MULTIDISCIPLINARY SCIENCE JOURNAL

RESEARCH ARTICLE
Published Online: June 27, 2025
https://doi.org/10.31893/multiscience.2025610

Estimation of NPK from soil data using a novel stacked ensemble model

Uditendu Sarkar[a] | **Gouravmoy Banerjee**[b] | **Indrajit Ghosh**[c] ✉

[a]National Informatics Centre (NIC), Ministry of Electronics & Information Technology, Government of India, West Bengal State Centre,Ground Floor, Vidyut Bhawan, DJ Block, Sector II, Salt Lake, Kolkata, West Bengal, India-700091, India.
[b]Eklavya Model Residential School, Swayem, Namok, Mangan, Sikkim, India-737116, India.
[c]Agro-Computing Research Laboratory, Department of Computer Science, Ananda Chandra College, Jalpaiguri, West Bengal, India-735101, India.

Abstract Soil fertility is defined in terms of various nutrients present in the soil and has a direct impact on the sustained production of crops. Amongst various nutrients, nitrogen (N), phosphorous (P), and potassium (K), all together NPK, play the most significant role in gross crop yield. Precise estimation of the fertility status of existing soil NPK is the most significant task for proper fertilizer management. Traditionally, the fertility status of NPK is estimated by chemical analysis, which is expert- dependent and expensive for field applications. Thus, rural farmers apply various fertilizers without estimating the current fertility status of the nutrients, which leads to a nutrient imbalance and a serious threat to sustainable agriculture. Instead, several traditional machine learning models have been suggested for estimating the fertility status of various soil nutrients. However, they have some major limitations with poor performances. Conversely, the stacked ensemble models are more stable and robust compared to base learners with better predictive accuracy. Several successful applications of such models have been reported. However, no stacked ensemble model has been designed for estimating the fertility status of NPK using a trustworthy soil data repository. This paper proposes a novel architecture of a stacked ensemble model for precise estimation of the fertility status of three macronutrients, NPK. The novelty of this work is that the architecture of the existing stacked ensemble models has been restructured by incorporating an additional preprocessing module along with a data feedback technique to boost up the performance. Undoubtedly, a web-based version of such a model will be beneficial for rural farmers and agricultural decision-makers when accessed using smartphones. Our proposed model was experimentally validated using authentic field datasets. Empirical results revealed that our proposed model outperformed all other existing models both in terms of *average accuracy* (0.8457) and *average Cohen's kappa* (0.6114).

Keywords: soil fertility, fertility status, NPK, stacked ensemble model

1. Introduction

Poor soil fertility is one of the most significant challenges that directly reduces the quality and quantity of agricultural yields (Sengupta & Sonwani, 2012). Fertility is defined in terms of various nutrients present in the soil. Amongst them, nitrogen (N), phosphorous (P), and potassium (K), all together NPK are called macronutrients, whereas sulphur (S), iron (Fe), boron (B), copper (Cu), and manganese (Mn) are needed in small quantities and are called micronutrients (Zekri and Obreza, 2003). The macronutrients are required in large quantities, while micronutrients are needed in small quantities, and the farmers are less concerned about the micronutrients. The macronutrients play the most significant role in the growth of a healthy crop, and their inadequate supply enforces a detrimental impact on gross crop yield (Kumar Chauhan et al., 2014). Nitrogen is an important constituent of chlorophyll, which is directly involved in the photosynthesis process (Nursu'aidah, et al., 2014) and helps to boost fruit and seed production, speed up plant growth, and improve the quality of forage crops and leaves (Mengel, 1987). Phosphorus is crucial for root development, flowering, and fruiting of a plant, and is a key component of DNA (deoxyribonucleic acid), RNA (ribonucleic acid), and ATP (adenosine triphosphate) (Marschner, 2011). Potassium is involved in various physiological processes that enhance photosynthesis, activate enzymes, regulate water balance, aid in the transportation of nutrients, and support overall plant growth and development (Pettigrew, 2008). For sustained production of crops, maintaining healthy soil necessitates the timely addition of adequate soil nutrients through proper management of fertilizers. Prior to cultivation, precise estimation of fertility status (or level) of existing macronutrients (NPK) in arable soil is one of the most significant and underpinning tasks for proper fertilizer management leading to better yield. As an agricultural practice, the fertility status of the various soil nutrients is generally estimated through chemical analysis in soil testing laboratories. The scarcity of soil testing laboratories, adequately skilled professionals, and the cost of expensive chemicals make the process practically infeasible to be deployed in the widespread and remote cultivations in most countries. Consequently, rural farmers typically rely on their own traditional fertilization practices without estimating the existing fertility

 Received: February 1, 2025 | Accepted: April 22, 2025

status of the nutrients (Fabregas et al., 2019), which, in turn, results in a nutrient imbalance in their cultivable soil and decreases the quality and quantity of crops. In addition, the nutrient imbalance in the arable soil is a serious threat to sustainable agriculture.

Apart from conventional chemical methods, near-infrared (NIR) and mid-infrared (MIR) spectroscopies coupled with various machine learning (ML) techniques were used for determining nutrient contents and other soil properties of agricultural soil (Jia et al., 2023; Metzger et al., 2024; Zhang et al., 2024; Horf et al., 2024; Chaudhry et al., 2024; Mammadov et al., 2024). Although the spectroscopic methods have high accuracy in the estimation of soil nutrient contents and subsequently soil fertility status, they suffer from two major drawbacks: (i) sample preparation is a challenging task that requires significant expertise, and a lack of it may lead to error in 60 to 70% of cases, and (ii) an appropriate determination of a reference model for calibration of the spectroscope (Nduwamungu et al., 2009). Furthermore, the high cost incurred for high-precision spectroscopes is unaffordable to the rural farmers, particularly in developing countries.

In the last two decades, the use of some machine learning (ML) techniques have been rapidly increasing to design various agricultural decision support systems. For estimating the fertility status of various nutrients, several automated systems were suggested throughout the world using conventional ML techniques, such as artificial neural networks (ANN), fuzzy inference systems (FIS), adaptive neuro-fuzzy inference systems (ANFIS), k-nearest neighbors (k-NN), support vector machines (SVM), regression trees, etc., which include (Zolfaghari et al., 2020; Hosseini et al., 2017; Sunori et al., 2022; Xu et al., 2017; Zhang et al.; 2017, Sunori et al., 2024; Pant et al., 2021; Ganesh et al., 2022; Escorcia-Gutierrez et al., 2022). The significant limitation of such systems was that for estimating the fertility status, all these systems used soil data of various nutrients as inputs, obtained through infield measurement using either expensive chemical analysis or complex spectroscopic methods. To provide a low-cost alternative, a few systems were suggested that used easily accessible soil data from a reliable soil data repository, rather than those collected through expensive infield methods (Suchithra and Pai, 2020; Reshma and Aravindhar, 2022). However, such systems ignored the comprehensive impact of the other coexistent nutrients on estimating the fertility status of a targeted nutrient, and the reported accuracies did not specify the real performance of those systems (Suchithra and Pai, 2020; Reshma and Aravindhar, 2022).

For several field applications, where the quantity of the soil nutrients is not directly measurable, precise estimation of the fertility status of a nutrient is the most challenging task. This issue was first addressed by Sirsat et al. (2017), who applied eight classical ML models to estimate the fertility status of NPK using a standard soil data repository (Sirsat et al., 2017). The performances of the eight ML models were evaluated in terms of Cohen's kappa (k) and were reported to be very poor (values of k ranging from 0.3185 to 0.3508 out of one). Sarkar et al. (2023) suggested another voting classifier-based hybrid system for estimating the fertility status of NPK using soil health card (SHC) data (Sarkar et al., 2023). Though their proposed system performed better than that of Sirsat et al. (2017), the performance was not remarkable in terms of Cohen's kappa (with k = 0.645 for N, k = 0.398 for P, and k = 0.832 for K).

On the other hand, it is evident that the ensemble models are more stable and can overcome the limitations of ordinary ML models (Zounemat-Kermani et al., 2020). The ensemble models combine multiple models rather than a single ML model and minimize the errors by reducing model bias, variance, and overfitting and thus increase the predictive accuracy (Dong et al., 2020). Different categories of ensemble models have been suggested that differ from each other based on both heuristics and learning principles (Kim et al., 2019). Wolpert (1992) first proposed a new architecture for designing an ensemble model, named the stacked ensemble machine. It is a collaborative collection of multiple ML models that can produce a robust model from poorer ones. The basic theoretical foundation and learning techniques of stacked ensemble model are available in Wolpert (1992). Several successful applications of stacked ensemble models have been reported for various domains, which include medical diagnosis (Akyol, 2020; Behera et al., 2019), evapotranspiration estimation (Elbeltagi et al., 2022; Wu et al., 2020; Fan et al., 2018), and credit risk prediction (Pavitha and Sugave, 2023). Particularly, in agriculture, a few stacked ensemble models were successfully applied for plant disease management (Enkvetchakul and Surinta, 2022), crop yield prediction (Renju and Brunda, 2024), crop quality assessment (Xu et al., 2017), and crop recommendation (Raju et al., 2024; Nti et al., 2023).

In the current year, several successful applications of stacked ensemble models in agriculture have been reported. Qaadan et al. (2025), suggested a stacked ensemble model comprising support vector machines, K-nearest neighbors, random forest, gradient-boosted trees, and artificial neural networks for sugarcane disease detection. The model achieved an accuracy of up to 98.89% for disease classification. Wang and Gao (2025), proposed such a model consisting of random forests, categorical boosted trees, light gradient boosted trees, and extremely gradient boosted trees for downscaling soil moisture data. They proved that their stacked ensemble model outperformed any single constituent model. For predicting soluble solid content (SSC) in apples using near-infrared spectroscopy, Zhang et al. (2025), designed another model. In their model, partial least squares regressor (PLS), support vector machines, random forest, and extremely gradient boosted trees were used as base learners (at level 1), and random forest was used as the meta learner (at level 2). The model outperformed the constituent models with RMSE = 0.4253, MAE = 0.3245, MAPE = 0.0271, and R^2 = 0.9520. Ramesh and Kumaresan (2025), developed a stacked ensemble model for crop yield prediction that used six base learners: linear regression, elastic net, extremely gradient boosted trees, AdaBoost, random forest, and k-nearest neighbor with a decision tree regressor as the meta-learner. Their

proposed model outperformed all the traditional models with RMSE = 124.78, MAE = 7.20, and R^2 = 0.98. However, no one has suggested a state-of-the-art stacked ensemble model for estimating the fertility status of a soil nutrient using a soil data repository.

Therefore, the aim of this paper is to propose an improved architecture of a stacked ensemble model for precise estimation of the fertility status of three macronutrients, NPK. The novelty of this work is that the general architecture of the existing stacked ensemble models has been restructured by incorporating an additional preprocessing module along with a data feedback technique to boost up the model performance.

Our proposed model, designed with novel architecture, was empirically validated using authentic field datasets obtained from the Dept. of Agriculture, Govt. of India (DAFW, 2021) for six districts: South 24 Parganas, West Midnapore, Purulia, Malda, Bankura, and Jalpaiguri in the state of West Bengal, India. The empirical results revealed that our proposed stacked ensemble model outperformed the base learners as well as other existing models when compared in terms of both accuracy (ranging from 0.8096 to 0.9336) and Cohen's kappa (ranging from 0.5181 to 0.8610).

The datasets used, suggested system architecture, and the metrics for performance analysis are described in Section 2. The empirical results, discussions, and a comparative study with existing models are presented in Section 3. Section 4 presents our conclusions in their entirety.

2. Materials and Methods

2.1. Datasets used

As case studies, our proposed stacked ensemble model was applied to estimate the fertility status of three targeted macronutrients, N, P, and K, for six districts: South 24 Parganas, West Midnapore, Purulia, Malda, Bankura, and Jalpaiguri, in the state of West Bengal, India. In order to have a proper representation of the soils found in the state of West Bengal, soil data were collected from six districts dispersed across the state with varied soil properties. The Jalpaiguri district is situated in the northern Terai regions of the state; Malda lies in the middle of the Gangetic plain, and South 24 Parganas is in the Sundarban delta area near the Bay of Bengal. The other three districts, West Midnapore, Purulia, and Bankura, lie in one of the westernmost parts of the state.

The relevant soil data were collected from the soil health card (SHC) repository (DAFW, 2021). The primary dataset consists of 12 input parameters, such as soil pH (pH), organic carbon content (OC), electrical conductivity (EC), and the nutrient contents of nitrogen (N), phosphorous (P), potassium (K), sulphur (S), zinc (Zn), iron (Fe), boron (B), manganese (Mn), and copper (Cu), and the targeted output class (fertility status). The soil pH is measured using the 1:2.5 H_2O method, the organic carbon is measured in percentage by the Walkley-Black method, and the electrical conductivity is measured in decisiemens per meter (dS/m). The three macronutrients, NPK, are measured in Kg/ha, while the other nutrients are measured in parts per million (ppm). The targeted output class (fertility status) of each of the macronutrients, NPK, was divided into three categories: *low*, *medium*, and *high*, following the standard guideline recommended by Patil et al., (2016). For convenience, these three output classes were numerically coded as 1 for *low*, 2 for *medium*, and 3 for *high*, respectively. A total of 14228 samples were collected from the soil health card repository (DAFW, 2021) for estimating the fertility status of NPK. Out of these 14228 samples, 60% were used for training and the remaining 40% were used for testing in a random stratified sampling manner (Liberty et al., 2016). The targeted nutrients (NPK), the input parameters used, and the number of samples that were available against each output class are summarized in Table 1.

Table 1 The targeted nutrients, the input parameters used, and the number of samples per output class.

Targeted nutrients	Input parameters used	Number of samples in each output class		
		Low (1)	Medium (2)	High (3)
Nitrogen (N)	pH, OC, EC, P, K, S, Zn, Fe, Cu, Mn, B	5634	8569	25
Phosphorus (P)	pH, OC, EC, N, K, S, Zn, Fe, Cu, Mn, B	2141	10061	2026
Potassium (K)	pH, OC, EC, N, P, S, Zn, Fe, Cu, Mn, B	486	3732	10010

2.2. Architecture of proposed stacked ensemble model

The basic theoretical foundation and learning techniques of stacked ensemble model are available in Wolpert, (1992). Such models use a two-level architecture where the initial inputs are given to the first level (Level-1), which is a stack of multiple base learners. The predictions (outputs) of Level-1 are then used as input to the second level (Level-2), designed with a single learner. The Level-2 provides the final output. This technique aids in designing an improved model with better predictions than the individual models (Rajagopal et al., 2020). In our proposed model, the traditional architecture of stacked ensemble models has been restructured by incorporating an additional preprocessing level (Level-0) and a data feedback mechanism to boost the model performance. Thus, the model is a modified stacked ensemble model consisting of three functional levels: Level-0 (preprocessing level), Level-1 (intermediate level), and Level-2 (final level), and a data feedback path at Level-1. Figure 1 presents the proposed novel architecture of our stacked ensemble model.

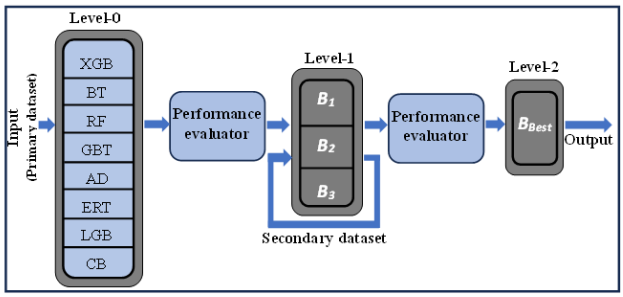

Figure 1 Schematic diagram of the proposed architecture.

In the Level-0, instead of a single learner, we considered a stack of eight ensemble decision tree-based ML models: extreme gradient boosted classifier (XGB), bagging classifier (BT), random forest classifier (RF), gradient boosted classifier (tree-based) (GBT), adaboost classifier (AD), extremely randomized trees classifier (ERT), light gradient boosted classifier (LGB), and categorical boosted classifier (CB) as base learners, rather than the ordinary ML models. These ensemble decision tree-based ML models were selected based on their merits and performance in solving various problems in the agricultural domain. They have been found to be more efficient in solving both classification and regression problems (Sheth et al., 2023). Furthermore, the limitations of conventional decision tree algorithms can be easily overcome by its ensemble variants like random forest (Sheth et al., 2023), gradient boosted trees, categorical boosted trees, etc. (Malek et al., 2022). A similar observation was also reported by the authors themselves (Sarkar et al., 2022). All these eight base learners were trained with the training dataset (60% of the primary dataset), and their hyperparameters were tuned using a fivefold grided search strategy (Owen, 2022). Hyperparameters are the internal variables that are set before the learning starts. In contrast to model parameters, the values of hyperparameters are obtained through training and must be provided explicitly. Appropriate values of hyperparameters can greatly improve model performance. They impact the rate at which a model converges to a solution, and the fine-tuning of the hyperparameters is very significant for designing efficient machine learning models. Five-fold grided search strategy is an effective method for fine-tuning the values of the hyperparameters (Owen, 2022). It is an iterative approach for adjusting the hyperparameters by performing an exhaustive search over a given subset of hyperparameter values. Initially, a grid (subset) of potential values for each hyperparameter is defined, and the algorithm then trains the model several times, taking each value from the defined subset of values of the hyperparameter, usually using cross-validation to ensure the highest performance. The best value of the subset is then chosen for the final model design. For all the base learners, the number of estimators was considered as the hyperparameter, and the set of initial values of each hyperparameter was {10, 50, 100, 150, 200, 250, 300}. The base learners and the best values of the hyperparameters after tuning for each targeted nutrient are summarized in Table 2.

Table 2 The base learners, and best values of hyperparameters for each targeted nutrient.

Base learners	Targeted nutrients	Best values of hyperparameters after tuning
Extreme Gradient Boosted Classifier (XGB)	N	250
	P	10
	K	150
Bagging Classifier (BT)	N	250
	P	250
	K	250
Random Forest Classifier (RF)	N	150
	P	200
	K	300
Gradient Boosted Classifier Tree-Based (GBT)	N	300
	P	250
	K	300
Adaboost Classifier (AD)	N	100
	P	50
	K	100
Extremely Randomized Trees Classifier (ERT)	N	200
	P	250
	K	300
Light Gradient Boosted Classifier (LGB)	N	300
	P	50
	K	200
Categorical Boosted Classifer (CB)	N	300
	P	300
	K	300

A performance evaluator evaluates the performance of these eight base learners in terms of five well-accepted statistical metrics: accuracy (*Ac*), weighted precision (*Pr*), weighted recall (*Rc*), weighted F-score (*F*), and Cohen's kappa (*k*) (Sokolova et al., 2006; Chicco et al., 2021; Behera et al., 2019) using the testing datasets, and the best three (B_1, B_2, and B_3) were selected for each targeted nutrient. The second level, Level-1, was designed with a stack of these best three (B_1, B_2, and B_3) base learners.

At the time of training, all these eight base learners generated an output dataset consisting of decision attributes (fertility status). The outputs generated by these three best base learners (B_1, B_2, and B_3) were segregated from this output dataset to construct the secondary training dataset. Similarly, the secondary testing dataset was constructed using the same technique. The three best learners (B_1, B_2, and B_3) in Level-1 were further trained and tested with these secondary training and testing datasets. The output datasets (training and testing) generated from the primary datasets by the best three base learners were again used to train and test themselves (data feedback path). The performances of the best three learners were evaluated again by a performance evaluator in terms of five performance metrics: *Ac*, *Pr*, *Rc*, *F*, and *k* (Sokolova et al., 2006; Chicco et al., 2021; Behera et al., 2019), and the best (B_{Best}) among the best three (B_1, B_2, and B_3) was selected. Finally, the Level-2 was designed with the best learner (B_{Best}) to give the best possible output.

2.3. Performance evaluation

Several measures have been proposed to evaluate a classifier's effectiveness (Tallón-Ballesteros et al., 2016). Five well-accepted metrics, accuracy score (*Ac*), weighted precision (*Pr*), weighted recall (*Rc*), weighted F-scores (*F*), and Cohen's kappa (*k*), were used by the performance evaluator to evaluate the performance of each learner (Sokolova et al., 2006; Chicco et al., 2021; Behera et al., 2019). The higher values of *Ac*, *Pr*, *Rc*, *F*, and *k* indicate better model performance. All these metrics are mathematically defined as (Chicco et al., 2021; Behera et al., 2019):

$$Accuracy\ (Ac) = \frac{Tp+Tn}{Tp+Tn+Fp+Fn} \quad (1)$$

$$Weighted\ Precision\ (Pr) = \frac{\sum_{i=1}^{m} |y_i| \frac{Tp_i}{Tp_i+fp_i}}{\Sigma_{i=1}^{m}|y_i|} \quad (2)$$

$$Weighted\ Recall\ (Rc) = \frac{\sum_{i=1}^{m} |y_i| \frac{Tp_i}{Tp_i+Fn_i}}{\Sigma_{i=1}^{m}|y_i|} \quad (3)$$

$$Weighted\ F\text{-}Score\ (F) = \frac{\sum_{i=1}^{m} |y_i| \frac{2Tp_i}{2Tp_i+Fp_i+Fn_i}}{\sum_{i=1}^{m}|y_i|} \quad (4)$$

Where the total number of classes is *m*, and the number of instances that belong to class *i* is represented by $|y_i|$. The *Tp*, *Tn*, *Fp*, and *Fn* are the true positive, true negative, false positive, and false negative counts obtained at testing.

Cohen's kappa (*k*) is a well-recognized and commonly used metric for evaluating the performance of a classifier. It expresses the degree of agreement or disagreement between two instances. The generalized expression to calculate the value of Cohen's kappa (*k*) for the m classes is given by (Chicco et al., 2021):

$$Cohen's\ kappa\ (k) = \frac{N.\sum_{i=1}^{m} C_{ii} - \sum_{i=1}^{m} C_{i_{cr}}.C_{i_{pr}}}{N^2 - \sum_{i=1}^{m} C_{i_{cr}}.C_{i_{pr}}} \quad (5)$$

Where $\sum_{i=1}^{m} C_{ii}$ and *N* is the total number of successfully predicted instances, and the total number of instances, respectively. $C_{i_{cr}}$ is the total number of correctly categorized instances for class *i*, and $C_{i_{pr}}$ is the total number of instances correctly predicted as belonging to class *i*.

3. Results and Discussion

3.1. Performance analysis

The empirical values of the performance metrics accuracy score (*Ac*), weighted precision (*Pr*), weighted recall (*Rc*), weighted F-scores (*F*), Cohen's kappa (*k*), and the average values (*Avg*) obtained at Level-1 for three macronutrients: N, P, and K are summarized in Table 3. The values of these five metrics and their average values at Level-2 are presented in Table 4.

In the case of N, it is observed that at Level-1, the LGB model achieved the best performance (with an average accuracy *Avg* = 0.9172), followed by the RF and XGB models (with *Avg* = 0.9126 and 0.9121, respectively). However, in the case of P, comparatively lower performances were observed for all these eight models. In terms of average accuracy, the best three models were RF, ERT, and BT (with *Avg* = 0.7157, 0.7139, and 0.7154, respectively). Similarly, for K, the RF, LGB, and ERT models were the best three models with *Avg* = 0.7410, 0.7373, and 0.7311, respectively. It is observed from Table 4 that for level-2,

the RF model was the outperformer in terms of average accuracy (with *Avg* = 0.9185 and 0.7433, respectively) for N and K, while in the case of P, the BT model was the best (with *Avg* = 0.7182).

Table 3 Summary of results of Level-1 models.

	Classifier	*Ac*	*Pr*	*Rc*	*F*	*K*	*Avg.*
Nitrogen (N)	XGB:	0.9283	0.9266	0.9283	0.9273	0.8499	0.9121
	BT:	0.9276	0.9261	0.9276	0.9265	0.8478	0.9111
	RF:	0.9288	0.9274	0.9288	0.9277	0.8503	0.9126
	GBT:	0.9209	0.9214	0.9209	0.9208	0.8343	0.9037
	AD:	0.8593	0.8625	0.8593	0.8559	0.6975	0.8269
	ERT:	0.9178	0.9168	0.9178	0.9164	0.8263	0.8990
	LGB:	0.9325	0.9308	0.9325	0.9315	0.8586	0.9172
	CB:	0.9245	0.9228	0.9245	0.9234	0.8415	0.9073
Phosphorus (P)	XGB:	0.7809	0.7809	0.7809	0.7496	0.3956	0.6976
	BT:	0.7920	0.7882	0.7920	0.7670	0.4377	0.7154
	RF:	0.7929	0.7907	0.7929	0.7665	0.4355	0.7157
	GBT:	0.7767	0.7659	0.7767	0.7499	0.3968	0.6932
	AD:	0.7461	0.7405	0.7461	0.7054	0.2972	0.6471
	ERT:	0.7925	0.7902	0.7925	0.7645	0.4296	0.7139
	LGB:	0.7836	0.7770	0.7836	0.7562	0.4116	0.7024
	CB:	0.7807	0.7673	0.7807	0.7599	0.4231	0.7024
Potassium (K)	XGB:	0.7995	0.7820	0.7995	0.7818	0.5089	0.7344
	BT:	0.8029	0.7844	0.8029	0.7822	0.5098	0.7365
	RF:	0.8082	0.7926	0.8082	0.7835	0.5124	0.7410
	GBT:	0.7901	0.7688	0.7901	0.7625	0.4609	0.7145
	AD:	0.7200	0.6886	0.7200	0.6917	0.3044	0.6249
	ERT:	0.8031	0.7876	0.8031	0.7725	0.4894	0.7311
	LGB:	0.8029	0.7868	0.8029	0.7833	0.5104	0.7373
	CB:	0.8018	0.7860	0.8018	0.7831	0.5099	0.7365

Table 4 The values of performance metrics for the best three models at Level-2 for NPK.

Nutrient	Classifier	Ac	Pr	Rc	F	K	Avg.
N	RF:	0.9336	0.9319	0.9336	0.9326	0.8610	0.9185
P	BT:	0.7939	0.7876	0.7939	0.7705	0.4451	0.7182
K	RF:	0.8096	0.7947	0.8096	0.7848	0.5181	0.7433

Comparing the values of performance metrics summarized in Tables 3 and 4, an improvement in the model performance is observed in Level-2 as compared to Level-1 for all three nutrients. As the best model, the value of *Avg* of the RF model was improved by 0.64% from 0.9126 to 0.9185 for N. Similarly, for P, the performance of the best model BT improved by 0.39% in terms of *Avg* from 0.7154 to 0.7182. For K, the RF model exhibited an improvement of 0.31% from 0.7410 to 0.7433, in terms of *Avg*. Therefore, it was observed that the performance of all the models at Level-2 improved when compared with their performances at Level-1. Thus, it is revealed that our proposed stacked ensemble model, consisting of a three-layer architecture and a data feedback mechanism, achieved higher accuracy as compared to a single-level machine learning model for the estimation of NPK using available soil datasets.

3.2. Comparative study with other models

We further compared our proposed stacked ensembled model with other prevailing models suggested for soil NPK estimation. These models were suggested by Sirsat et al. (2017) (Sirsat et al., 2017), Suchitra and Pai (2020) (Suchitra and Pai, 2020), Reshma and Aravindhar (2022) (Reshma and Aravindhar, 2022), and Sarkar et al. (2025) (Sarkar et al., 2023). Each of these models has been assigned a unique model number (Model-1 to Model-4) for better understanding. In order to compare these models, two performance metrics, average accuracy (*Avg-Acc*) and average Cohen's kappa (*Avg-k*), were considered because either of them or both of them were reported in their respective papers. It is observed that for Model-1, the average accuracy was not reported, and for Model-3, the average value of Cohen's kappa was not derived. Table 5 presents the values of these metrics.

From the two comparative graphs, presented in figures 2 and 3, it is observed that in terms of average accuracy, Model 3 was the worst (*Avg-Acc* = 0.3900), and in terms of average Cohen's kappa, Model-1 showed the lowest performance (*Avg-k* = 0.3739). The second-best position was secured by Model-2 and Model-4, respectively, when compared in terms of average values of accuracy and Cohen's kappa. Empirical results revealed that our proposed model outperformed all other models both in terms of average accuracy (*Avg-Acc* = 0.8457) and average Cohen's kappa (*Avg-k* = 0.6114).

Table 5 The average values of accuracy and Cohen's kappa for different models.

Sl. No.	Suggested by	Models	Nutrient	*Avg-Acc*	*Avg-k*
1.	Sirsat et al., 2017	Model-1	Nitrogen (N) Phosphorous (P) Potassium (K)	NR	0.3739
2.	Suchitra and Pai, 2019	Model-2	Nitrogen Phosphorous (P) Potassium (K)	0.8422	0.5084
3.	Reshma and Aravindhar, 2022	Model-3	Nitrogen (N) Phosphorous (P) Potassium (K)	0.3900	NR
4.	Sarkar et al., 2023	Model-4	Nitrogen (N) Phosphorous (P) Potassium (K)	0.6234	0.5619
5.	Proposed model (Sarkar et al., 2025)	Proposed model	Nitrogen (N) Phosphorous (P) Potassium (K)	0.8457	0.6114

*NR: Not reported.

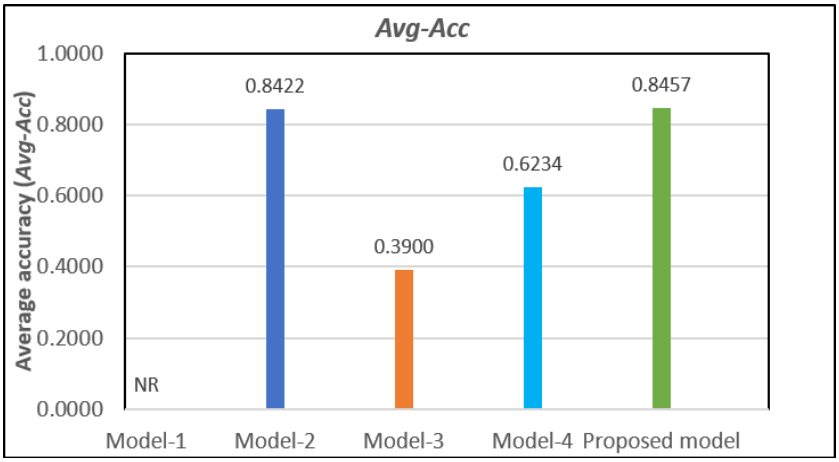

Figure 2 Comparative graph in terms of average accuracy (*Avg-Acc*).

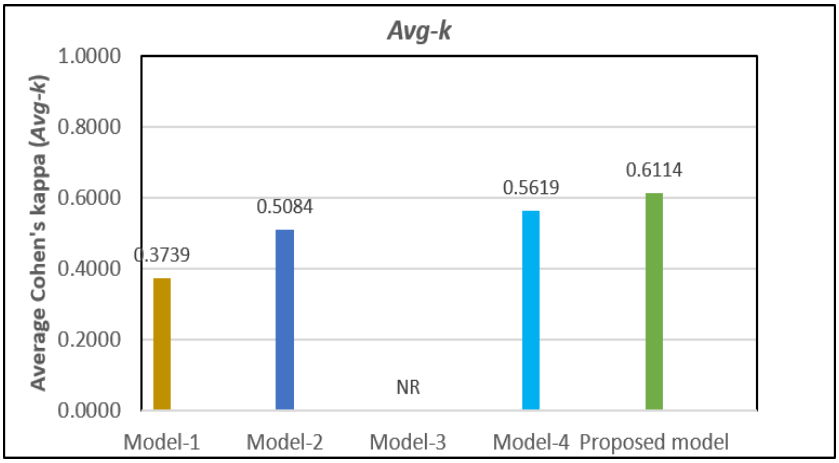

Figure 3 Comparative graph in terms of average Cohen's kappa (*Avg-k*).

4. Conclusions

Soil fertility has a great impact on crop production. Precise estimation of macronutrient level prior to planting any crop is one of the major challenges for better production of crops, both in terms of quality and yield. Our proposed stacked ensemble model with novel architecture addressed this issue with the highest performance reported to date.

The main focus of our present study is to suggest an improved architecture of a stacked ensemble model for precise estimation of the fertility status of soil nutrients. As a case study, the performance of our proposed model was empirically validated with three macronutrients, NPK. The novelty of this work is that the general architecture of the existing stacked

ensemble models has been restructured by incorporating an additional preprocessing module along with a data feedback technique to boost the model performance.

Rather than a random selection of machine learning models in each level, a more scientific performance-based selection was adopted to select the best classifier(s) in each level using the Soil Health Card data collected from the Indian Council of Agricultural Research. This approach helped to explore and utilize the context-dependent performance of the best classifier from eight popular classifiers arranged in a three-level stack and a secondary data feedback technique to estimate the NPK content in agricultural soil.

The experimental results revealed that the stacked ensemble model achieved better accuracy than the common ensemble models, particularly in this agricultural domain. Furthermore, the comparative study with other existing models exposed that the proposed model attained a 15% improvement over the existing models.

Though our proposed model outperformed other existing models, there is still a lot of scope for future improvement by applying other hyperparameter optimization techniques. Moreover, it is to be noted that since the soil samples are collected from West Bengal, the model trained and tested with this dataset will be applicable only in West Bengal. Therefore, it is acknowledged that the model is not a generalized one; however, like all other machine learning models, it can be trained and tested with data available for other states for more generalization. This will be the focus of our future attempts.

Ethical considerations

Not Applicable.

Declaration of interest

The authors declare no conflicts of interest

Funding

No funds were received to conduct this research.

References

Akyol, K. (2020). Stacking ensemble based deep neural networks modeling for effective epileptic seizure detection. *Expert Systems with Applications*, *148*, 113239. https://doi.org/10.1016/j.eswa.2020.113239

Behera, B., Kumaravelan, G., & Kumar, P. (2019). Performance evaluation of deep learning algorithms in biomedical document classification. In *Proceedings of the 11th International Conference on Advanced Computing, ICoAC 2019* (p.220-224). Chennai, India. https://doi.org/10.1109/ICoAC48765.2019.246843

Chaudhry, H., Vasava, H. B., Chen, S., Saurette, D., Beri, A., Gillespie, A., & Biswas, A. (2024). Evaluating the soil quality index using three methods to assess soil fertility. *Sensors, 24*(3), 864. https://doi.org/10.3390/s24030864

Chicco, D., Tötsch, N., & Jurman, G. (2021). The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining, 14*, 1–22. https://doi.org/10.1186/s13040-021-00244-z

Department of Agriculture & Farmers Welfare (DAFW). (2021). *Soil health card*. Ministry of Agriculture and Farmers Welfare, Government of India. https://soilhealth.dac.gov.in/home. Accessed on August 20, 2022.

Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science*, *14*, 241-258. https://doi.org/10.1007/s11704-019-8208-z

Elbeltagi, A., Kushwaha, N. L., Rajput, J., Vishwakarma, D. K., Kulimushi, L. C., Kumar, M., Zhang, J., Pande, C. B., Choudhari, P., Meshram, S. G., Pandey, K., Sihag, P., Kumar, N., & Abd-Elaty, I. (2022). Modelling daily reference evapotranspiration based on stacking hybridization of ANN with meta-heuristic algorithms under diverse agro-climatic conditions. *Stochastic Environmental Research and Risk Assessment*, *36*(10), 3311-3334. https://doi.org/10.1007/s00477-022-02196-0

Enkvetchakul, P., & Surinta, O. (2022). Stacking ensemble of lightweight convolutional neural networks for plant leaf disease recognition. *ICIC Express Letters*, *16*(5), 521-528. https://doi.org/10.24507/icicel.16.05.521

Escorcia-Gutierrez, J., Gamarra, M., Soto-Diaz, R., Pérez, M., Madera, N., & Mansour, R. F. (2022). Intelligent Agricultural Modelling of Soil Nutrients and pH Classification Using Ensemble Deep Learning Techniques. *Agriculture*, *12*(7), 977. https://doi.org/10.3390/agriculture12070977

Fabregas, R., Kremer, M., & Schilbach, F. (2019). Realizing the potential of digital development: The case of agricultural advice. *Science, 366*, eaay3038. https://doi.org/10.1126/science.aay3038

Fan, J., Yue, W., Wu, L., Zhang, F., Cai, H., Wang, X., Lu, X., & Xiang, Y. (2018). Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. *Agricultural and forest meteorology*, *263*, 225-241. https://doi.org/10.1016/j.agrformet.2018.08.019

Ganesh, D., Yeshwanth, K. J. A., Satheesh, M., Guru, M., Reddy, S. V., Chirudeep, T., Naveen, S., & Polisetty, K. (2022). Extreme learning mechanism for classification & prediction of soil fertility index. *Journal of Pharmaceutical Negative Results, 13,* 37-43. https://doi.org/10.47750/pnr.2022.13.S06.006

Horf, M., Gebbers, R., Olfs, H. W., & Vogel, S. (2024). Determining nutrients, dry matter, and pH of liquid organic manures using visual and near-infrared spectrometry. *Science of the Total Environment, 908,* 168045. https://doi.org/10.1016/j.scitotenv.2023.168045

Hosseini, M., Rajabi Agereh, S., Khaledian, Y., Jafarzadeh Zoghalchali, H., Brevik, E. C., & Movahedi Naeini, S. A. R. (2017). Comparison of multiple statistical techniques to predict soil phosphorus. *Applied Soil Ecology, 114*, 123–131. https://doi.org/10.1016/j.apsoil.2017.02.011

Jia, X., Fang, Y., Hu, B., Yu, B., & Zhou, Y. (2023). Development of soil fertility index using machine learning and visible-near-infrared spectroscopy. *Land, 12*(12), 2155. https://doi.org/10.3390/land12122155

Kim, D., Yu, H., Lee, H., Beighley, E., Durand, M., Alsdorf, D. E., & Hwang, E. (2019). Ensemble learning regression for estimating river discharges using satellite altimetry data: Central Congo River as a Test-bed. *Remote sensing of environment*, *221*, 741-755. https://doi.org/10.1016/j.rse.2018.12.010

Kumar Chauhan, D., Kumar Tripathi, D., Pratap Singh, V., Mohan Prasad, S., & Dubey, N. K. (2014). Role of macronutrients in plant growth and acclimation: Recent advances and future prospective. In *Improvement of Crops in the Era of Climatic Changes* (p. 197-216). Springer New York. https://doi.org/10.1007/978-1-4614-8824-8_8

Liberty, E., Lang, K., & Shmakov, K. (2016). Stratified sampling meets machine learning. In *International conference on machine learning* (p. 2320-2329). PMLR.

Malek, N. H. A., Yaacob, W. F. W., Nasir, S. a. M., and Shaadan, N. (2022). Prediction of water quality classification of the Kelantan River Basin, Malaysia, using machine learning techniques. *Water*, *14*(7), 1067. https://doi.org/10.3390/w14071067

Mammadov, E., Denk, M., Mamedov, A. I., & Glaesser, C. (2024). Predicting Soil Properties for Agricultural Land in the Caucasus Mountains Using Mid-Infrared Spectroscopy. *Land*, *13*(2), 154-154. https://doi.org/10.3390/land13020154

Marschner, H. (2011). *Marschner's mineral nutrition of higher plants* (3rd ed.). Academic Press.

Mengel, K. (1987). *Principles of plant nutrition*. International Potash Institute.

Metzger, K., Liebisch, F., Herrera, J. M., Guillaume, T., Walder, F., & Bragazza, L. (2024). The use of visible and near-infrared spectroscopy for in-situ characterization of agricultural soil fertility: A proposition of best practice by comparing scanning positions and spectrometers. *Soil Use and Management, 40*(1), e12952. https://doi.org/10.1111/sum.12952

Nduwamungu, C., Ziadi, N., Parent, L. É., Tremblay, G. F., & Thuriès, L. (2009). Opportunities for, and limitations of, near infrared reflectance spectroscopy applications in soil analysis: A review. Canadian Journal of Soil Science, *89*(5), 531-541.

Nti, I. K., Zaman, A., Nyarko-Boateng, O., Adekoya, A. F., & Keyeremeh, F. (2023). A predictive analytics model for crop suitability and productivity with tree-based ensemble learning. *Decision Analytics Journal*, *8*, 100311. https://doi.org/10.1016/j.dajour.2023.100311

Nursu'aidah, H., Motior, M. R., Nazia, A. M., and Islam, M. A. (2014). Growth and photosynthetic responses of long bean (*Vigna unguiculata*) and mung bean (*Vigna radiata*) to fertilization. *The Journal of Animal and Plant Sciences, 24*(2), 573–578. https://www.researchgate.net/publication/286770664_Growth_and_photosynthetic_responses_of_long_bean_Vigna_unguiculata_and_mung_bean_Vigna_radiata_response_to_fertilization

Owen, L. (2022). *Hyperparameter tuning with Python: Boost your machine learning model's performance via hyperparameter tuning*. Packt Publishing Ltd.

Pant, J., Pant, P., Pant, R. P., Bhatt, A., Pant, D., & Juyal, A. (2021). Soil quality prediction for determining soil fertility in Bhimtal block of Uttarakhand (India) using machine learning. *International Journal of Analysis and Applications, 19*(1), 91-109. https://doi.org/10.28924/2291-8639

Patil, P. L., Kuligod, V. B., Gundlur, S. S., Katti, J., Nagaral, I. N., Shikrashetti, P., Geetanjali, H. M., & Dasog, G. S. (2016). Soil fertility mapping in Dindur sub-watershed of Karnataka for site-specific recommendations. *Journal of the Indian Society of Soil Science, 64*(4), 381-390. https://doi.org/10.5958/0974-0228.2016.00050.5

Pavitha, N., & Sugave, S. (2023). A Novel Multi-Stage Stacked Ensemble Classifier using Heterogeneous Base Learners. *International Journal of Engineering Trends and Technology, 71*(4)*,* 65-71. https://doi.org/10.14445/22315381/IJETT-V71I4P206

Pettigrew, W. T. (2008). Potassium influences on yield and quality production for maize, wheat, soybean, and cotton. *Physiologia Plantarum, 133*(4), 670–681. https://doi.org/10.1111/j.1399-3054.2008.01073.x

Qaadan, S., Alshare, A., Ahmed, A., and Altartouri, H. (2025). Stacked ensembles powering smart farming for imbalanced sugarcane disease detection. *Applied Sciences*, *15*(5), 2788. https://doi.org/10.3390/app15052788

Rajagopal, S., Kundapur, P. P., & Hareesha, K. S. (2020). A stacking ensemble for network intrusion detection using heterogeneous datasets. *Security and Communication Networks*, *2020*(1), 4586875. https://doi.org/10.1093/bib/bbab172

Raju, C., Ashoka, D. V., & BV, A. P. (2024). CropCast: Harvesting the future with interfused machine learning and advanced stacking ensemble for precise crop prediction. *Kuwait Journal of Science*, *51*(1), 100160. https://doi.org/10.1016/j.kjs.2023.11.009

Ramesh, V., and Kumaresan, P. (2025). Stacked ensemble model for accurate crop yield prediction using machine learning techniques. *Environmental Research Communications*, *7*(3), 035006. https://doi.org/10.1088/2515-7620/adb9c0

Renju, K., & Brunda, V. (2024). Optimizing crop yield prediction through multiple models: an ensemble stacking approach. *International Journal of Data Informatics and Intelligent Computing*, *3*(2), 52-58. https://doi.org/10.59461/ijdiic.v3i2.120

Reshma, S. J., & Aravindhar, D. J. (2022). A systematic approach of classifying soil & crop nutrient using machine learning algorithms. *International Journal of Intelligent Systems and Applications in Engineering*, *10*(2S), 174-179. Retrieved from https://www.ijisae.org/index.php/IJISAE/article/view/2380

Sarkar, U., Banerjee, G., and Ghosh, I. (2022). A machine learning model for estimation of village level soil nutrient index. *Indian Journal of Science and Technology*, *15*(36), 1815–1822. https://doi.org/10.17485/ijst/v15i36.851

Sarkar, U., Sharma, M., Dixit, S., Das, J., Banerjee, G., & Ghosh, I. (2023). Estimation of soil macronutrients using voting classifiers. In *Proceedings of International Conference on Communication and Computational Technologies: Algorithms for Intelligent Systems* (p. 779-788). Jaipur, India. https://doi.org/10.1007/978-981-19-3951-8_59

Sengupta, A. & Sonwani, D. (2012). Sustainable development in India with reference to agricultural sector. *International Journal of Emerging Research in Management & Technology*, *1*(2), 24-29.

Sheth, V., Tripathi, U., and Sharma, A. (2022). A Comparative analysis of Machine learning Algorithms for classification purpose. *Procedia Computer Science*, *215*, 422–431. https://doi.org/10.1016/j.procs.2022.12.044

Sirsat, M. S., Cernadas, E., Fernández-Delgado, M., & Khan, R. (2017). Classification of agricultural soil parameters in India. *Computers and Electronics in Agriculture, 135*, 269-279. https://doi.org/10.1016/j.compag.2017.01.019

Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation. In *Australasian joint conference on artificial intelligence* (p. 1015-1021). Berlin, Heidelberg.

Suchithra, M. S., & Pai, M. L. (2020). Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters. *Information Processing in Agriculture, 7*(1), 72-82. https://doi.org/10.1016/j.inpa.2019.05.003

Sunori, S. K., Negi, P. B., Garia, P., Arora, S., Lohani, M. C., Mittal, A., & Juneja, P. (2022). Design of ANN-based classifiers for soil fertility of Uttarakhand. In *3rd International Conference for Emerging Technology (INCET)* (p. 1-5). Belgaum, India. https://doi.org/10.1109/INCET54531.2022.9825273

Sunori, S. K., Negi, P. B., Joshi, N. C., Mittal, A., & Juneja, P. (2024). Soil fertility assessment using ensemble methods in machine learning. In *Proceedings of the 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks, ICICV 2024*, (p. 17-21). Tirunelveli, India. https://doi.org/10.1109/ICICV62344.2024.00010

Tallón-Ballesteros, A. J., & Riquelme, J. C. (2014). Data mining methods applied to a digital forensics task for supervised machine learning. In Muda, A., Choo, YH., Abraham, A., N. Srihari, S. (eds) Computational Intelligence in Digital Forensics: Forensic Investigation and Applications. *Studies in Computational Intelligence* (pp. 413-428). Springer. https://doi.org/10.1007/978-3-319-05885-6_17

Wang, L., and Gao, Y. (2025). Estimating and downscaling ESA-CCI soil moisture using Multi-Source remote sensing images and Stacking-Based ensemble learning algorithms in the Shandian River Basin, China. *Remote Sensing*, *17*(4), 716. https://doi.org/10.3390/rs17040716

Wolpert, D. H. (1992). Stacked generalization. *Neural networks*, *5*(2), 241-259. https://doi.org/10.1016/S0893-6080(05)80023-1

Wu, T., Zhang, W., Jiao, X., Guo, W., & Hamoud, Y. A. (2020). Comparison of five Boosting-based models for estimating daily reference evapotranspiration with limited meteorological variables. *Plos One*, *15*(6), e0235324. https://doi.org/10.1371/journal.pone.0235324

Xu, C., Xu, X., Liu, M., Yang, J., Zhang, Y., & Li, Z. (2017). Developing pedotransfer functions to estimate the S-index for indicating soil quality. *Ecological Indicators, 83*, 338-345. https://doi.org/10.1016/j.ecolind.2017.08.011

Zekri, M., & Obreza, T. A. (2003). Plant Nutrients for Citrus Trees: SL 200/SS419, 1/2003. EDIS, 2003(2). https://doi.org/10.32473/edis-ss419-2003

Zhang, H., Wu, P., Yin, A., Yang, X., Zhang, M., & Gao, C. (2017). Prediction of soil organic carbon in an intensively managed reclamation zone of eastern China: A comparison of multiple linear regressions and the random forest model. *Science of the Total Environment, 592*, 704-713. https://doi.org/10.1016/j.scitotenv.2017.02.146

Zhang, L., Huang, Z., and Zhang, X. (2025). Stacking Ensemble Learning Method for Quantitative Analysis of Soluble Solid Content in Apples. Journal of Chemometrics, 39(1), e3635.

Zhang, X., Liu, D., Ma, J., Wang, X., Li, Z., & Zheng, D. (2024). Visible near-infrared hyperspectral soil organic matter prediction based on combinatorial modeling. *Agronomy, 14*(4), 789. https://doi.org/10.3390/agronomy14040789

Zolfaghari, A. A., Abolkheiryan, M., Soltani-Toularoud, A. A., Taghizadeh-Mehrjardi, R., & Weldeyohannes, A. O. (2020). Prediction of soil macronutrients using fractal parameters and artificial intelligence methods. *Spanish Journal of Agricultural Research, 18*(2), 1-13. https://doi.org/10.5424/sjar/2020182-15460

Zounemat-Kermani, M., Stephan, D., Barjenbruch, M., & Hinkelmann, R. (2020). Ensemble data mining modeling in corrosion of concrete sewer: A comparative study of network-based (MLPNN & RBFNN) and tree-based (RF, CHAID, & CART) models. *Advanced Engineering Informatics*, *43*, 101030. https://doi.org/10.1016/j.aei.2019.101030