SAINS TANAH – Journal of Soil Science and Agroclimatology

Journal homepage: http://jurnal.uns.ac.id/tanah

Modeling the activity ratio of soil potassium using machine learning approach

Desi Nadalia[1,2], Arief Hartono[2]*, Heru Bagus Pulunggono[2], Bambang Hendro Trisasongko[2], Widiatmaka[2], Muhammad Fuady Emzir[3]

[1] Doctoral Program of Soil Science, Graduate School, IPB University, Bogor, 16680, Indonesia
[2] Department of Soil Science and Land Resources, Faculty of Agriculture, IPB University, Bogor, 16680, Indonesia
[3] Department of Control and Instrumentation Engineering, King Fahd University of Petroleum and Minerals, Saudi Arabia

ARTICLE INFO

Keywords:
Cubist
MLR
Predictor
RF
SVM

Article history
Submitted: 2025-03-05
Revised: 2025-07-28
Accepted: 2025-08-08
Available online: 2025-09-27
Published regularly: December 2025

* *Corresponding Author*
Email address:
hartono@apps.ipb.ac.id

ABSTRACT

The potassium (K) Quantity-Intensity (Q-I) relationship results in important parameters, including the activity ratio of potassium at equilibrium (ARe^K), which indicates potassium availability in soil. Experiments to observe soil Q-I K relationship parameters are often complex, time-consuming, and do not include environmental variables. This research aims to model ARe^K using a machine learning (ML) approach. ML models applied are Random Forest (RF), Cubist, and Support Vector Machine (SVM) as the primary approaches, with Multiple Linear Regression (MLR) serving as a baseline. The dataset was derived from sixty-one observation points in Brebes, Central Java. The predictors were pH, organic carbon, clay, cation exchange capacity (CEC), exchangeable cations (Exc-Ca, Mg, K, Na), water soluble K, available K, K saturation, potential K, non-exchangeable K (NE-K), elevation, and slope. The response variable was the ARe^K. Variable selection was performed using Pearson correlation to eliminate highly correlated predictors and reduce multicollinearity. Exactly 75% of the data was utilized as the training set and 25% as the test set. Three metrics, i.e., MAE, RMSE, and R^2, were used in model evaluation. The results showed that the Cubist model could predict ARe^K with high accuracy (R^2=0.9437) and low RMSE (0.5701) and MAE (0.3514). Based on the Cubist model, Exc-K, Exc-Mg, CEC, and Exc-Ca were the most important variables for predicting ARe^K. This model can be employed to support site-specific fertilizer recommendation strategies. To improve the performance of the model, it is necessary to add other predictor variables (e.g., soil physical properties, clay minerals, rainfall, temperature and soil moisture).

How to Cite: Nadalia, D., Hartono, A., Pulunggono, H. B., Trisasongko, B. H., Widiatmaka, & Emzir, M. F. (2025). Modeling the activity ratio of soil potassium using machine learning approach. Sains Tanah Journal of Soil Science and Agroclimatology, 22(2), 243-254. https://doi.org/10.20961/stjssa.v22i2.100102

1. INTRODUCTION

The dynamics of Potassium (K) in soils are affected by its interactions with calcium (Ca) and magnesium (Mg). Understanding of soil K dynamics can be seen from the thermodynamic Quantity-Intensity (Q-I) K relationship (Beckett, 1964). The Q-I K relationship results in important parameters, including the activity ratio of K at equilibrium (ARe^K). ARe^K is the activity ratio of K in soil solution at equilibrium (ΔK=0). The activity ratio of K is calculated by dividing the activity of K by the square root of the combined activities of Ca and Mg. This value reflects the intensity of K availability by considering the influence of Ca and Mg, which can compete with K in the soil solution. ARe^K tends to rise with higher levels of K concentration in the soil solution (Ajiboye et al., 2015; Al-Hamandi et al., 2019; Bilias & Barbayiannis, 2019; Lalitha & Dhakshinamoorthy, 2015; Lumbanraja et al., 2020; Zhu et al., 2020). Recent research by Nadalia et al. (2024) observed variations in ARe^K in Brebes farmland soils. The findings suggest that K is predominantly adsorbed at edge positions of 2:1 clay mineral, which contributes to efficient K release during depletion. However, farmers in the Brebes regency study area typically apply synthetic inorganic K fertilizer at 150 kg/ha K_2O, exceeding the recommended amount for shallot plants (Balittanah, 2021). The excessive use of fertilizer has an impact on the accumulation of K residues in the soil. This is supported by the results of Muliana et al. (2018), who found that K extracted using 25% HCl in several shallot cultivation areas averaged of 55 mg/100 g K_2O, a level classified as high. Thus, assessing K nutrient availability

STJSSA, p-ISSN 1412-3606 e-ISSN 2356-1424
http://dx.doi.org/10.20961/stjssa.v22i2.100102

in these intensively managed shallot cultivation soils is crucial for optimizing fertilizer application and enhancing nutrient management efficiency.

Experiments to observe soil Q-I K relationship parameters are often complex, costly, and time-consuming. Additionally, Q-I K relationship parameters have not yet incorporated environmental variables. Controlling environment, such as climate and topography data, brings a better understanding on the relationship between environmental conditions as well as soil nutrients, revealing gradual changes through both linear and nonlinear associations. As a result, there is an urgent demand for faster and more accurate methods to predict soil nutrients, minimize nutrient losses, and improve the efficiency of fertilizer management (Song et al., 2018).

Survey of the literature, however, indicates that Q-I studies are yet to be intertwined with machine learning (ML) as a modern data extraction. Currently, linear regression models, particularly Multiple Linear Regression (MLR), are frequently utilized to examine, interpret, and forecast various soil science data (Rossiter, 2018; Sharma et al., 2015). The use of ML in soil science investigations has also been progressing (Li et al., 2022; Padarian et al., 2020; Rossiter, 2018; Wang et al., 2023). Geostatistical methods for spatial interpolation are widely used in nutrient assessment research. Despite extensive testing, these approaches often struggle to capture the relationship between nutrients and their environmental context. This limitation has encouraged the development of ML-based methods. However, they have mainly focused on nutrients other than K. Machine learning models can effectively process and identify patterns and handle multicollinearity, heteroscedasticity, and nonlinear relationship problems (Chen et al., 2019; Feng et al., 2019; Padarian et al., 2020). Typically, ML models have at least one parameter requiring adjustment, which should be calibrated from predictive estimates of errors. Random Forest (RF) model works effectively with minimal tuning. Probst et al. (2019), for instance, proved that RF tuning and improved regression tree models can optimize predictive results. Adjustments, however, should be cautiously applied as the models are prone to overfitting (Ließ et al., 2016). Li et al. (2014) research presented a model for evaluating soil nutrients using support vector machines (SVM) and MLR. In the study, independent variables were the content of the organic matter, total nitrogen, available potassium, and available phosphorus. The study's results indicated that the SVM algorithm produced an average accuracy of 77.87%, while the MLR algorithm reached 83.00%. These results provide information that SVM can effectively predict nutrient content in the soil. In addition, the Cubist model for soil organic carbon estimation showed the best performance compared to a simpler model, namely multivariate linear regression, and was easier to interpret than the artificial neural network (ANN) model (John et al., 2020; John et al., 2021). Covariates or environmental variables serve as predictors in machine learning models. These covariates can explain the underlying physical and chemical processes that drive spatial variability in soil characteristics. The most common covariates are soil properties, average annual rainfall (long term), temperature, remote sensing imagery (e.g., vegetation index obtained from satellite imagery), elevation, and terrain-related metrics (e.g., slope, local curvature, topographic wetness index) (Gomes et al., 2019; Szatmári et al., 2019).

This study is the first to apply ML algorithms (RF, Cubist, and SVM) to model ARe^K based on soil properties and environmental variables. Although past studies have explored both linear regression and ML applications for general nutrient prediction, none has specifically modeled the activity ratio of K (ARe^K) using ML algorithms. The models also generate variable importance that help identity key factors influencing ARe^K. Soil properties and environmental factors that potentially influence ARe^K include exchangeable K, Ca, Mg, Na, cation exchange capacity (CEC), clay content, organic carbon, soil pH, non-exchangeable K, elevation, and slope. The resulting variables importance can subsequently be used to predict ARe^K through ML models. Therefore, this research aims to model ARe^K using a ML approach and assess the performance of ML-based approaches in predicting ARe^K within cultivated lands of Brebes, Central Java. The ML approach can help assess soil nutrients, including ARe^K, which can be employed to support site-specific fertilizer recommendation strategies to optimize K management.

2. MATERIAL AND METHODS

2.1. Soil characteristic analysis

This study was conducted on agricultural land in Brebes, Central Java, Indonesia (Fig. 1). Soil orders at the study site include Inceptisols, Alfisols, and Andisols (according to USDA soil taxonomy classification) (ICALRD, 2017). Soil sampling was conducted using a random composite method at a 0–30 cm depth. Soil sampling was carried out at 61 distinct sites, producing 61 composite samples. The sampled soils represent potential areas for shallot cultivation in Brebes Regency. The selected sampling sites also reflect the three dominant soil types in the region: Inceptisols, Alfisols, and Andisols. Most of the sampling points were in intensively managed agricultural lands, especially shallot cultivation, which is one of the region's horticultural commodities. Samples were air-dried, crushed, and sieved (<2 mm and 2 mm) before being subjected to laboratory analysis. The soil characterization involved measurements of selected chemical properties and soil texture. Analysis of selected soil properties shown in Table 1.

Saturation percentage of potassium (K saturation) was determined by dividing Exc-K by CEC and multiplying by 100%. Available potassium (Avail-K) was determined by combining Water-K and Exc-K measurements. The difference between the HNO_3-extracted potassium and Avail-K was defined as non-exchangeable potassium (NE-K). Beckett's method determined the potassium activity ratio at equilibrium (ARe^K) (Beckett, 1964).

2.2. Modeling approaches

This study applied four machine learning models: Random Forest (RF), Cubist, Support Vector Machine (SVM), and Multiple Linear Regression (MLR). R software, with the controlling package being the 'caret' (Kuhn et al., 2020), was used for the entire process of modeling. A brief overview of each ML method used in the current research is as follows:

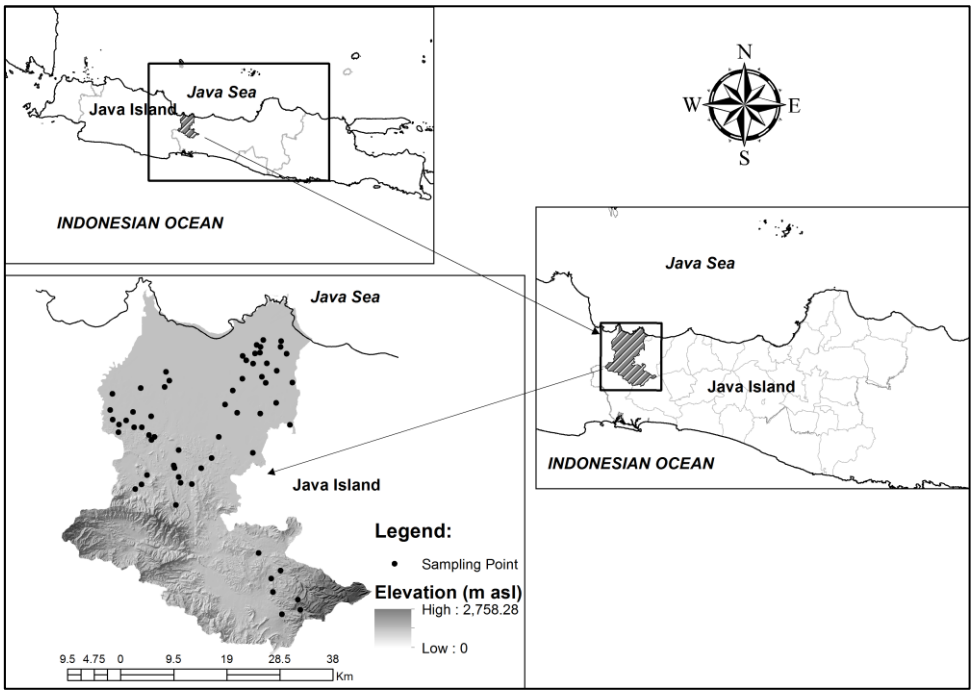

Figure 1. The geographical location of the study area in Brebes Regency, Central Java

Table 1. Analysis of selected soil properties

No.	Soil Properties	Methods/ Measurements	References
1	pH H_2O (1:5)	pH-meter	Thomas (1996)
2	Organic-C (org-C)	Walkley and Black	Nelson and Sommers (1996)
3	Potential K	HCl 25%	Eviati et al. (2023)
4	Cation Exchange Cation (CEC)	NH_4OAc 1 N at pH 7	Sumner and Miller (1996)
5	Exchangeable K (Exc-K) and Na (Exc-Na)	Flamephotometer	Helmke and Sparks (1996)
6	Exchangeable Ca (Exc-Ca) and Mg (Exc-Mg)	Atomic Absorption Spectrophotometry (AAS)	Suarez (1996)
7	Water-soluble K (Water-K)	Soil-water at a 1:5 ratio	Soil and Plant Analysis Council Inc. (2000)
8	K-HNO_3	Boiling with HNO_3	Helmke and Sparks (1996)
9	Soil texture (sand, silt, and clay fractions)	Pipette	van Reeuwijk (2002)

1) Random Forest (RF)

RF is a supervised and integrated learning model that aggregates results from single decision trees. RF is an ensemble method that improves prediction efficiency by averaging multiple decision trees and is widely used for classification and regression purposes (Breiman, 2001). Two hyperparameters can be modified in the RF model to adjust the model's complexity, namely the number of trees (or iterations) (*ntree*) and *mtry*. In this study, the *mtry* value was tested repeatedly from one to the total number of variables, while the *ntree* was set at 500. The model performance of each hyperparameter tuning combination was achieved using the combination of grid search and the K-fold cross-validation method (K-fold CV) (Zhou et al., 2019).

2) Cubist

Quinlan (1992) introduced the Cubist model as an approach to regression modeling. This model is applied in R by performing hyperparameter tuning on instance (neighbors) and committee (committee) parameters. The instance and committee parameters tend to have the greatest influence on the overall performance of the Cubist model.

3) Support Vector Machine (SVM)

SVM is an effective and widely used learning model for classifying linear and nonlinear regression problems (Cortes & Vapnik, 1995; Onyelowe et al., 2022). Regression SVM with a radial basis function (RBF) kernel method is applied in this study. The radial-based kernel approach is utilized to generate optimal performance from the SVM model by discovering the best combination of parameters (Sigma and Cost/ C) for the ARe^K training data set.

4) Multiple Linear Regression (MLR)

MLR is one of the algorithms used in machine learning that regresses the response variable against the predictor variables. Equation 1 presents a simple MLR equation.

$$y = a + \sum_{i=1}^{n} bi \times xi \pm \varepsilon i \quad \text{.....[1]}$$

with *n* indicating the total number of predictor variables; *y* indicates the dependent variable (ARe^K); *xi* indicates the predictor variables (indicator variables of several soil and environmental properties); *a* indicates the intercept (constant); *bi* indicates the regression coefficient for each independent variable; εi indicates the error (residual) in the model.

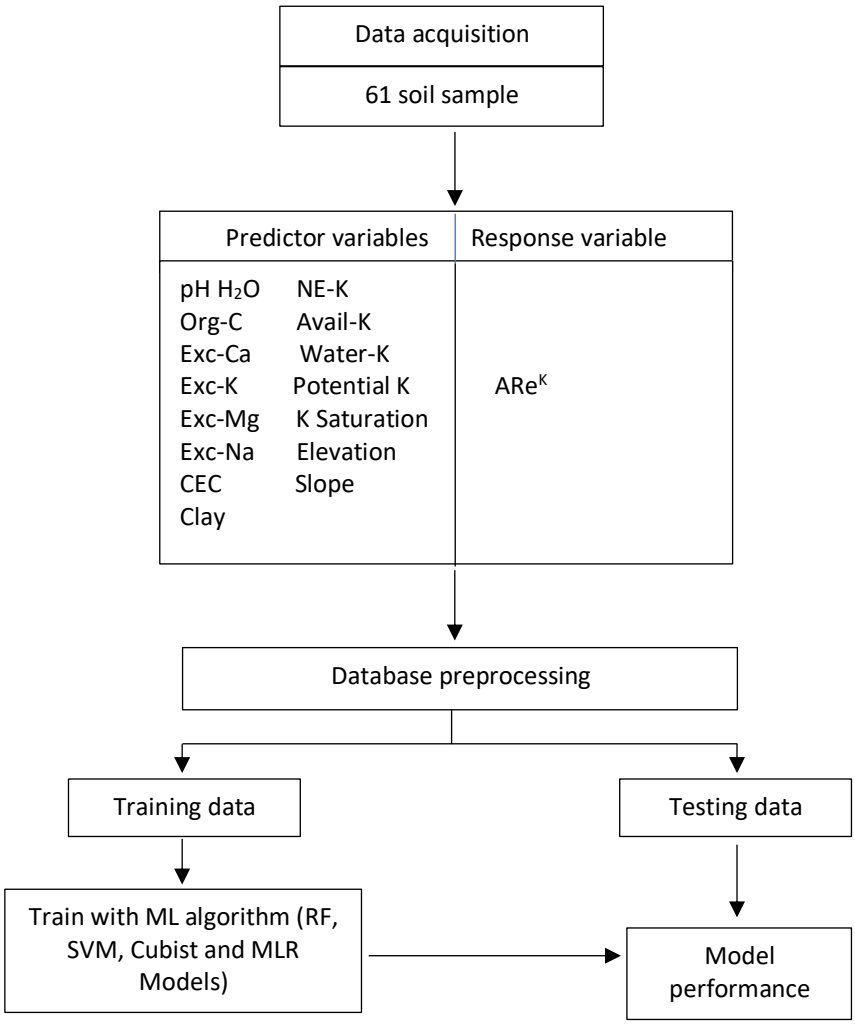

Figure 2. Methodological flowchart

K-fold cross-validation is one of the most widely used cross-validation methods in machine learning, and a value of K = 5 with five repetitions was applied in this study. The methodology for estimating the parameter involved in the dynamics of K (ARe^K) using machine learning methods is illustrated in Figure 2. Variable selection was then performed to determine the number of predictor variables used for modeling. Variable selection was performed using Pearson correlation to eliminate highly correlated predictors and reduce multicollinearity. The data were split into 75% for training and 25% for testing the model.

2.3. Model performance

Model performance is an integral part of the model design process, helping to identify the model that can provide the best predictive results. The measures used to evaluate the prediction accuracy and the level of model fit included the mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2). R^2 was used to assess the extent to which the model was able to capture variations in the observed data. The MAE explains the general absolute error between predictions and actual values, giving equal weight to each error. It is less sensitive to large errors compared to R^2 and RMSE. Meanwhile, RMSE calculates errors by squaring each one, thus giving larger errors were calculated with a higher weight proportion. The level of estimation error in the observed variable units was measured using the RMSE and MAE values. A value of zero for both metrics indicates a very good model fit and reflects strong predictive performance, as it corresponds to minimal estimation error. An accurate model prediction shows low RMSE and MAE, and an R^2 value close to 1. Model evaluation was carried out using test data. The numbers obtained from the evaluation of the test data provided an accurate assessment of how well the model performed on new data that was not involved in model training. All MAE, RMSE, and R^2 were utilized to evaluate the performance of the model as outlined by Equations 2, 3, and 4.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|ARe^K(y_i) - ARe^K(\hat{y}_i)| \quad \text{.....[2]}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[ARe^K(y_i) - ARe^K(\hat{y}_i)]^2} \quad \text{.....[3]}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}[ARe^K(y_i) - ARe^K(\hat{y}_i)]^2}{\sum_{i=1}^{n}[ARe^K(y_i) - ARe^K(\bar{y}_i)]^2} \quad \text{.....[4]}$$

where $ARe^K(y_i)$ indicates the measured response, $ARe^K(\hat{y}_i)$ represent the prediction response of ARe^K, $ARe^K(\bar{y}_i)$ represent the average value of ARe^K observation, the number of samples is denoted by *n*, and *i* indicates the order of observation (*i* = 1,2,3, . . . , *n*).

Table 2. Descriptive statistics of ARe^K

	n	Mean	Median	SD	Min	Max	1st Quartile	3rd Quartile	CV
ARe^K $(mol/L)^{1/2}$	61	0.004	0.003	0.004	0.0003	0.018	0.0014	0.005	96.71

Remarks: n = number of samples; SD = standar deviation; Min = minimum; Max = maximum; CV = Coefficient of Variation

3. RESULTS

3.1. Descriptive statistics

Descriptive statistics from ARe^K at the research locations are presented in Table 2**.** According to the coefficient of variation (CV) value, ARe^K exhibited high variability (CV > 35%). The mean and median values of ARe^K indicated that, in most locations, the activity of K in soil solutions was relatively medium or K is adsorbed at the edge sites of 2:1 clay mineral. However, the high CV score suggested a large variation between locations. This variation can be referred to as factors that influence ARe^K, such as soil mineralogy, soil properties, and environmental factors.

3.2. Variable selection

Table 3 presents the analysis of the correlation using Pearson correlation. Based on Table 3, out of 15 predictor variables, four variables exhibited high correlation (r > 0.73) with each other, namely Avail-K, Water-K, Potential K, and K Saturation. These four predictor variables were excluded from model development to minimize data redundancy and boost the model's consistency and reliability of the constructed ML models. In machine learning models, high correlation between variables can lead to information redundancy, which may affect model interpretation, although some machine learning algorithms are relatively robust to the situation.

3.3. Hyperparameters optimization results

RF, Cubist, and SVM models were applied to estimate ARe^K. The original dataset, consisting of various soil and environmental properties affecting ARe^K, was randomly split into a training and testing dataset. In this study, as much as 75% of the data was used as a randomly selected training set, and 25% as a testing set. The performance of the regression model was evaluated using the testing dataset. The amount of training data determines the performance of the ARe^K.
The ARe^K dataset, consisting of 61 observations, was randomly split into 47 for training and 14 for testing. The regression model was built using data from the training dataset to assess the parameters of the RF, Cubist, and SVM regression methods. To tune optimal hyperparameters and reduce potential bias, a five-fold cross-validation method was applied to the RF, Cubist, and SVM models.

The RF model using the "rf" method by default produced parameters (*mtry*) of 2, 6, and 11 (Table 4). The optimal model was obtained at *mtry*=6 based on the smallest RMSE. In addition to using the "rf" method, the RF model was also tested using the "ranger" method. The results of the "ranger" method showed that setting *mtry*=11 with *splitrule* = extratrees and min.node.size=5 provided the best results in terms of R^2, RMSE and MAE (Table 5). To analyze the behavior of hyperparameters in RF, namely *ntree* and *mtry*, was carried out by tuning the model using a grid search approach combined with a cross-validation method. The *mtry* parameters tested for the RF model using the "rf" method were 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 (Table 6). The optimal model was obtained at *mtry* = 2 based on the smallest RMSE. For the "ranger" method, the tuning parameters tested included *mtry* = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11; ntree = 200, 300, 400, 500 and node.size = seq(3, 20, by = 2). The best values were obtained with *mtry* = 2, *ntree* = 500, and node.size = 7 (Table 7).

Table 8 presents the results of an SVM model. Based on the smallest RMSE, the optimal model was obtained at C=1, with Sigma=0.1734. Table 9 shows the results of parameter tuning in the SVM model, with sigma values tested at 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, and C values tested at 0.2, 0.3, 0.4, and 0.5. RMSE was employed to choose the optimal model based on the lowest value. The values sigma = 0.05 and C = 0.5 were set as the final model parameters. The performance of the Cubist in the default configuration is shown in Table 10, and the performance after tuning is presented in Table 11. In Table 11, the tuned model presented a reduction in the number of committees and neighbors, indicating a simpler model structure that resulted in lower MAE and RMSE values.

3.4. Evaluating measured and predicted ARe^K using different machine learning models

Evaluation of the performance of the predictive model was carried out using common validation metrics, namely RMSE, MAE, and R^2. The validation datasets' results are provided in Table 12. Using the validation dataset, suggested ML models demonstrated their capacity to estimate ARe^K in 14 test datasets. The Cubist model exhibited the smallest RMSE and MAE values among the three machine learning models studied. The tuned RF model showed lower predictive ability in comparison with the default model, as seen from the increasing RMSE and MAE values, as well as a lower R^2 value. Meanwhile, the SVM model showed the highest MAE value compared to other machine learning models. The Cubist model was superior to the others, achieving the lowest MAE and the highest R^2 value. MLR model predictive ability was comparable, which belongs to the Cubist model.

3.5. Importance factors in the individual models

Figure 3a presents the RF model prediction using the "rf" method. The figure indicates that Exc-K is the best predictor, followed by Exc-Mg and NE-K, in explaining soil ARe^K variability. Similarly, in the RF model using the "ranger" method with the default training dataset (Fig. 3b), Exc-K remained the most reliable predictor, then Exc-Mg. For the Cubist model prediction, trained using the tuned training dataset (Fig. 3c), Exc-K and Exc-Mg were identified as the most significant predictors, followed by CEC and Exc-Ca for explaining soil ARe^K variability. The best predictors identified by the MLR model in predicting ARe^K variability were Exc-K, followed by NE-K and slope (Fig. 3d).

Table 3. Correlation analysis results

Correlation (n=61 samples)	Exc-Ca	Exc-Mg	CEC	Clay	Org-C	pH	Exc-Na	Water-K	Exc-K	NE-K	Avail-K	K Saturation	Potential K	Elevation	Slope
Exc-Ca	1.00														
Exc-Mg	-0.06	1.00													
CEC	0.50**	0.41**	1.00												
Clay	0.54**	0.51**	0.50**	1.00											
Org-C	-0.29*	-0.48**	-0.55**	-0.51**	1.00										
pH	0.67**	0.10	0.12	0.33**	-0.26*	1.00									
Exc-Na	0.07	0.39**	0.41**	0.31*	-0.23	0.13	1.00								
Water-K	0.19	0.07	0.13	0.10	-0.08	0.01	0.39**	1.00							
Exc-K	0.31*	-0.02	0.12	0.13	0.01	0.15	0.42**	0.85**	1.00						
NE-K	0.24	0.32*	0.34**	0.39**	-0.50**	0.26*	0.47**	0.38**	0.20	1.00					
Avail-K	0.31*	-0.02	0.12	0.13	0.01	0.14	0.42**	0.86**	1.00	0.21	1.00				
K Saturation	0.09	-0.21	-0.27*	-0.11	0.23	0.07	0.21	0.67**	0.87**	-0.03	0.87**	1.00			
Potential K	0.36**	0.14	0.21	0.30*	-0.16	0.28*	0.52**	0.66**	0.82**	0.37**	0.82**	0.65**	1.00		
Elevation	-0.54**	-0.49**	-0.59**	-0.63**	0.59**	-0.33**	-0.24	-0.15	-0.11	-0.54**	-0.11	0.19	-0.32*	1.00	
Slope	-0.44**	-0.42**	-0.44**	-0.56**	0.37**	-0.30*	-0.29*	-0.19	-0.19	-0.55**	-0.19	-0.02	-0.36**	0.73**	1.00
ARe^{K}	0.03	-0.07	-0.10	-0.08	0.02	0.004	0.26	0.84**	0.79**	0.21	0.80**	0.70**	0.59**	0.15	0.11

Remarks: *) p-value <0.05; **) p-value < 0.01

Table 4. Results of the Random Forest model of the "rf" method of default

Parameter (*mtry*)	RMSE	R^2	MAE
2	0.724	0.287	0.520
6	0.718	0.294	0.512
11	0.723	0.267	0.513

Table 5. Results of the Random Forest model of the "ranger" method of default

Parameter (*mtry*)	Splitrule	RMSE	R^2	MAE
2	variance	0.725	0.303	0.520
2	extratrees	0.688	0.358	0.501
6	variance	0.723	0.288	0.516
6	extratrees	0.666	0.379	0.480
11	variance	0.719	0.276	0.510
11	extratrees	0.647	0.397	0.468

Table 6. Results of the Random Forest model of the "rf" method of tuned

Parameter (*mtry*)	RMSE	R^2	MAE
1	0.7162	0.3061	0.5207
2	0.7144	0.3073	0.5129
3	0.7205	0.2881	0.5155
4	0.7175	0.2983	0.5105
5	0.7187	0.3132	0.5107
6	0.7188	0.2905	0.5106
7	0.7197	0.2793	0.5131
8	0.7187	0.2866	0.5115
9	0.7181	0.2874	0.5096
10	0.7153	0.2826	0.5091
11	0.7236	0.2715	0.5107

4. DISCUSSION

The RF, Cubist, and SVM model approaches were used to predict the activity ratio K in equilibrium (ARe^K), and the MLR model served as the baseline. The Cubist model showed the smallest average RMSE and MAE, and the greatest model fit (R^2) compared to SVM and RF. The greatest average MAE value was obtained from the SVM model, while the RF model showed the highest average RMSE among all tested models. Considering these evaluation results, the Cubist model emerged as the best-performing model in predicting ARe^K. In this study, the Cubist and RF models showed different performances in predicting response variables.

This result differs from the findings of Zhou et al. (2019), who predicted the shear strength of rockfill materials, stating that the Cubist and RF models had the same prediction performance. While some studies have shown RF to outperform Cubist under specific conditions (Drzewiecki, 2016), others have demonstrated superior accuracy of the Cubist model in predicting environmental and soil variables such as total nitrogen, organic carbon, and CEC (Ludwig et al., 2022; Wu et al., 2022).

Table 7. Results of the Random Forest model of the "ranger" method of tuned

mtry	Node.size	*ntree*	OOB_RMSE
2	7	500	0.7339
2	7	400	0.7355
5	13	400	0.7356
5	11	400	0.7358
5	9	400	0.7363
5	15	400	0.7364
8	5	500	0.7365
8	9	500	0.7366
8	7	500	0.7367
8	7	500	0.7369

Table 8. Results of the SVM model of default

C	RMSE	R^2	MAE
0.25	0.7135	0.2697	0.5179
0.50	0.7002	0.2646	0.5132
1.00	0.6916	0.2757	0.5078

The superior performance of the Cubist model in this study is likely attributed to its combined modeling approach, which integrates rule-based decision trees with linear regression models applied at the terminal segments of the tree structure. This design allows the Cubist model to effectively capture both linear and piecewise-linear relationships, which is particularly advantageous for soil datasets that often exhibit local linear patterns within broader variability. Additionally, the relatively small sample size (n=61) used in this study may have favored Cubist, as it tends to generalize better than more complex models such as RF or SVM in data-limited conditions. This performance is consistent with findings from other studies reporting the advantage of Cubist modeling in soil properties estimation using relatively small and linear datasets (John et al., 2020; Ludwig et al., 2022).

The MLR model performed comparably to Cubist model in this study, suggesting that a simpler and more interpretable model is sufficient to capture most of the variability in ARe^K under predominantly linear conditions. Thus, in context with limited data and clear linear structure, MLR may serve as an efficient and robust alternative. However, ML approaches such as Cubist remain valuable in broader applications involving nonlinear dynamics or complex interactions between predictors and response variables, especially in areas with spatial and soil-related variability. Furthermore, the Cubist model produces outputs that are more interpretable compared to models with less algorithmic transparency, such as SVM, thus offering practical advantages for decision-making in agricultural land management. These results confirm that the Cubist model not only outperforms RF, SVM, and MLR in this specific context, but also stands out as the most suitable modeling approach for predicting ARe^K in relatively small and heterogeneous soil datasets.

Table 9. Results of the SVM model of tuned

Sigma	C	RMSE	R^2	MAE
0.05	0.2	0.7066	0.3714	0.5166
0.05	0.3	0.6864	0.3763	0.5000
0.05	0.4	0.6720	0.3822	0.4901
0.05	0.5	0.6619	0.3855	0.4843
0.06	0.2	0.7077	0.3530	0.5156
0.06	0.3	0.6895	0.3553	0.5015
0.06	0.4	0.6785	0.3588	0.4943
0.06	0.5	0.6714	0.3585	0.4907
0.07	0.2	0.7092	0.3372	0.5160
0.07	0.3	0.6921	0.3381	0.5028
0.07	0.4	0.6850	0.3380	0.4992
0.07	0.5	0.6791	0.3377	0.4965
0.08	0.2	0.7108	0.3234	0.5168
0.08	0.3	0.6950	0.3235	0.5046
0.08	0.4	0.6899	0.3219	0.5027
0.08	0.5	0.6849	0.3212	0.5012
0.09	0.2	0.7122	0.3119	0.5177
0.09	0.3	0.6976	0.3114	0.5063
0.09	0.4	0.6937	0.3089	0.5056
0.09	0.5	0.6893	0.3085	0.5049
0.10	0.2	0.7137	0.3023	0.5186
0.10	0.3	0.6997	0.3017	0.5081
0.10	0.4	0.6964	0.2986	0.5077
0.10	0.5	0.6925	0.2985	0.5076

Table 10. Cubist model results (default)

Model	Cases	Attributes	Committees	Neighbors	MAE	RMSE	R^2
Cubist	48	12	10	5	0.4738	0.6823	0.8863

Table 11. Results of the Cubist model (tuned)

Model	Cases	Attributes	Committees	Neighbors	MAE	RMSE	R^2
Cubist	48	12	5	4	0.3514	0.5701	0.9437

Table 12. The validation datasets results

Model		R^2	RMSE	MAE
RF (rf method)				
	Default Model	0.6193	1.2044	0.7438
	Tuned Model	0.3009	1.3511	0.8386
RF (ranger method)				
	Default Model	0.8214	0.9944	0.6513
	Tuned Model	0.3534	1.3396	0.8325
SVM				
	Default Model	0.3538	1.3583	0.8490
	Tuned Model	0.3734	1.2842	0.8395
Cubist				
	Default Model	0.8863	0.6823	0.4738
	Tuned Model	0.9437	0.5701	0.3514
MLR		0.9076	0.5255	0.3784

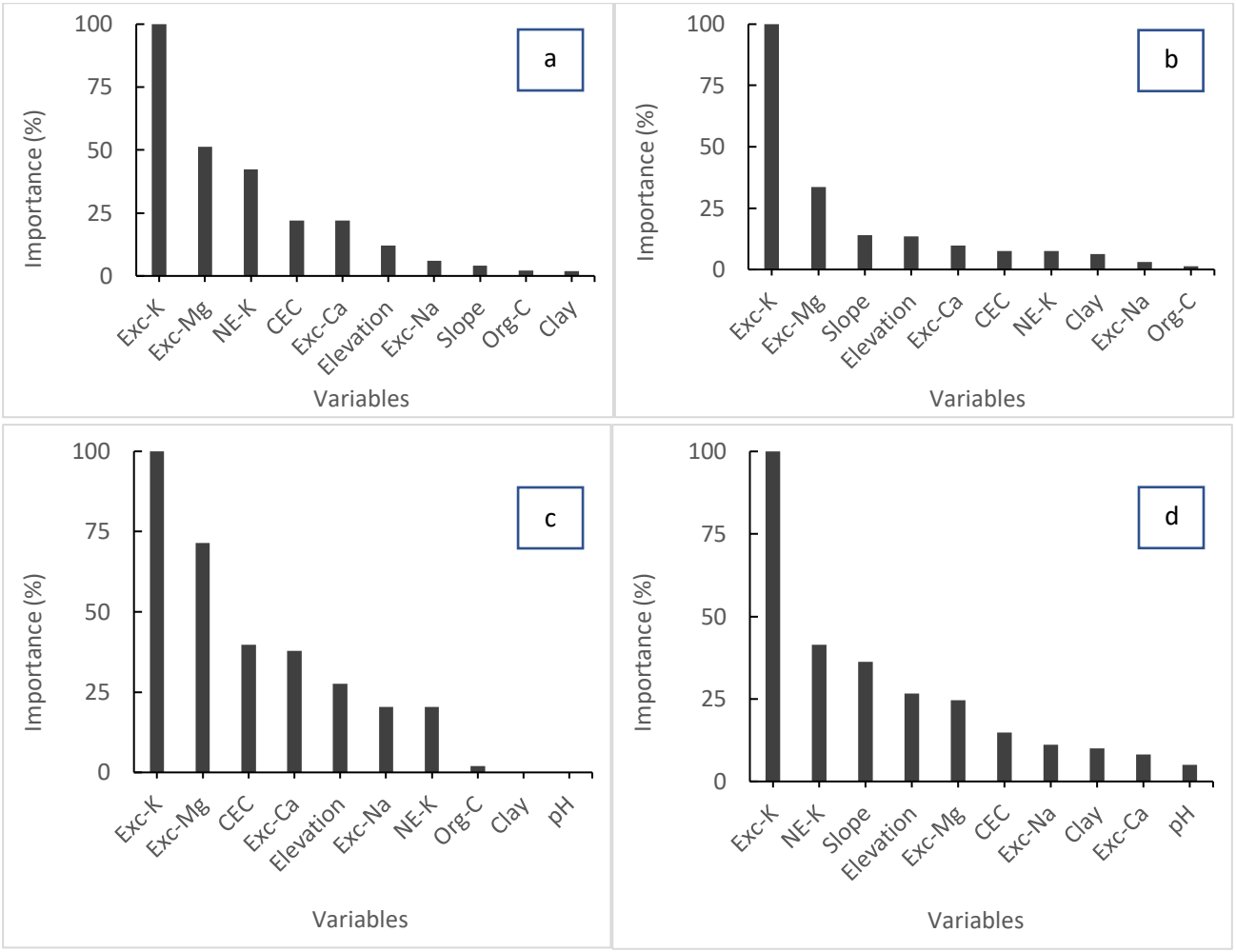

Figure 3. Relative importance variable for ARe^K using RF model "rf" method (a), RF model "ranger" method (b), Cubist (c), and MLR model (d)

Based on the results of four machine learning models, soil properties variables such as Exc-K, Exc-Mg, CEC, Exc-Ca, and NE-K contribute to predicting ARe^K. These results indicate that soil properties variables provide a greater contribution compared to environmental variables, in this case, elevation and slope gradient. One factor that supports this finding is that most of the research data was obtained from locations with flat topography, so environmental variability has no or negligible effects. These results were consistent with previous research (John et al., 2020), which found that environmental variables do not affect soil organic carbon levels on flat topography in Calabar, Cross River State.

Soil properties serve an important function in determining the forms of soil K (Li et al., 2021), including ARe^K. One of the main soil properties that influences is CEC. The soil at the research location had an average soil CEC in the high category, has a clay texture, and contains 2:1 type clay mineral (Nadalia et al., 2024). Clay affects soil CEC, where soil dominated by clay fractions had a higher CEC and water-holding capacity. The type of clay mineral also plays an important role in determining soil CEC. At the study site, Inceptisols contain smectite or vermiculite minerals, which contribute to elevated CEC values than soils dominated by kaolinite or other mineral combinations. Soil containing illite and vermiculite minerals had a high capacity to bind K into the non-exchangeable K form (NE-K), so in soil containing these minerals, ARe^K could be lower even though the total K was high. In addition to Exc-K, NE-K could make a reasonably high contribution to increasing ARe^K in the soil (Islam et al., 2023; Panda & Patra, 2018) and K nutrient uptake (Al Viandari et al., 2024). Hartati et al. (2018) reported a positive correlation between K uptake with total K and available K.

In addition, magnesium nutrients also affect the dynamics of adsorption and desorption of other cations, including K, thus affecting the availability of K for plants (Al–Obaidi & Syan, 2022; Laekemariam et al., 2018; Xie et al., 2021). This condition is relevant in soil with a 2:1 clay mineral type content, where the space between clay mineral layers can trap K into NE-K. The presence of high amounts of exchangeable Ca and Mg may displace K from exchange sites, leading to greater desorption of K into the soil solution and thereby increasing ARe^K (Kassa et al., 2021). However, as reported by Han et al. (2019) and Schneider et al. (2016), high Ca concentrations in the soil solution can reduce K uptake by plants. This is caused due to K displacement from the exchange site, which potentially causing K leaching and promoting selective Ca uptake through mass flow. This dual mechanism illustrates that, while Ca and Mg may enhance ARe^K in the soil, excessive concentrations can suppress K availability to plants through ionic competition.

The use of RF, SVM, and Cubist approaches to predict ARe^K had limitations that should be resolved in continued research. First, one of the limitations of the RF, SVM, and Cubist models was the relatively small dataset size; only 61 data were used in the modeling. To improve the validity and reliability of the model, analysis on a greater dataset is utterly needed. Second, other local environmental parameters (e.g., *in situ* temperature, rainfall, soil moisture) and clay mineral composition can affect ARe^K, as well as further exploration of other hyperparameters of the RF, SVM, and Cubist models, which can improve model performance. Finally, other supervised machine learning techniques with superior results in modeling nonlinear relationships, such as Gaussian Process Regression (GPR), and Artificial Neural Networks (ANN), have not entirely been studied and compared for ARe^K prediction. Optimal prediction results are highly dependent on the data quantity and quality. Although the Cubist model provides the most accurate prediction results, this study did not evaluate its transparency or usability for non-technical users like farmers or extension workers. The model was intended to simplify complex laboratory procedures for determining ARe^K using easily measurable soil and environmental variables. To ensure broader field applicability, particularly for site-specific fertilizer recommendations, future research focus on developing decision-support tools or user-friendly applications that translate the model outputs into practical guidance. This is essential to ensure that the predictive advantages of the Cubist model can be effectively utilized in nutrient management strategies tailored for agricultural stakeholders.

5. CONCLUSION

The research found that the Cubist model (R^2= 0.9437; RMSE= 0.5701; MAE= 0.3514) predicted ARe^K better than MLR, SVM, and RF in agricultural land in Brebes Regency, Central Java. Soil properties (Exc-K, Exc-Mg, CEC, Exc-Ca, and NE-K) were considered important variables in ARe^K prediction in this study. This ML approach in predicting ARe^K can support site-specific K fertilizer recommendations for the agricultural land of Brebes regency. Future studies should improve the accuracy of the model by expanding the dataset, adjusting hyperparameters, or incorporating new predictive variables (e.g., soil properties, clay minerals, climate factors).

Acknowledgements

The author would like to thank the Directorate General of Higher Education (DIKTI), Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia for supporting this research through the Doctoral Dissertation Research Scheme.

Declaration of Competing Interest

The authors declare that no competing financial or personal interests may appear to influence the work reported in this paper.

References

Ajiboye, A. G., Azeez, J. O., & Omotunde, A. J. (2015). Potassium forms and quantity–intensity relationships in some wetland soils of Abeokuta, Southwestern Nigeria. *Archives of Agronomy and Soil Science*, *61*(10), 1393-1408. https://doi.org/10.1080/03650340.2015.1004319.

Al-Hamandi, H., Al-Obaidi, M., & Aljumaily, M. (2019). Study on quantity and intensity of potassium in the alluvial soils in Baghdad. *Plant Archives*, *19*(2), 123-130. https://www.plantarchives.org/SPL%20ISSUE%20SUPP%202,2019/21%20(123-130).pdf.

Al–Obaidi, M. A. J., & Syan, C. M. H. (2022). Binary exchange of calcium, magnesium, potassium and sodium in calcareous soils in Kurdistan region of Iraq. *IOP Conference Series: Earth and Environmental Science*, *1120*(1), 012016. https://doi.org/10.1088/1755-1315/1120/1/012016.

Al Viandari, N., Harsanti, E. S., Supraptomo, E., & Wihardjaka, A. (2024). Utilization of cattle manure and potassium fertilizer on soil potassium availability and yield of cowpea (*Vigna unguiculata* L. Walp) in rainfed rice. *Sains Tanah Journal of Soil Science and Agroclimatology*, *21*(2), 9. https://doi.org/10.20961/stjssa.v21i2.85535.

Balittanah. (2021). *Rekomendasi pemupukan tanaman hortikultura per kabupaten se-Indonesia* [Fertilization recommendations for horticultural crops by districts across Indonesia]. Soil Research Institute, Indonesian Center for Agricultural Land Resources Research and Development, Indonesian Agency for Agricultural Research and Development, Ministry of Agriculture, Republic of Indonesia. https://kikp-pertanian.id/polbangtanmalang/opac/detail-opac?id=20342

Beckett, P. H. T. (1964). Studies on soil potassium: i. confirmation of the ratio law: measurement of potassium potential. *Journal of Soil Science*, *15*(1), 1-8. https://doi.org/10.1111/j.1365-2389.1964.tb00239.x.

Bilias, F., & Barbayiannis, N. (2019). Potassium availability: An approach using thermodynamic parameters derived from quantity-intensity relationships. *Geoderma*, *338*, 355-364. https://doi.org/10.1016/j.geoderma.2018.12.026.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32. https://doi.org/10.1023/A:1010933404324.

Chen, L., Ren, C., Li, L., Wang, Y., Zhang, B., Wang, Z., & Li, L. (2019). A Comparative assessment of geostatistical, machine learning, and hybrid approaches for mapping topsoil organic carbon content. *ISPRS International Journal of Geo-Information*, *8*(4), 174. https://doi.org/10.3390/ijgi8040174.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *machine learning*, *20*(3), 273-297. https://doi.org/10.1007/BF00994018.

Drzewiecki, W. (2016, 2-4 June 2016). Comparison of selected machine learning algorithms for sub-pixel imperviousness change assessment. 2016 Baltic Geodetic Congress (BGC Geomatics), https://doi.org/10.1109/BGC.Geomatics.2016.21

Eviati, Sulaeman, Herawaty, L., Anggria, L., Usman, Tantika, H. E., . . . Wuningrum, P. (2023). *Petunjuk teknis analisis*

kimia tanah, tanaman, air, dan pupuk [Soil, plant, water, and fertilizer chemical analysis] (3rd ed.). Center for Standardization of Agricultural Land Resources Instruments Testing, Agricultural Instrument Standardization Agency, Ministry of Agriculture of the Republic of Indonesia. https://repository.pertanian.go.id/server/api/core/bitstreams/9313cd49-5eaa-4a45-9709-5776ce28e99f/content

Feng, S., Zhou, H., & Dong, H. (2019). Using deep neural network with small dataset to predict material defects. *Materials & Design*, *162*, 300-310. https://doi.org/10.1016/j.matdes.2018.11.060.

Gomes, L. C., Faria, R. M., de Souza, E., Veloso, G. V., Schaefer, C. E. G. R., & Filho, E. I. F. (2019). Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma*, *340*, 337-350. https://doi.org/10.1016/j.geoderma.2019.01.007.

Han, T., Cai, A., Liu, K., Huang, J., Wang, B., Li, D., . . . Zhang, H. (2019). The links between potassium availability and soil exchangeable calcium, magnesium, and aluminum are mediated by lime in acidic soil. *Journal of Soils and Sediments*, *19*(3), 1382-1392. https://doi.org/10.1007/s11368-018-2145-6.

Hartati, S., Minardi, S., Hartatik, W., & Haniati, I. L. (2018). The effects of inorganic fertilizer and mineral leucite residues on K uptake and maize yields (*Zea mays* L.) in Oxisols. *Sains Tanah Journal of Soil Science and Agroclimatology*, *15*(2), 8. https://doi.org/10.15608/stjssa.v15i2.18604.

Helmke, P. A., & Sparks, D. L. (1996). Lithium, sodium, potassium, rubidium, and cesium. In *Methods of Soil Analysis* (pp. 551-574). https://doi.org/10.2136/sssabookser5.3.c19

ICALRD. (2017). *Soil map of Brebes regency 1:50.000*. Indonesian Center for Agricultural Land Resources Research and Development.

Islam, S., Gathala, M. K., Timsina, J., Dutta, S., Salim, M., & Majumdar, K. (2023). Potassium supplying capacity and contribution of non-exchangeable potassium in wetland rice soils in Bangladesh. *Communications in Soil Science and Plant Analysis*, *54*(20), 2745-2762. https://doi.org/10.1080/00103624.2023.2240853.

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*(12), 487. https://doi.org/10.3390/land9120487.

John, K., Kebonye, N. M., Agyeman, P. C., & Ahado, S. K. (2021). Comparison of Cubist models for soil organic carbon prediction via portable XRF measured data. *Environmental Monitoring and Assessment*, *193*(4), 197. https://doi.org/10.1007/s10661-021-08946-x.

Kassa, M., Kebede, F., & Haile, W. (2021). Forms and dynamics of soil potassium in acid soil in the Wolaita zone of southern Ethiopia. *Applied and Environmental Soil Science*, *2021*(1), 9917316. https://doi.org/10.1155/2021/9917316.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., . . . Hunt, T. (2020). *Package 'caret'*. https://cran.radicaldevelop.com/web/packages/caret/caret.pdf

Laekemariam, F., Kibret, K., & Shiferaw, H. (2018). Potassium (K)-to-magnesium (Mg) ratio, its spatial variability and implications to potential Mg-induced K deficiency in Nitisols of Southern Ethiopia. *Agriculture & Food Security*, *7*(1), 13. https://doi.org/10.1186/s40066-018-0165-5.

Lalitha, M., & Dhakshinamoorthy, M. (2015). Quantity-intensity characteristics of potassium (K) in relation to potassium availability under different cropping system in alluvial soils. *African Journal of Agricultural Research*, *10*(19), 2097-2103. https://doi.org/10.5897/AJAR2014.8947.

Li, H., Leng, W., Zhou, Y., Chen, F., Xiu, Z., & Yang, D. (2014). Evaluation models for soil nutrient based on support vector machine and artificial neural networks. *The Scientific World Journal*, *2014*(1), 478569. https://doi.org/10.1155/2014/478569.

Li, T., Liang, J., Chen, X., Wang, H., Zhang, S., Pu, Y., . . . Liu, X. (2021). The interacting roles and relative importance of climate, topography, soil properties and mineralogical composition on soil Potassium variations at a national scale in China. *CATENA*, *196*, 104875. https://doi.org/10.1016/j.catena.2020.104875.

Li, Y., Rahardjo, H., Satyanaga, A., Rangarajan, S., & Lee, D. T.-T. (2022). Soil database development with the application of machine learning methods in soil properties prediction. *Engineering Geology*, *306*, 106769. https://doi.org/10.1016/j.enggeo.2022.106769.

Ließ, M., Schmidt, J., & Glaser, B. (2016). Improving the spatial prediction of soil organic carbon stocks in a complex tropical mountain landscape by methodological specifications in machine learning approaches. *PLOS ONE*, *11*(4), e0153673. https://doi.org/10.1371/journal.pone.0153673.

Ludwig, B., Klüver, K., Filipinski, M., Greenberg, I., Piepho, H.-P., & Cordsen, E. (2022). Description and prediction of copper contents in soils using different modeling approaches—Results of long-term monitoring of soils of northern Germany. *Journal of Plant Nutrition and Soil Science*, *185*(6), 876-887. https://doi.org/10.1002/jpln.202200075.

Lumbanraja, R., Lumbanraja, J., Norvpriansyah, H., & Utomo, M. (2020). Perilaku pertukaran kalium (K) dalam tanah, K terangkut serta produksi jagung (*Zea mays* L.) akibat olah tanah dan pemupukan di tanah Ultisol Gedung Meneng pada musim tanam ketiga [Potassium (K) exchange behavior in soil, harvested K and production of corn (*Zea mays* L.) *Affected by Tillage and Fertilizers in* Ultisol *Soil* of Gedung Meneng at the 3rd planting season]. *Journal of Tropical Upland Resources (J. Trop.*

Upland Res.), *2*(1). https://doi.org/10.23960/jtur.vol2no1.2020.69.

Muliana, Anwar, S., Hartono, A., D. Susila, A., & Sabiham, S. (2018). Pengelolaan dan pemupukan fosfor dan kalium pada pertanian intensif bawang merah di empat desa di Brebes [Management and fertilization of phosphorus and potassium on shallot intensive cultivation in four village in Brebes]. *Jurnal Hortikultura Indonesia (JHI)*, *9*(1), 27-37. https://doi.org/10.29244/jhi.9.1.27-37.

Nadalia, D., Hartono, A., & Pulunggono, H. B. (2024). Quantity-intensity relationship of soil potassium observed in a shallot production center, Central Java, Indonesia. *J. SSAAS*, *30*(2), 145-167.

Nelson, D. W., & Sommers, L. E. (1996). Total carbon, organic carbon, and organic matter. In *Methods of Soil Analysis* (pp. 961-1010). https://doi.org/10.2136/sssabookser5.3.c34

Onyelowe, K. C., Gnananandarao, T., & Ebid, A. M. (2022). Estimation of the erodibility of treated unsaturated lateritic soil using support vector machine-polynomial and -radial basis function and random forest regression techniques. *Cleaner Materials*, *3*, 100039. https://doi.org/10.1016/j.clema.2021.100039.

Padarian, J., Minasny, B., & McBratney, A. B. (2020). Machine learning and soil sciences: a review aided by machine learning tools. *SOIL*, *6*(1), 35-52. https://doi.org/10.5194/soil-6-35-2020.

Panda, R., & Patra, S. K. (2018). Quantity-intensity relations of potassium in representative coastal soils of eastern India. *Geoderma*, *332*, 198-206. https://doi.org/10.1016/j.geoderma.2018.07.014.

Probst, P., Wright, M. N., & Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for random forest. *WIREs Data Mining and Knowledge Discovery*, *9*(3), e1301. https://doi.org/10.1002/widm.1301.

Quinlan, J. R. (1992). Learning with continuous classes. AI '92 (Proceedings of the 5th Australian Joint Conference on Artificial Intelligence), https://doi.org/10.1142/9789814536271

Rossiter, D. G. (2018). Past, present & future of information technology in pedometrics. *Geoderma*, *324*, 131-137. https://doi.org/10.1016/j.geoderma.2018.03.009.

Schneider, A., Augusto, L., & Mollier, A. (2016). Assessing the plant minimal exchangeable potassium of a soil. *Journal of Plant Nutrition and Soil Science*, *179*(4), 584-590. https://doi.org/10.1002/jpln.201600095.

Sharma, A., Weindorf, D. C., Wang, D., & Chakraborty, S. (2015). Characterizing soils via portable X-ray fluorescence spectrometer: 4. Cation Exchange Capacity (CEC). *Geoderma*, *239-240*, 130-134. https://doi.org/10.1016/j.geoderma.2014.10.001.

Soil and Plant Analysis Council Inc. (2000). Major cations (potassium, calcium, magnesium, and sodium). In *Soil Analysis handbook of reference methods* (1st ed.). CRC Press, . https://doi.org/10.1201/9780203739433

Song, Y.-Q., Zhao, X., Su, H.-Y., Li, B., Hu, Y.-M., & Cui, X.-S. (2018). Predicting spatial variations in soil nutrients with hyperspectral remote sensing at regional scale. *Sensors*, *18*(9), 3086. https://doi.org/10.3390/s18093086.

Suarez, D. L. (1996). Beryllium, magnesium, calcium, strontium, and barium. In *Methods of Soil Analysis* (pp. 575-601). https://doi.org/10.2136/sssabookser5.3.c20

Sumner, M. E., & Miller, W. P. (1996). Cation exchange capacity and exchange coefficients. In *Methods of Soil Analysis* (pp. 1201-1229). https://doi.org/10.2136/sssabookser5.3.c40

Szatmári, G., Pirkó, B., Koós, S., Laborczi, A., Bakacsi, Z., Szabó, J., & Pásztor, L. (2019). Spatio-temporal assessment of topsoil organic carbon stock change in Hungary. *Soil and Tillage Research*, *195*, 104410. https://doi.org/10.1016/j.still.2019.104410.

Thomas, G. W. (1996). Soil pH and soil acidity. In *Methods of Soil Analysis* (pp. 475-490). https://doi.org/10.2136/sssabookser5.3.c16

van Reeuwijk, L. P. (2002). *Procedures for soil analysis* (6th ed.). International Soil Reference and Information Centre (ISRIC), Food and Agriculture Organization of the United Nations, Wageningen. https://www.isric.org/sites/default/files/ISRIC_TechPap09.pdf

Wang, X., Yang, Y., Lv, J., & He, H. (2023). Past, present and future of the applications of machine learning in soil science and hydrology. *Soil and Water Research*, *18*(2), 67-80. https://doi.org/10.17221/94/2022-SWR.

Wu, Z., Chen, Y., Yang, Z., Zhu, Y., & Han, Y. (2022). Mapping soil organic carbon in low-relief farmlands based on stratified heterogeneous relationship. *Remote Sensing*, *14*(15), 3575. https://doi.org/10.3390/rs14153575.

Xie, K., Cakmak, I., Wang, S., Zhang, F., & Guo, S. (2021). Synergistic and antagonistic interactions between potassium and magnesium in higher plants. *The Crop Journal*, *9*(2), 249-256. https://doi.org/10.1016/j.cj.2020.10.005.

Zhou, J., Li, E., Wei, H., Li, C., Qiao, Q., & Armaghani, D. J. (2019). Random forests and cubist algorithms for predicting shear strengths of rockfill materials. *Applied Sciences*, *9*(8), 1621. https://doi.org/10.3390/app9081621.

Zhu, D., Lu, J., Cong, R., Ren, T., Zhang, W., & Li, X. (2020). Potassium management effects on quantity/intensity relationship of soil potassium under rice-oilseed rape rotation system. *Archives of Agronomy and Soil Science*, *66*(9), 1274-1287. https://doi.org/10.1080/03650340.2019.1663830.