment for Knowledge Analysis (WEKA) (version 3.8.3, The University of Waikato, Hamilton, New Zealand, https://www.cs.waikato.ac.nz/ml/weka/). Within each element, samples were individually assigned to one of five nutrient classification ranges (deficient, low, sufficient, high, or excessive) based on ranges established by the nutrient distribution curves. The single element being classified was assigned the corresponding interpretation range and used as the class variable. Then, two decision trees (J48 and RF) and four different pattern-recognition ML algorithms were used to analyze the

Table 1. Sources of petunia leaf tissue nutrient data used in the development of the sufficiency range approach (SRA) distribution model.

Source	Sample size	Sample type	Notes/reference
North Carolina Department of Agriculture Laboratory	151	Diagnostic	Grower submitted diagnostic and predictive samples (unpublished)
North Carolina State University	120	Research	Unpublished electrical conductivity rate study
North Carolina State University	395	Predictive	Grower survey samples
North Carolina State University	52	Research[i]	Phosphorus rate study. Henry et al. (2016)
US Department of Agriculture-Agricultural Research Service	702	Research[ii]	Boldt J, unpublished and published data

[i] Citations of leaf tissue analysis methods used for data: see Henry et al. (2016)
[ii] Multiple published and unpublished studies: see Boldt (2018), Boldt et al. (2018), Boldt and Altland (2019), and Boldt and Altland (2021).

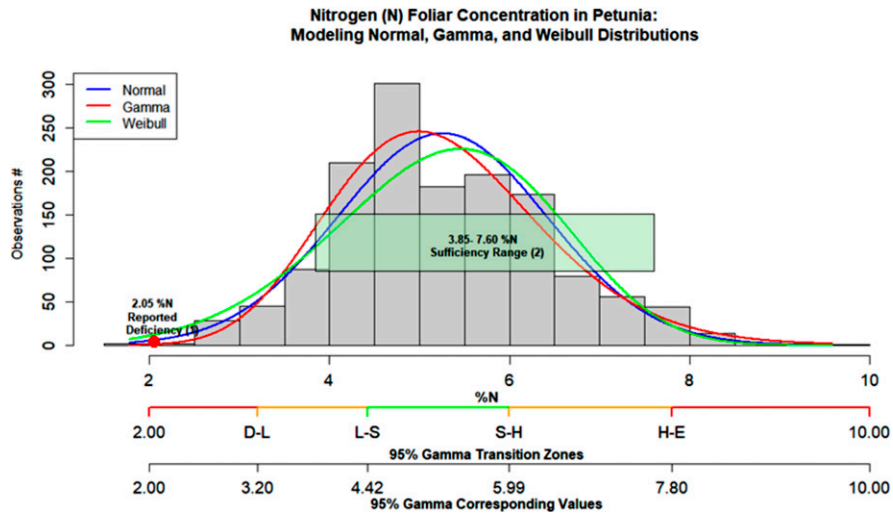

Fig. 1. Distribution of nitrogen (N) foliar concentrations in petunia (n = 1420) modeled using normal, Gamma, and Weibull distributions. Interpretation ranges based on the Gamma distribution define the following four transition zones: deficient to low (D-L), low to sufficient (L-S), sufficient to high (S-H), and high to excessive (H-E), which correspond to N concentrations of 3.20%, 4.42%, 5.99%, and 7.80%, respectively. Previously reported N sufficiency and deficiency ranges are based on studies by Pitchay et al. (2002) and Bryson and Mills (2015) and are reported for comparison.

Table 2. Percent correct classification (PCC) of nitrogen (N) values using four machine learning algorithms (MLP, SMO, J48, and RF) with two cross-validation methods (10-fold and 66% split). Models were first run using N alone, and then they progressively incorporated additional elements until all 11 were included. PCC represents the percentage of samples correctly classified based on the N concentration, indicating the model's ability to accurately determine the N classification of deficient, low, sufficient, high, or excessive.

Percent correct classification												
Elements included in the model												
Algorithm	Cross-validation	1 N	2 Cu	3 Zn	4 Mn	5 Ca	6 Fe	7 Mg	8 S	9 P	10 K	11 B
MLP	10-fold	92.46	92.25	92.54	95.28	94.86	97.04	94.23	94.58	93.94	94.01	90.85
	66% Split	91.10	90.06	94.00	93.17	93.58	89.86	89.65	92.96	90.68	88.62	89.65
SMO	10-fold	86.69	86.07	85.85	85.92	85.49	85.56	85.21	84.72	84.51	85.14	84.79
	66% Split	81.57	81.57	81.99	81.78	81.78	81.99	82.19	81.16	80.75	79.71	80.12
J48	10-fold	99.79	99.79	99.79	99.79	99.79	99.79	99.79	99.79	99.79	99.79	99.79
	66% Split	99.79	99.79	99.79	99.79	99.79	99.79	99.79	99.79	99.79	99.79	99.79
RF	10-fold	99.72	99.72	99.72	99.72	99.72	99.72	99.51	99.72	99.58	99.37	99.44
	66% Split	99.59	99.59	99.59	99.59	99.59	99.38	99.17	99.59	99.17	99.38	99.38

B = boron; Ca = calcium; Cu = copper; Fe = iron; K = potassium; Mg = magnesium; MLP = multilayer perceptron; Mn = manganese; N = nitrogen; P = phosphorus; RF = random forest; S = sulfur; SMO = sequential minimal optimization; Zn = zinc.

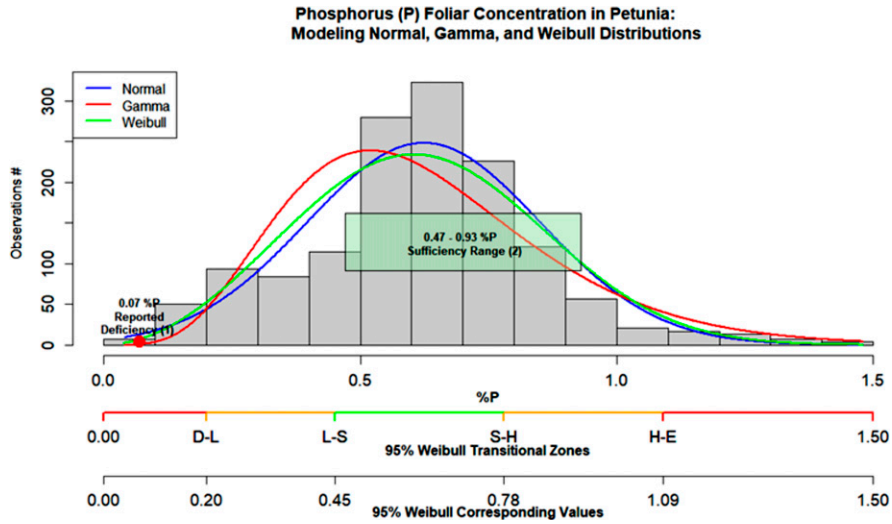

Fig. 2. Phosphorus (P) foliar concentrations of petunia (n = 1420) modeled using normal, Gamma, and Weibull distributions. Interpretation ranges based on the Gamma distribution define the following four transition zones: deficient to low (D-L), low to sufficient (L-S), sufficient to high (S-H), and high to excessive (H-E), which correspond to P concentrations of 0.20%, 0.45%, 0.78%, and 1.09%, respectively. Previously reported P sufficiency and deficiency ranges are based on studies by Pitchay et al. (2002) and Bryson and Mills (2015) and are reported for comparison.

dataset: an SMO, an SVM, an MLP, and an artificial neural network (ANN) (Witten and Frank 2005).

The ML algorithms were compared based on whether they could correctly classify a foliar nutrient concentration interpretation range. To create a model specific to each of the examined nutrients, all 11 essential elements from the dataset (N, P, K, Ca, Mg, S, B, Cu, Fe, Mn, and Zn) were ranked based on Shannon entropy (information gain) in the dichotomous classification assignment by SVMs (Eibe et al. 2016; Keerthi et al. 2001). Then, information gain ranking was used to identify those elements that were most relevant to the assignment of each element to a classification range to determine the inclusion order. Reduction of data dimensionality for each ML algorithm was performed by the sequential exclusion of elements least relevant to the class assignment until one element was remaining. This step eliminates the overfitting of the ML classifiers. To identify the minimum number of elements required for classification of foliar concentration patterns, each element that contributed an information gain value >0.0 was removed independently. This step identified the underfitting of the ML classifiers. The point of optimal classification was determined to be the least number of elements that yielded the greatest percentage of correctly classified instances.

Class assignment of all ML algorithms was evaluated independently by two cross-validation strategies. The first was a percentage

Table 3. Percent correct classification (PCC) of phosphorus (P) values using four machine learning algorithms (MLP, SMO, J48, and RF) with two cross-validation methods (10-fold and 66% split). Models were first run using P alone, and then they progressively incorporated additional elements until all 11 were included. PCC represents the percentage of samples correctly classified based on the P concentration, indicating the model's ability to accurately determine the P classification of deficient, low, sufficient, high, or excessive.

Percent correct classification												
Elements included in the model												
Algorithm	Cross-validation	1 P	2 Mg	3 N	4 Ca	5 Mn	6 Fe	7 Cu	8 S	9 B	10 K	11 Zn
MLP	10-fold	92.46	92.82	92.39	93.03	92.18	92.68	94.51	92.82	92.11	92.96	92.11
	66% Split	92.55	91.10	95.24	92.13	92.34	95.24	92.75	95.03	94.41	92.34	93.79
SMO	10-fold	83.87	85.14	84.86	85.14	84.93	84.86	84.51	85.35	84.51	85.35	85.42
	66% Split	86.34	85.51	86.54	85.51	85.92	85.92	85.92	86.13	86.54	86.13	86.13
J48	10-fold	99.86	99.86	99.86	99.86	99.86	99.86	99.86	99.86	99.86	99.86	99.86
	66% Split	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
RF	10-fold	100.00	100.00	100.00	100.00	100.00	99.93	99.86	100.00	99.86	99.79	99.65
	66% Split	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.79	99.79

B = boron; Ca = calcium; Cu = copper; Fe = iron; K = potassium; Mg = magnesium; MLP = multilayer perceptron; Mn = manganese; N = nitrogen; P = phosphorus; RF = random forest; S = sulfur; SMO = sequential minimal optimization; Zn = zinc.

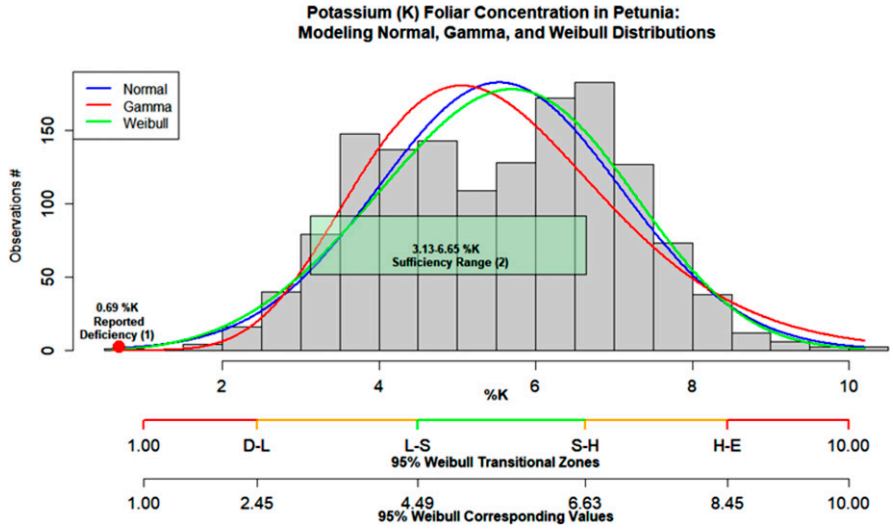

Fig. 3. Potassium (K) foliar concentrations of petunia (n = 1420) modeled using normal, Gamma, and Weibull distributions. Interpretation ranges based on the Gamma distribution define the following four transition zones: deficient to low (D-L), low to sufficient (L-S), sufficient to high (S-H), and high to excessive (H-E), which correspond to K concentrations of 2.45%, 4.49%, 6.63%, and 8.45%, respectively. Previously reported K sufficiency and deficiency ranges are based on studies by Pitchay et al. (2002) and Bryson and Mills (2015) and are reported for comparison.

split, whereby 66% of the total data were randomly used for training and the remaining 34% of the data were used for testing. The second cross-validation was a stratified hold-out (n-fold) method with 10-fold data, with nine-fold of the randomized foliar concentration data used for training and one-fold used for testing. This was repeated eight times so that all replicate samples were used at least once for testing and the average model performance was recorded for each algorithm evaluated.

The performances of the four ML algorithms, SVM, MLP, and two decision trees were determined using the percentage correct classification (PCC) during the cross-validations. The PCC indicates the likelihood that each sample could be accurately assigned to the respective nutrient category based on the foliar nutrient concentration data provided. Kappa statistics and receiver-operating characteristic scores were also recorded. Any kappa statistic >0 and receiver-operating characteristic score >0.5 indicated that the ML classifier performed better than random chance.

Results and Discussion

Nitrogen. Of the examined models, the Gamma distribution provided the best representation of the N foliar concentrations because it had the lowest BIC value and visually represented the tails of the data (Fig. 1). A recommended sufficiency range of 4.42% to 5.99% N narrowed the previously recommended range of 3.85% to 7.60% N reported by Bryson and Mills (2015). The lowest 2.5% of the represented samples yielded a deficiency value of 3.20% N, which encompassed the previously reported deficiency value of 2.05% N reported by Pitchay et al. (2002). Although N toxicity is rare and values have not been reported for petunia, toxicity can occur when high concentrations of ammonium (NH_4^+) are supplied and temperatures are low (<20 °C) or excessive (>40 °C), the substrate is waterlogged, or a substrate pH is <5.6 (Handreck and Black 2002). Luxury consumption of N can inhibit flowering and induce potential antagonistic relationships with other essential nutrients (Marschner 1995). This work established that an excessive concentration was >7.80% N and offered an initial value for future refinement.

All ML algorithms for N yielded a PCC >79.71%, which was a large increase over the random chance of 20% (Table 2). However, J48 provided the best classification of N with a minimum PCC of 99.79% (Table 2). The MLP yielded a PCC range of 88.62% to 97.04%, and the SMO algorithm yielded a range between 79.71% and 86.69%. A decision tree containing four to 10 elements would provide the greatest PCC while accounting for the reported interaction of N × K.

Phosphorus. Phosphorus foliar concentrations were best represented using a Weibull distribution (Fig. 2). Although a smaller BIC value was achieved by the normal distribution, the Weibull distribution provided a better representation of the left and right tails of the sample data. Based on the Weibull distribution, a recommended sufficiency range of 0.45% to 0.78% P would narrow the previously reported sufficiency range of 0.47% to 0.93% P recommended by Bryson and Mills (2015). Additionally, a deficiency range of <0.20% P encompassed the previously reported deficiency value of 0.07% P (Pitchay et al. 2002). Although P toxicity in petunia has not been reported, a P foliar concentration exceeding 2% can be considered toxic for most species (Marschner 1995). Additionally, excessive P concentrations can antagonize the uptake of Cu, Fe, and Zn. The Weibull distribution established >1.09% P as excessive for petunia.

The P foliar tissue concentrations were best classified by the decision tree algorithms J48 and RF, which yielded minimum PCCs of 99.85% and 99.65%, respectively, for both cross-validations (Table 3). The SMO yielded

Table 4. Percent correct classification (PCC) of potassium (K) values using four machine learning algorithms (MLP, SMO, J48, and RF) with two cross-validation methods (10-fold and 66% split). Models were first run using K alone, and then they progressively incorporated additional elements until all 11 were included. PCC represents the percentage of samples correctly classified based on the K concentration, indicating the model's ability to accurately determine the K classification of deficient, low, sufficient, high, or excessive.

		Percent correct classification										
		Elements included in the model										
Algorithm	Cross-validation	1 K	2 S	3 Mn	4 Cu	5 Mg	6 Ca	7 N	8 P	9 Zn	10 B	11 Fe
MLP	10-fold	95.56	95.63	95.42	95.28	95.28	95.28	94.72	94.86	95.35	95.63	94.01
	66% Split	93.58	94.00	94.20	92.96	94.10	93.79	93.58	94.00	92.13	93.17	91.20
SMO	10-fold	95.85	93.17	90.63	90.14	90.00	89.01	88.59	88.45	87.75	88.17	88.17
	66% Split	93.37	91.72	89.44	89.86	89.44	88.82	86.75	86.96	87.75	87.16	86.75
J48	10-fold	99.72	99.72	99.72	99.72	99.72	99.72	99.72	99.72	99.72	99.72	99.72
	66% Split	99.79	99.79	99.79	99.79	99.79	99.79	99.79	99.79	99.79	99.79	99.79
RF	10-fold	99.86	99.86	99.86	99.86	99.79	99.79	99.65	99.79	99.79	99.72	99.51
	66% Split	99.79	99.79	99.79	99.79	99.79	99.79	99.79	99.79	99.17	99.17	98.34

B = boron; Ca = calcium; Cu = copper; Fe = iron; K = potassium; Mg = magnesium; MLP = multilayer perceptron; Mn = manganese; N = nitrogen; P = phosphorus; RF = random forest; S = sulfur; SMO = sequential minimal optimization; Zn = zinc.

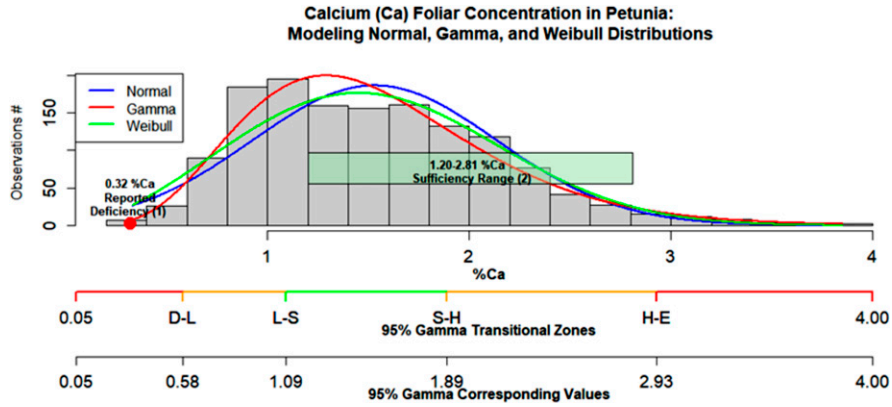

Fig. 4. Calcium (Ca) foliar concentrations of petunia (n = 1420) modeled using normal, Gamma, and Weibull distributions. Interpretation ranges based on the Gamma distribution define the following four transition zones: deficient to low (D-L), low to sufficient (L-S), sufficient to high (S-H), and high to excessive (H-E), which correspond to Ca concentrations of 0.58%, 1.09%, 1.89%, and 2.93%, respectively. Previously reported Ca sufficiency and deficiency ranges are based on studies by Pitchay et al. (2002) and Bryson and Mills (2015) and are reported for comparison.

and Gamma distributions for K foliar concentrations (Fig. 3). A recommended sufficiency range of 4.49% to 6.63% K would narrow the previously reported sufficiency range of 3.13% to 6.65% K (Bryson and Mills 2015). A deficiency range of <2.45% K encompassed the previously reported deficiency value of 0.69% K reported by Pitchay et al. (2002). The threshold for excessive K was established at 8.45% K. When K foliar concentrations become excessive, antagonistic interactions with Ca, Mg, and B have been observed (Marschner 1995). High K levels can compete with Ca and Mg for uptake, potentially leading to deficiencies that affect cell wall stability, enzyme activation, and photosynthetic efficiency (Marschner 1995).

All algorithms yielded a PCC classification >86.75% when additional elements, other than K, were incorporated (Table 4). However, similar to other elements, SMO

Table 5. Percent correct classification (PCC) of calcium (Ca) values using four machine learning algorithms (MLP, SMO, J48, and RF) with two cross-validation methods (10-fold and 66% split). Models were first run using Ca alone, and then they progressively incorporated additional elements until all 11 were included. PCC represents the percentage of samples correctly classified based on the Ca concentration, indicating the model's ability to accurately determine the Ca classification of deficient, low, sufficient, high, or excessive.

Percent correct classification												
		Elements included in the model										
Algorithm	Cross-validation	1 Ca	2 Mn	3 Fe	4 Mg	5 N	6 S	7 P	8 Cu	9 Zn	10 K	11 B
MLP	10-fold	94.79	94.51	95.07	96.55	95.92	95.07	94.23	94.44	93.87	93.10	93.38
	66% Split	95.65	96.48	95.65	94.82	94.41	94.41	94.00	95.45	93.79	92.75	94.20
SMO	10-fold	92.75	92.25	92.32	92.75	92.39	93.03	93.54	85.35	92.32	91.20	91.27
	66% Split	92.96	91.51	91.72	92.32	91.72	91.30	91.10	86.13	90.68	89.44	89.44
J48	10-fold	99.79	99.79	99.79	99.79	99.79	99.79	99.79	99.86	99.00	98.79	99.79
	66% Split	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
RF	10-fold	99.79	99.79	99.79	99.79	99.79	99.79	99.65	99.79	99.51	99.58	99.58
	66% Split	100.00	100.00	100.00	100.00	100.00	99.79	100.00	100.00	99.79	99.38	98.55

B = boron; Ca = calcium; Cu = copper; Fe = iron; K = potassium; Mg = magnesium; MLP = multilayer perceptron; Mn = manganese; N = nitrogen; P = phosphorus; RF = random forest; S = sulfur; SMO = sequential minimal optimization; Zn = zinc.

the lowest PCC (approximately 85.48%) averaged across the two cross-validation methods (Table 3). Using an RF algorithm that contained between seven and 11 elements (Table 3) allowed for a very high PCC (>99.65%) while still accounting for reported antagonistic interactions of P × K, P × Cu, P × Fe, and P × Zn (Marschner 1995).

Potassium. A Weibull distribution yielded a smaller BIC value than that of the normal yielded the lowest PCC when compared with J48, RF, and MLP. As additional elements were incorporated in the SMO model, a general negative trend for PCC was observed (Table 4). This suggested that a reduction of data dimensionality is required to achieve the greatest accuracy while also preventing underfitting. J48 achieved the greatest PCC (99.76%) when averaged across the 66% split and for the 10-fold cross-validation (99.79%), and it could be reduced to include four to eight elements to account for nutrient interactions while still achieving a PCC >99% (Table 4).

Calcium. Calcium foliar concentrations were best represented by the Gamma distribution, which yielded the smallest BIC value of the three models (Fig. 4). A sufficiency range of 1.09% to 1.89% Ca would decrease and narrow the previously reported sufficiency range of 1.20% to 2.81% Ca (Bryson and Mills, 2015). The Ca deficiency foliar concentration threshold of 0.58% Ca encompassed a previously reported value of 0.32% Ca (Pitchay et al. 2002). There are no published excessive or toxic Ca values for petunia. However, luxury consumption of Ca can

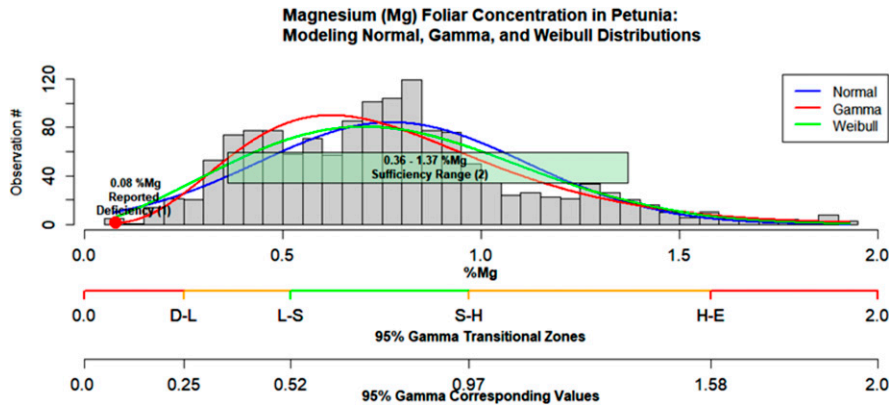

Fig. 5. Magnesium (Mg) foliar concentrations of petunia (n = 1420) modeled using normal, Gamma, and Weibull distributions. Interpretation ranges based on the Gamma distribution define the following four transition zones: deficient to low (D-L), low to sufficient (L-S), sufficient to high (S-H), and high to excessive (H-E), which correspond to Mg concentrations of 0.25%, 0.52%, 0.97%, and 1.58%, respectively. Previously reported Mg sufficiency and deficiency ranges are based on studies by Pitchay et al. (2002) and Bryson and Mills (2015) and are reported for comparison.

Table 6. Percent correct classification (PCC) of magnesium (Mg) values using four machine learning algorithms (MLP, SMO, J48, and RF) with two cross-validation methods (10-fold and 66% split). Models were first run using Mg alone, and then they progressively incorporated additional elements until all 11 were included. PCC represents the percentage of samples correctly classified based on the Mg concentration, indicating the model's ability to accurately determine the Mg classification of deficient, low, sufficient, high, or excessive.

Percent correct classification												
		Elements included in the model										
Algorithm	Cross-validation	1 Mg	2 Ca	3 Mn	4 P	5 B	6 N	7 S	8 Fe	9 Zn	10 K	11 Cu
MLP	10-fold	93.94	94.37	96.06	94.86	94.15	95.07	93.41	92.68	93.31	92.54	92.82
	66% Split	91.10	90.89	91.10	91.10	90.68	92.34	91.72	91.51	91.10	94.82	92.34
SMO	10-fold	89.58	89.72	89.43	89.79	90.00	90.00	89.58	89.58	89.37	89.44	89.30
	66% Split	86.54	88.41	88.41	87.58	87.37	88.61	87.78	87.58	87.99	87.99	87.99
J48	10-fold	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	66% Split	99.59	99.59	99.59	99.59	99.59	99.59	99.59	99.59	99.59	99.59	99.59
RF	10-fold	100.00	100.00	100.00	100.00	100.00	99.79	99.72	99.86	99.72	99.72	99.51
	66% Split	99.59	99.59	99.59	99.59	99.59	99.59	99.17	99.59	99.59	99.38	98.96

B = boron; Ca = calcium; Cu = copper; Fe = iron; K = potassium; Mg = magnesium; MLP = multilayer perceptron; Mn = manganese; N = nitrogen; P = phosphorus; RF = random forest; S = sulfur; SMO = sequential minimal optimization; Zn = zinc.

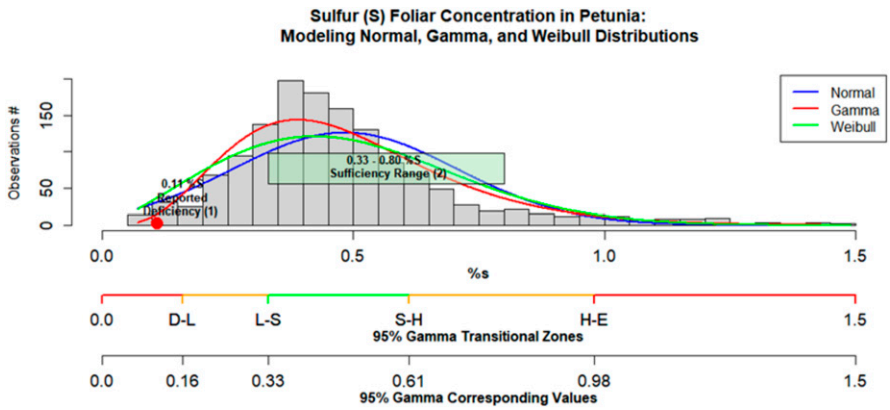

Fig. 6. Sulfur (S) foliar concentrations of petunia (n = 1420) modeled using normal, Gamma, and Weibull distributions. Interpretation ranges based on the Gamma distribution define the following four transition zones: deficient to low (D-L), low to sufficient (L-S), sufficient to high (S-H), and high to excessive (H-E), which correspond to S concentrations of 0.16%, 0.33%, 0.61%, and 0.98%, respectively. Previously reported S sufficiency and deficiency ranges are based on studies by Pitchay et al. (2002) and Bryson and Mills (2015) and are reported for comparison.

occur when abundant Ca is supplied, and this may be reflected in the higher previously recommended range of 1.20% to 2.81% Ca (Bryson and Mills 2015). Luxury consumption should be monitored for the possibility of interference with P, K, Mg, Fe, B, Mn, and Zn uptake (Marschner 1995). Using the Gamma distribution, the upper 2.5% of samples set the excessive range threshold at 2.93% Ca (Fig. 4). The proposed excessive range established an upper threshold to minimize the occurrence of decreased K and Mg uptake because of excessively high Ca foliar concentrations.

Additionally, Ca was best classified by the decision tree algorithms J48 and RF, which yielded minimum PCCs of 99.79% and 99.38%, respectively (Table 5). The MLP yielded a PCC range of 92.75% to 96.55%, and the SMO algorithm yielded a PCC range between 85.35% and 93.54% (Table 5). Although all algorithms yielded greater than a random chance of 20%, RF consistently yielded the greatest PCC. An algorithm that contained between four and 11 elements (Table 5) would account for known interactions of Ca × Mg, Ca × P, Ca × K, Ca × Fe, Ca × B, and Ca × Mn (Marschner 1995).

Magnesium. A Gamma distribution yielded the lowest BIC compared with that of the other two examined distributions for foliar Mg (Fig. 5). The identified sufficiency range of 0.52% to 0.97% Mg was within the previously suggested sufficiency range of 0.36% to 1.37% Mg (Bryson and Mills 2015) and offered a refined range. A deficiency range of <0.25% Mg encompassed the reported deficiency concentration of 0.08% Mg (Pitchay et al. 2002). This established the first reported excessive Mg concentration for petunia of 1.58% Mg.

Table 7. Percent correct classification (PCC) of sulfur (S) values using four machine learning algorithms (MLP, SMO, J48, and RF) with two cross-validation methods (10-fold and 66% split). Models were first run using S alone, and then they progressively incorporated additional elements until all 11 were included. PCC represents the percentage of samples correctly classified based on the S concentration, indicating the model's ability to accurately determine the S classification of deficient, low, sufficient, high, or excessive.

Percent correct classification												
		Elements included in the model										
Algorithm	Cross-validation	1 S	2 Mn	3 B	4 Cu	5 Zn	6 K	7 Ca	8 P	9 Mg	10 N	11 Fe
MLP	10-fold	95.28	89.30	94.93	94.23	94.23	94.08	93.52	92.89	93.59	93.66	92.54
	66% Split	92.34	90.27	93.58	93.58	93.79	93.17	92.34	92.55	91.30	91.30	90.68
SMO	10-fold	73.62	73.59	73.03	73.52	75.42	77.96	78.52	79.94	79.37	80.21	80.28
	66% Split	75.78	75.16	74.74	74.53	74.47	79.71	79.92	79.30	80.95	81.16	80.95
J48	10-fold	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	66% Split	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
RF	10-fold	100.00	100.00	100.00	100.00	100.00	99.93	99.93	99.93	99.93	99.93	99.86
	66% Split	100.00	100.00	100.00	100.00	100.00	99.79	99.59	99.59	99.79	99.38	99.38

B = boron; Ca = calcium; Cu = copper; Fe = iron; K = potassium; Mg = magnesium; MLP = multilayer perceptron; Mn = manganese; N = nitrogen; P = phosphorus; RF = random forest; S = sulfur; SMO = sequential minimal optimization; Zn = zinc.

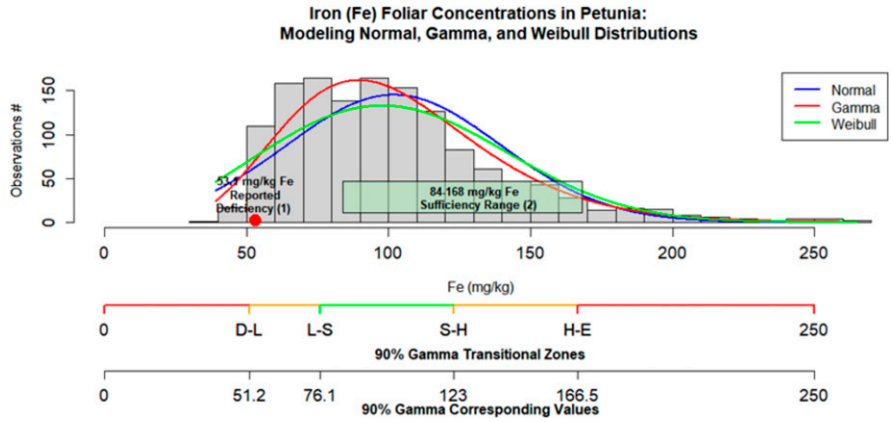

Fig. 7. Iron (Fe) foliar concentrations of petunia (n = 1420) modeled using normal, Gamma, and Weibull distributions. Interpretation ranges based on the Gamma distribution define the following four transition zones: deficient to low (D-L), low to sufficient (L-S), sufficient to high (S-H), and high to excessive (H-E), which correspond to Fe concentrations of 51.2, 76.1, 123.0, and 166.5 mg·kg^{-1}, respectively. Previously reported Fe sufficiency and deficiency ranges are based on studies by Pitchay et al. (2002) and Bryson and Mills (2015) and are reported for comparison.

The Mg foliar tissue concentrations were best classified by the decision tree algorithms J48 and RF and yielded minimum PCCs of 99.59% and 99.17%, respectively, with J48 consistently yielding an average PCC of 99.79% across both cross-validations regardless of the number of elements included in the model (Table 6). The SMO yielded the lowest PCC, on average, of approximately 88.72% across the two cross-validation methods (Table 6). Using an RF algorithm that contains between four and 10 elements (Table 6) allows for optimal PCC while still accounting for reported antagonistic interactions of Mg × K and Mg × Ca (Marschner 1995).

Sulfur. Of the examined models, a Gamma distribution optimally represented S foliar tissue concentrations (Fig. 6). A recommended sufficiency range of 0.33% to 0.61% S would narrow the previous sufficiency range of 0.33% to 0.80% S (Bryson and Mills 2015). Additionally, a deficiency range of <0.16% S encompassed the previously reported 0.11% S deficiency value at which visual symptoms were observed (Pitchay et al. 2002), although no S toxicity values have been previously reported for luxury consumption. This study defined >0.98% S as excessive.

Additionally, S was best classified by the decision tree algorithms J48 and RF, which collectively yielded a minimum PCC of 99.38% (Table 7). The MLP yielded a PCC range of 89.30% to 95.28%, and the SMO algorithm yielded a range between 73.03% and 81.16% (Table 5). Although all algorithms developed yields greater than what was expected with a random chance of 20%, J48 consistently yielded the greatest PCC across the algorithm types evaluated. A J48 algorithm containing between four and 11 elements (Table 7) reduced data dimensionality while still providing a similar PCC.

Iron. The Fe foliar tissue concentrations were best represented by a Gamma distribution, which yielded the smallest BIC values

Table 8. Percent correct classification (PCC) of iron (Fe) values using four machine learning algorithms (MLP, SMO, J48, and RF) with two cross-validation methods (10-fold and 66% split). Models were first run using Fe alone, and then they progressively incorporated additional elements until all 11 were included. PCC represents the percentage of samples correctly classified based on the Fe concentration, indicating the model's ability to accurately determine the Fe classification of deficient, low, sufficient, high, or excessive.

		Percent correct classification										
		Elements included in the model										
Algorithm	Cross-validation	1 Fe	2 Mn	3 Ca	4 Cu	5 P	6 N	7 B	8 Mg	9 S	10 K	11 Zn
MLP	10-fold	68.31	69.23	70.70	68.31	69.65	70.28	69.86	68.52	69.15	68.87	67.75
	66% Split	81.57	76.81	81.57	78.67	79.30	77.64	82.82	79.92	78.88	75.98	74.53
SMO	10-fold	47.32	50.92	60.42	60.70	60.00	60.35	61.90	62.25	63.10	62.18	62.11
	66% Split	45.55	45.55	60.66	61.08	60.66	59.42	61.28	61.49	60.87	60.25	60.04
J48	10-fold	99.72	99.72	99.72	99.72	99.72	99.72	99.72	99.72	99.72	99.72	99.72
	66% Split	99.59	99.59	99.59	99.59	99.59	99.59	99.59	99.59	99.59	99.59	99.59
RF	10-fold	99.86	99.86	99.79	99.86	99.79	99.72	99.65	99.65	99.65	99.51	99.30
	66% Split	99.79	99.79	99.79	99.79	99.79	99.59	99.59	99.59	99.79	99.59	99.79

B = boron; Ca = calcium; Cu = copper; Fe = iron; K = potassium; Mg = magnesium; MLP = multilayer perceptron; Mn = manganese; N = nitrogen; P = phosphorus; RF = random forest; S = sulfur; SMO = sequential minimal optimization; Zn = zinc.

Monitoring Mg foliar concentrations is essential because Mg deficiency disrupts the loading of sucrose into the phloem (Guo et al. 2016) and excessive foliar Mg concentrations inhibit photosynthesis and plant growth (Rao et al. 1987).

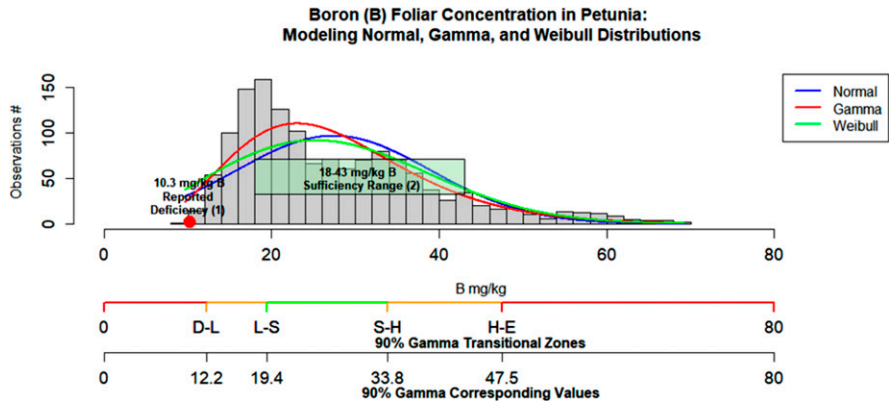

Fig. 8. Manganese (Mn) foliar concentrations of petunia modeled using normal, Gamma, and Weibull distributions. Interpretation ranges based on the Gamma distribution define the following four transition zones: deficient to low (D-L), low to sufficient (L-S), sufficient to high (S-H), and high to excessive (H-E), which correspond to Mn concentrations of 19.2, 44.2, 108.4, and 180.2 mg·kg^{-1}, respectively. Previously reported Mn sufficiency and deficiency ranges are based on studies by Pitchay et al. (2002) and Bryson and Mills (2015) and are reported for comparison.

Table 9. Percent correct classification (PCC) of manganese (Mn) values using four machine learning algorithms (MLP, SMO, J48, and RF) with two cross-validation methods (10-fold and 66% split). Models were first run using Mn alone, and then they progressively incorporated additional elements until all 11 were included. PCC represents the percentage of samples correctly classified based on the Mn concentration, indicating the model's ability to accurately determine the Mn classification of deficient, low, sufficient, high, or excessive.

		Percent correct classification										
		Elements included in the model										
Algorithm	Cross-validation	1 Mn	2 Ca	3 Fe	4 B	5 Mg	6 Cu	7 N	8 S	9 K	10 Zn	11 P
MLP	10-fold	96.20	96.06	95.07	95.07	94.86	95.28	94.08	94.51	93.52	93.45	93.45
	66% Split	97.52	97.10	95.65	94.41	96.27	95.86	94.82	94.00	93.37	94.00	92.55
SMO	10-fold	88.59	83.03	83.03	84.72	84.30	84.79	85.14	85.99	86.27	86.20	86.20
	66% Split	91.30	79.71	79.71	82.40	81.57	81.16	82.40	85.09	84.47	84.27	84.68
J48	10-fold	99.79	99.79	99.79	99.79	99.79	99.79	99.79	99.79	99.79	99.79	99.79
	66% Split	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
RF	10-fold	99.79	99.79	99.79	99.79	99.79	99.79	99.72	99.86	99.86	99.79	99.65
	66% Split	99.79	99.79	99.79	99.79	99.79	99.79	99.79	99.79	99.79	99.38	99.17

B = boron; Ca = calcium; Cu = copper; Fe = iron; K = potassium; Mg = magnesium; MLP = multilayer perceptron; Mn = manganese; N = nitrogen; P = phosphorus; RF = random forest; S = sulfur; SMO = sequential minimal optimization; Zn = zinc.

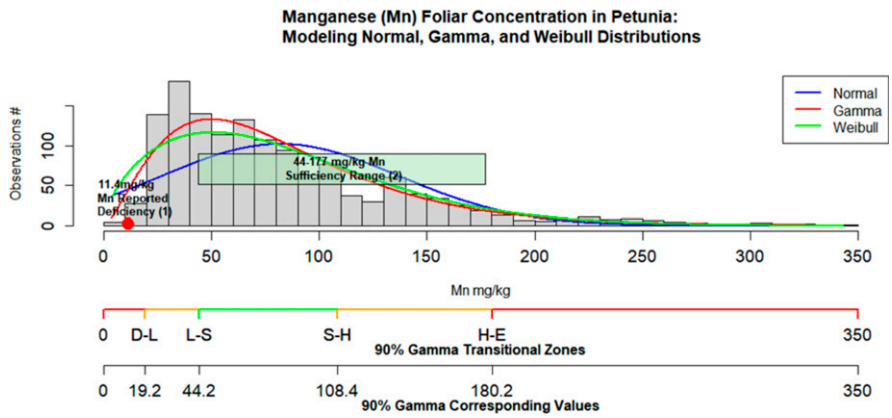

Fig. 9. Boron (B) foliar concentrations of petunia (n = 1420) modeled using normal, Gamma, and Weibull distributions. Interpretation ranges based on the Gamma distribution define the following four transition zones: deficient to low (D-L), low to sufficient (L-S), sufficient to high (S-H), and high to excessive (H-E), which correspond to B concentrations of 12.2, 19.4, 33.8, and 47.5 mg·kg^{-1}, respectively. Previously reported B sufficiency and deficiency ranges are based on studies by Pitchay et al. (2002) and Bryson and Mills (2015) and are reported for comparison.

compared with those of normal and Weibull distributions (Fig. 7). Based on this curve, a recommended sufficiency range of 76.1 to 123.0 mg·kg^{-1} Fe would decrease the current suggested Fe range of 84 to 168 mg·kg^{-1} Fe (Bryson and Mills 2015). The Fe deficiency foliar concentration of 51.2 mg·kg^{-1} Fe, which was based on the lowest 5% of the samples, was lower than 55.1 mg·kg^{-1} Fe, which was reported previously (Pitchay et al. 2002). Petunias are considered Fe-inefficient and can often experience Fe deficiency when substrate pH is high (>6.5) even if adequate Fe is supplied to the root zone (Smith et al. 2004). Currently, there are no reported values of visual Fe toxicity symptoms in petunia; however, a decrease in plant dry weight when the Fe foliar concentrations were greater than 757 mg·kg^{-1} Fe has been reported (Lee et al. 1992). This value is well above the excessive zone of >166.5 mg·kg^{-1} Fe established by our research.

Additionally, Fe was best classified by the decision tree algorithms J48 and RF, which both yielded a minimum PCC of 99.30% (Table 8). The MLP yielded a PCC range of 68.31% to 82.82%, and the SMO algorithm yielded a range of 45.55% to 63.10% (Table 8). Although all algorithms developed PCCs greater than the 20% expected with random chance, RF consistently yielded the greatest PCC across the four algorithms evaluated. An algorithm containing five elements (Table 8) would allow for a reduction of data dimensionality while accounting for the known interaction of Fe × P (Marschner 1995).

Manganese. Of the three examined models, a Gamma distribution had the lowest BIC and the best visual representation of the tails (Fig. 8). A recommended sufficiency range of 44.2 to 108.4 mg·kg^{-1} Mn narrowed and

Table 10. Percent correct classification (PCC) of boron (B) values using four machine learning algorithms (MLP, SMO, J48, and RF) with two cross-validation methods (10-fold and 66% split). Models were first run using B alone, and then they progressively incorporated additional elements until all 11 were included. PCC represents the percentage of samples correctly classified based on the B concentration, indicating the model's ability to accurately determine the B classification of deficient, low, sufficient, high, or excessive.

		Percent correct classification										
		Elements included in the model										
Algorithm	Cross-validation	1 B	2 Mn	3 S	4 Fe	5 Mg	6 K	7 Cu	8 N	9 P	10 Ca	11 Zn
MLP	10-fold	96.13	96.20	94.72	94.58	94.37	94.79	94.15	94.01	93.38	94.08	92.46
	66% Split	94.62	96.69	95.03	94.20	94.20	93.58	94.20	92.75	90.89	92.55	89.65
SMO	10-fold	85.56	77.39	77.39	77.61	78.31	77.54	77.89	76.62	76.34	76.90	77.18
	66% Split	54.45	67.49	68.53	67.91	66.25	66.87	66.87	67.49	68.53	68.74	69.57
J48	10-fold	99.86	99.86	99.86	99.86	99.86	99.86	99.86	99.86	99.86	99.86	99.86
	66% Split	99.59	99.59	99.59	99.59	99.59	99.59	99.59	99.59	99.59	99.59	99.59
RF	10-fold	100.00	100.00	100.00	100.00	100.00	99.86	99.86	99.93	99.86	99.86	99.86
	66% Split	99.59	99.59	99.79	99.38	99.38	99.59	99.38	99.59	99.79	99.79	99.38

B = boron; Ca = calcium; Cu = copper; Fe = iron; K = potassium; Mg = magnesium; MLP = multilayer perceptron; Mn = manganese; N = nitrogen; P = phosphorus; RF = random forest; S = sulfur; SMO = sequential minimal optimization; Zn = zinc.

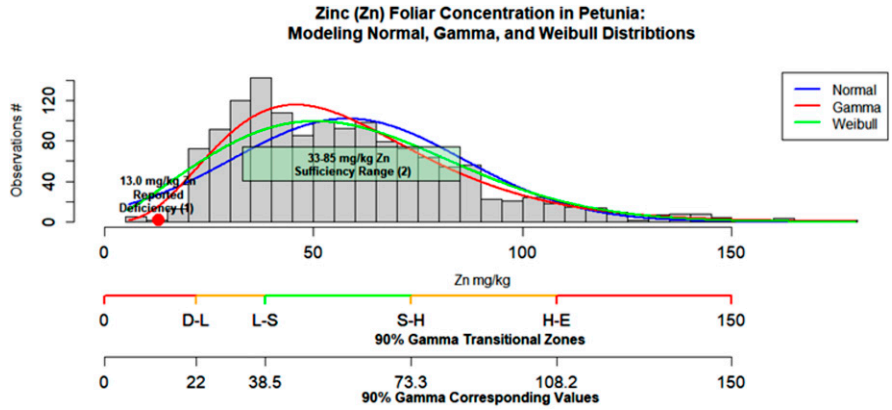

Fig. 10. Zinc (Zn) foliar concentrations of petunia (n = 1420) modeled using normal, Gamma, and Weibull distributions. Interpretation ranges based on the Gamma distribution define the following four transition zones: deficient to low (D-L), low to sufficient (L-S), sufficient to high (S-H), and high to excessive (H-E), which correspond to Zn concentrations of 22.0, 38.5, 73.3, and 108.2 mg·kg^{-1}, respectively. Previously reported Zn sufficiency and deficiency ranges are based on studies by Pitchay et al. (2002) and Bryson and Mills (2015) and are reported for comparison.

chlorophyll concentration when a Mn foliar tissue concentration of 2560 mg·kg^{-1} Mn was observed (Lee et al. 1992). Our research decreased the transition between high and excessive zones to >180.2 mg·kg^{-1} Mn.

The Mn foliar tissue concentrations were best classified by the decision tree algorithms J48 and RF. J48 yielded an average PCC of 99.89% across the two cross-validation types, whereas RF consistently yielded a PCC of 99.74% when averaged across both cross-validations (Table 9). The SMO yielded the lowest PCC (approximately 84.31%) averaged across the two cross-validation methods (Table 9). Using a J48 algorithm that contains between four and eight elements will allow for optimal PCC while still accounting for the reported antagonistic interaction of Mn × Fe (Marschner 1995) and reducing data dimensionality.

Boron. The Gamma distribution had the lowest BIC and best represented the tails and

Table 11. Percent correct classification (PCC) of zinc (Zn) values using four machine learning algorithms (MLP, SMO, J48, and RF) with two cross-validation methods (10-fold and 66% split). Models were first run using Zn alone, and then they progressively incorporated additional elements until all 11 were included. PCC represents the percentage of samples correctly classified based on the Zn concentration, indicating the model's ability to accurately determine the Zn classification of deficient, low, sufficient, high, or excessive.

		Percent correct classification										
		Elements included in the model										
Algorithm	Cross-validation	1 Zn	2 Cu	3 N	4 S	5 Mn	6 Ca	7 Mg	8 P	9 K	10 B	11 Fe
MLP	10-fold	92.39	94.15	94.01	94.08	94.15	88.94	93.52	88.29	88.10	87.96	91.06
	66% Split	87.37	93.58	88.41	89.65	87.37	87.78	86.13	86.96	86.13	85.92	85.92
SMO	10-fold	86.13	86.06	84.51	84.94	85.21	83.87	83.52	83.24	82.32	82.68	82.68
	66% Split	84.89	83.02	81.57	82.40	82.40	79.92	78.09	79.71	77.85	78.88	78.47
J48	10-fold	99.86	99.86	99.86	99.86	99.86	99.86	99.86	99.86	99.86	99.86	99.86
	66% Split	99.17	99.17	99.17	99.17	99.17	99.17	99.17	99.17	99.17	99.17	99.17
RF	10-fold	99.72	99.72	99.65	99.65	99.65	99.58	99.58	99.58	99.65	99.51	99.30
	66% Split	99.59	99.59	99.59	99.59	99.59	99.59	99.59	99.59	99.59	99.38	99.38

B = boron; Ca = calcium; Cu = copper; Fe = iron; K = potassium; Mg = magnesium; MLP = multilayer perceptron; Mn = manganese; N = nitrogen; P = phosphorus; RF = random forest; S = sulfur; SMO = sequential minimal optimization; Zn = zinc.

lowered the current sufficiency range of 44 to 177 mg·kg^{-1} Mn suggested by Bryson and Mills (2015). Additionally, the deficiency threshold of 19.2 mg·kg^{-1} Mn encompassed the critical value of 11.3 mg·kg^{-1} Mn reported previously (Pitchay et al. 2002). Although there are no reported visual Mn toxicity foliar values for petunia, a decrease in the leaf center compared with the other two examined distributions (Fig. 9). A recommended sufficiency range of 19.4 to 33.8 mg·kg^{-1} B narrowed the current recommendation of 18 to 43 mg·kg^{-1} B (Bryson and Mills 2015). A deficiency range of <12.2 mg·kg^{-1} B encompassed the deficiency value of 10.3 mg·kg^{-1} B previously reported (Pitchay et al. 2002). Lee et al. (1992) reported leaf edge burn when B foliar concentrations exceeded 651 mg·kg^{-1} B and reduced flower formation when B foliar tissue concentrations exceeded 1051 mg·kg^{-1} B. Our research established the transition between high and excessive zones as >47.5 mg·kg^{-1} B.

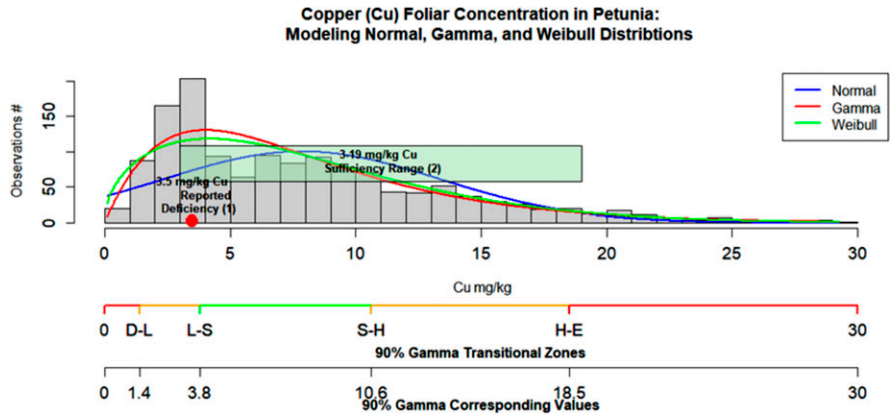

Fig. 11. Copper (Cu) foliar concentrations of petunia (n = 1420) modeled using normal, Gamma, and Weibull distributions. Interpretation ranges based on the Gamma distribution define the following four transition zones: deficient to low (D-L), low to sufficient (L-S), sufficient to high (S-H), and high to excessive (H-E), which correspond to Cu concentrations of 1.4, 3.8, 10.6, and 18.5 mg·kg^{-1}, respectively. Previously reported Cu sufficiency and deficiency ranges are based on studies by Pitchay et al. (2002) and Bryson and Mills (2015) and are reported for comparison.

The B foliar tissue concentrations were best classified by the decision tree algorithms J48 and RF, which both yielded a minimum PCC of 99.38%; however, the RF consistently had less variability in PCC (0.62%) across both cross-validations (Table 10). The SMO yielded the lowest PCC (approximately 72.39%) averaged across the two cross-validation methods (Table 10). Using an RF algorithm that contains 10 elements will allow for optimal PCC while still accounting for reported antagonistic

Table 12. Percent correct classification (PCC) of copper (Cu) values using four machine learning algorithms (MLP, SMO, J48, and RF) with two cross-validation methods (10-fold and 66% split). Models were first run using Cu alone, and then they progressively incorporated additional elements until all 11 were included. PCC represents the percentage of samples correctly classified based on the Cu concentration, indicating the model's ability to accurately determine the Cu classification of deficient, low, sufficient, high, or excessive.

Percent correct classification												
		Elements included in the model										
Algorithm	Cross-validation	1 Cu	2 N	3 Zn	4 Mn	5 S	6 Ca	7 Fe	8 P	9 K	10 Mg	11 B
MLP	10-fold	86.20	87.46	92.75	90.85	90.49	90.07	89.86	90.28	87.96	88.66	86.90
	66% Split	89.23	89.65	90.68	91.51	90.89	92.13	87.58	90.06	89.03	87.37	87.99
SMO	10-fold	41.27	52.61	55.63	59.15	60.35	59.79	59.79	61.48	65.45	65.56	65.49
	66% Split	40.79	48.48	44.86	54.45	57.35	58.39	57.97	60.87	63.56	66.25	67.08
J48	10-fold	99.44	99.44	99.44	99.44	99.44	99.44	99.44	99.44	99.44	99.44	99.44
	66% Split	99.59	99.59	99.59	99.59	99.59	99.59	99.59	99.59	99.59	99.59	95.59
RF	10-fold	99.58	99.08	99.51	99.58	99.58	99.44	99.44	99.44	99.51	99.44	99.37
	66% Split	99.79	99.76	99.59	99.17	99.59	99.79	99.60	99.79	99.60	99.59	98.59

B = boron; Ca = calcium; Cu = copper; Fe = iron; K = potassium; Mg = magnesium; MLP = multilayer perceptron; Mn = manganese; N = nitrogen; P = phosphorus; RF = random forest; S = sulfur; SMO = sequential minimal optimization; Zn = zinc.

interactions of B × Ca, B × K, and B × N (Marschner 1995).

Zinc. A Gamma distribution was used for Zn foliar concentrations and best represented the middle and tails of the observations across all three models; additionally, it yielded the lowest BIC (Fig. 10). Based on this distribution, a recommended sufficiency range of 38.5 to 73.3 mg·kg^{-1} Zn would narrow the current Zn sufficiency range of 33 to 85 mg·kg^{-1} Zn suggested by Bryson and Mills (2015). The Zn deficiency foliar concentration of 22.0 mg·kg^{-1} Zn, based on the lowest 5% of the samples, included the 13.0 mg·kg^{-1} Zn previously reported (Pitchay et al. 2002). Lee et al. (1992) reported decreased plant dry weight and flower development when Zn foliar concentrations of 1630 mg·kg^{-1} Zn were observed. Our research decreases the transition between high and excessive zones to >108.2 mg·kg^{-1} Cu.

The Zn foliar tissue concentrations were best classified by the decision tree algorithms J48 and RF. Both yielded a minimum PCC of 99.17%; however, RF yielded a more consistent PCC across both cross-validations because additional elements were included in the model (Table 11). The SMO yielded the lowest PCC (approximately 82.37%) averaged across the two cross-validation methods (Table 11).

Copper. The Gamma distribution for the Cu concentration had the lowest BIC and best represented the tails and center compared with the other two examined distributions. A recommended sufficiency range of 3.8 to 10.6 mg·kg^{-1} Cu narrowed the current recommendations of 3 to 19 mg·kg^{-1} (Bryson and Mills 2015) (Fig. 11). A deficiency range of <1.4 mg·kg^{-1} Cu is below the reported deficiency value of 3.5 mg·kg^{-1} (Pitchay et al. 2002). Pitchay et al. (2002) reported a Cu deficiency value in asymptomatic plants that did not receive Cu fertility after 8 weeks of growth, which may contribute to the difference in values. This discrepancy merits additional investigation to confirm the critical Cu deficiency concentration of petunia. The Cu toxicity symptoms included yellowing, interveinal chlorosis, and decreased plant dry weight, which were observed in petunia with a foliar concentration of 149 mg·kg^{-1} Cu (Lee et al. 1992). This current research lowered the transition between high and excessive zones to >18.5 mg·kg^{-1} Cu.

Additionally, Cu was best classified by the decision tree algorithms J48 and RF, which both yielded a minimum PCC of 99.08% (Table 12). The SMO algorithm yielded the lowest PCC range of 40.79% to 67.08% (Table 12). While all algorithms yielded greater than the 20% expected with random chance, RF consistently yielded the greatest PCC across all algorithm types. An algorithm containing seven elements (Table 12) would allow for a reduction of data dimensionality while accounting for the known interaction of Cu × Fe (Marschner 1995).

The creation of five nutrient interpretation ranges is a critical step to providing data-driven diagnostics. Previous work highlighted sufficiency ranges or critical values of small datasets; however, because of the economic value of petunias, a more refined system was needed. This study used a larger dataset and fit appropriate distribution models using an SRA method to provide more defined ranges beyond the sufficiency zone to enable the identification of samples that are deficient, low, sufficient, high, or excessive. Additionally, by using a standard commercial laboratory analysis, ML can accurately provide diagnostics to a wider range of users. Although all examined algorithms can be used for the classification of petunia foliar nutrient concentrations, their architectures greatly impact the level of accuracy. The two decision trees that were evaluated (J48 and RF) routinely performed better than MLP and SMO. This is likely because of the decision trees subgrouping architecture compared with SMO, which is intended for binary classification using a hyperplane to separate data.

Conclusion

There is a continued need for refined leaf tissue nutrient standards for horticultural crops. Previously reported deficiency and sufficiency ranges, which used included a limited number of samples, offered an initial baseline but resulted in the need for increased accuracy. A more refined system was needed to diagnose nutritional problems in petunia and determine appropriate corrective procedures. This study used a larger dataset (n = 1420) compared with those previously used (n = 25 to 30) and fit appropriate distribution models using an SRA method to provide more defined ranges beyond the sufficiency zone and also identify zones of deficient, low, sufficient, high, or excessive concentrations. This work also established that ML algorithms can accurately classify leaf tissue samples and account for interactions among elements. Decision trees (J48 and RF) routinely yielded a greater PCC compared with those yielded by MLP and SMO for all examined elements, likely because of the algorithm architecture. Although additional work is needed to confirm this method for other crop species, this research demonstrated the capabilities of ML for crop nutrient diagnostics using traditional tissue analysis methods and the ability to reduce errors in the interpretation of leaf tissue analysis reports.

References Cited

Alem P, Thomas PA, van Iersel MW. 2015. Substrate water content and fertilizer rate affect growth and flowering of potted petunia. HortScience. 50(4):582–589. https://doi.org/10.21273/HORTSCI.50.4.582.

Amatya S, Karkee M, Gongal A, Zhang Q, Whiting MD. 2016. Detection of cherry tree branches with full foliage in planar architecture for automated sweet-cherry harvesting. Biosystems Engineering. 146:3–15. https://doi.org/10.1016/j.biosystemseng.2015.10.003.

Beaufils ER. 1973. Diagnosis and recommendation integrated system (DRIS). Soil Sci. Bull. 1. Univ. of Natal, Pietermaritzburg, South Africa.

Boldt JK. 2018. Short-term reductions in irradiance and temperature minimally affect growth and development of five floriculture species. HortScience. 53:33–37. https://doi.org/10.21273/HORTSCI10289-17.

Boldt JK, Altland JE. 2019. Timing of a short-term reduction in temperature and irradiance affects growth and flowering of four annual

bedding plants. Horticulturae. 5(1):15. https://doi.org/10.3390/horticulturae5010015.
Boldt JK, Altland JE. 2021. Petunia (*Petunia ×hybrida*) cultivars vary in silicon accumulation and distribution. HortScience. 56:305–312. https://doi.org/10.21273/HORTSCI15486-20.
Boldt JK, Locke JC, Altland JE. 2018. Silicon accumulation and distribution in petunia and sunflower grown in a rice hull-amended substrate. HortScience. 53:698–703. https://doi.org/10.21273/HORTSCI12325-17.
Bryson GM, Mills HA. 2015. Plant analysis handbook IV. Micro-Macro Publishing, Athens, GA USA.
Cera A, Montserrat Martí G, Drenovsky RE, Ourry A, Brunel Muguet S, Palacio S. 2022. Gypsum endemics accumulate excess nutrients in leaves as a potential constitutive strategy to grow in grazed extreme soils. Physiol Plant. 174(4):e13738. https://doi.org/10.1111/ppl.13738.
Chung C, Huang K, Chen S, Lai M, Chen Y, Kuo Y. 2016. Detecting Bakanae disease in rice seedlings by machine vision. Comput Electron Agric. 121:404–411. https://doi.org/10.1016/j.compag.2016.01.008.
Ebrahimi MA, Khoshtaghaza MH, Minaei S, Jamshidi B. 2017. Vision-based pest detection based on SVM classification method. Comput Electron Agric. 137:52–58. https://doi.org/10.1016/j.compag.2017.03.016.
Ennaji O, Vergutz L, El Allali A. 2023. Machine learning in nutrient management: A review. Artificial Intelligence in Agriculture. 9:1–11. https://doi.org/10.1016/j.aiia.2023.06.001.
Eibe F, Hall MA, Witten IH. 2016. The WEKA workbench. Online appendix for data mining: practical machine learning tools and techniques. Morgan Kaufmann Publishers, San Francisco, CA, USA.
Griffel LM, Delparte D, Whitworth J, Bodily P, Hartley D. 2023. 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 Ag Tech. 3:100101. https://doi.org/10.1016/j.atech.2022.100101.
Guggenmoos-Holzmann I. 1996. The meaning of kappa: Probabilistic concepts of reliability and validity revisited. J Clin Epidemiol. 49(7):775–782. https://doi.org/10.1016/0895-4356(96)00011-x.
Guo W, Nazim H, Liang Z, Yang D. 2016. Magnesium deficiency in plants: An urgent problem. Crop J. 4(2):83–91. https://doi.org/10.1016/j.cj.2015.11.003.
Handreck KA, Black ND. 2002. Growing media for ornamental plants and turf. University of New South Wales Press, Sydney, Australia.
Henry JB. 2017. Beneficial and adverse effects of low phosphorus fertilization of floriculture species (MS Thesis). North Carolina State University, Raleigh, NC, USA.
Henry JB, Whipker BE, McCall I. 2016. Phosphorus restriction as an alternative method of growth control for petunia. 43rd Proceedings of the Annual Plant Growth Regulation Society of America Conference, Raleigh, NC, USA, 16–21 Jul 2016. 43:52–57.
Keerthi SS, Shevade SK, Bhattacharyya C, Murthy KRK. 2001. Improvements to Platt's SMO algorithm for SVM classifier design. Neural Comput. 13(3):637–649. https://doi.org/10.1162/089976601300014493.
Lee CW, Pak C, Choi J, Self JR. 1992. Induced micronutrient toxicity in *Petunia hybrida*. J Plant Nutr. 15(3):327–339. https://doi.org/10.1080/01904169209364322.
Li D, Miao Y, Ransom CJ, Bean GM, Kitchen NR, Fernández FG, Sawyer JE, Camberato JJ, Carter PR, Ferguson RB, Franzen DW, Laboski CAM, Nafziger ED, Shanahan JF. 2022. Corn nitrogen nutrition index prediction improved by integrating genetic, environmental, and management factors with active canopy sensing using machine learning. Remote Sensing. 14(2):394. https://doi.org/10.3390/rs14020394.
Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D. 2018. Machine learning in agriculture: A review. Sensors. 18(8):2674. https://doi.org/10.3390/s18082674.
Marschner H. 1995. Mineral nutrition of higher plants (2nd ed). Academic Press, New York, NY, USA.
Mhango JK, Hartley W, Harris WE, Monaghan JM. 2021. Comparison of potato (*Solanum tuberosum* L.) tuber size distribution fitting methods and evaluation of the relationship between soil properties and estimated distribution parameters. J Agric Sci. 159(9-10):643–657. https://doi.org/10.1017/S0021859621000952.
Mingers J. 1989. An empirical comparison of selection measures for decision-tree induction. Mach Learn. 3(4):319–342. https://doi.org/10.1007/BF00116837.
Pantazi X, Moshou D, Bravo C. 2016. Active learning system for weed species recognition based on hyperspectral sensing. Biosystems Engineering. 146:193–202. https://doi.org/10.1016/j.biosystemseng.2016.01.014.
Pantazi XE, Tamouridou AA, Alexandridis TK, Lagopodi AL, Kashefi J, Moshou D. 2017. Evaluation of hierarchical self-organising maps for weed mapping using UAS multispectral imagery. Comput Electron Agric. 139:224–230. https://doi.org/10.1016/j.compag.2017.05.026.
Parent LE, Dafir M. 1992. A theoretical concept of compositional nutrient diagnosis. J Am Soc Hortic Sci. 117(2):239–242. https://doi.org/10.21273/JASHS.117.2.239.
Pitchay DS, Gibson JL, Nelson PV, Walls FR, Whipker BE, Cleveland B. 2002. Petunia foliar analysis standards and deficiency symptoms of 11 nutrients (PhD Diss.). North Carolina State University, Raleigh, NC, USA.
Ramos PJ, Prieto FA, Montoya EC, Oliveros CE. 2017. Automatic fruit count on coffee branches using computer vision. Comput. Electron. Agric. 137:9–22. https://doi.org/10.1016/j.compag.2017.03.010.
Rao IM, Sharp RE, Boyer JS. 1987. Leaf magnesium alters photosynthetic response to low water potentials in sunflower. Plant Physiol. 84(4):1214–1219. https://doi.org/10.1104/pp.84.4.1214.
Reuter D, Robinson JB. 1997. Plant analysis: An interpretation manual. CSIRO Publishing, Victoria, Australia. https://doi.org/10.1071/9780643101265.
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. https://doi.org/10.1016/j.compag.2020.105860.
Slaton NA, Drescher GL, Parvej MR, Roberts TL. 2021. Dynamic critical potassium concentrations in soybean leaves and petioles for monitoring potassium nutrition. Agron J. 113(6):5472–5482. https://doi.org/10.1002/agj2.20819.
Smith BR, Fisher PR, Argo WR. 2004. Growth and pigment content of container-grown impatiens and petunia in relation to root substrate pH and applied micronutrient concentration. HortScience. 39(6):1421–1425. https://doi.org/10.21273/HORTSCI.39.6.1421.
Soltanpour PN, Malakouti MJ, Ronaghi A. 1995. Comparison of diagnosis and recommendation integrated system and nutrient sufficiency range for corn. Soil Sci Soc Am J. 59(1):133–139. https://doi.org/10.2136/sssaj1995.03615995005900010021x.
Sumner ME. 1990. Advances in the use and application of plant analysis. Commun Soil Sci Plant Anal. 21(13–16):1409–1430. https://doi.org/10.1080/00103629009368313.
van Iersel MW, Beverly RB, Thomas PA, Latimer JG, Mills HA. 1998. Fertilizer effects on the growth of impatiens, petunia, salvia, and vinca plug seedlings. HortScience. 33(4):678–682. https://doi.org/10.21273/HORTSCI.33.4.678.
Weibull W. 1951. A statistical distribution function of wide applicability. J Appl Mech. 18(3):293–297. https://doi.org/10.1115/1.4010337.
Wickham H. 2011. ggplot2. WIREs Computational Stats. 3(2):180–185. 10.18637/jss.v040.i01.
Witten IH, Frank E. 2005. Data mining: Practical machine learning tools and techniques (2nd ed). Morgan Kaufmann, San Francisco, CA, USA.