DOI : 10.5958/2349-4433.2024.00118.1

J Krishi Vigyan 2024, 12(3) : 691-694

Soil Nutrient Status Prediction Using Machine Learning Algorithms

Gigi Annee Mathew[1*], Varsha Jotwani[1] and A K Singh[2]
Department of CS & IT, Rabindranath Tagore University, Bhopal, Madhya Pradesh
Krishi Vigyan Kendra, JNKVV, Jabalpur, Madhya Pradesh

Soil nutrients play a pivotal role in facilitating optimal plant development and enhancing agricultural yield. The precise assessment of soil nutrient levels is paramount for making informed agricultural choices, encompassing crop selection, land preparation, and fertilizer application. This study incorporates diverse supervised machine learning approaches, including K-Nearest Neighbour, Decision Tree, Random Forest, Support Vector Machine and Naive Bayes, to analyse soil nutrient profiles. A total of 12 soil parameters were employed to categorize soil nutrients into low, medium, and high classifications. Post pre-processing, the dataset underwent a division into training and testing datasets. Algorithms were applied to the training collection and then assessed with the test dataset, using Python for coding. The random forest model achieved the highest accuracy, reaching 99%, thus surpassing alternative methodologies. The research highlights that the application of machine learning strategies, notably the random forest method, can greatly advance the accuracy of soil nutrient forecasts, allowing farmers to make wiser decisions that increase productivity and optimize land management.

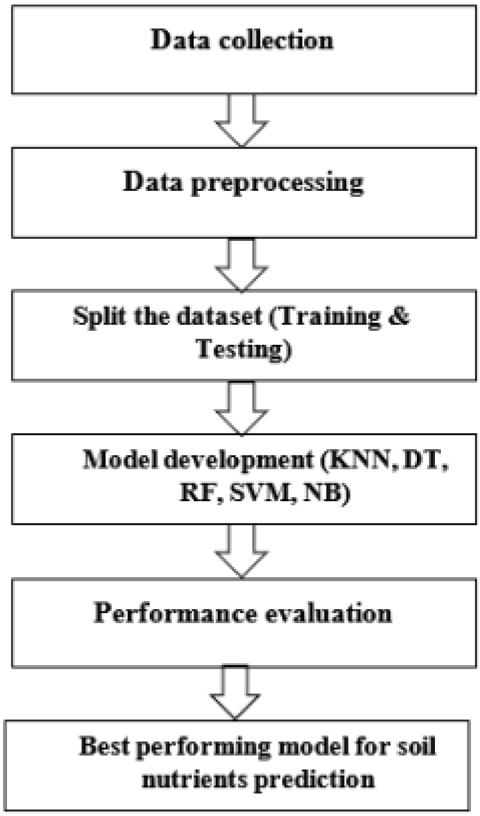
Key Words: K-Nearest Neighbour, Machine Learning, Naive Bayes, Prediction, Random Forest, Soil Nutrients, Support Vector Machine,

INTRODUCTION

Soil nutrients are vital for the healthy growth of plants. Maintaining adequate nutrient levels ensures optimal plant growth, high yields, and resilience against diseases and pests. Deficiencies or excesses in soil nutrients can compromise plant health and reduce agricultural output (Kumar, 2020). Prediction of soil nutrient status is the foundation of all input-based agricultural production systems. It is the initial and most crucial step, influencing several key agricultural decisions. These include the selection of suitable crops, the preparation of land, the choice of seeds, anticipated crop yields, and the appropriate application of fertilizers or manure (Sharma *et al,* 2020; Pandith *et al,* 2020). Farmers can make informed decisions that enhance crop production, conserve resources, and protect the environment by accurately assessing soil nutrient status.

Machine learning (ML) utilizes algorithms to analyse data, detect patterns, and make decisions, enabling systems to perform tasks similar to human activities. These systems learn and improve autonomously from experience without needing explicit programming. (Sarker, 2021; Abraham *et al,* 2024). With recent advancements in machine learning, the use of ML models is likely to become more widespread. Effectively used in support and decision-making, ML algorithms have the potential to revolutionize fields, enhancing efficiency and effectiveness (Folorunso Olusegun *et al,* 2023). Machine learning algorithms are primarily classified into supervised, unsupervised, and reinforcement learning paradigms. Supervised learning uses labelled datasets where each input has a corresponding output. These algorithms learn from this data to establish relationships and generalize predictions for new inputs. When applied to predict categorical values, they are known as classification algorithms (Sharma *et al,* 2020). It involves a two-step process: analysing dataset attributes to build a model with predetermined class labels, and estimating the model's accuracy. Before applying classification techniques, data cleaning, transformation, and

Corresponding Author's Email - gigiannee@gmail.com



relevance analysis are essential to refine the dataset for optimal model training and prediction. Classification model results are evaluated based on scalability, speed, predictive accuracy, robustness, and interpretability (Awasthi and Bansal, 2017).

Using machine learning techniques for soil nutrient assessment can optimize fertilizer application, predict pest and disease outbreaks, and suggest precise irrigation strategies. This enhances agricultural productivity and efficient land resource management (Folorunso Olusegun *et al*, 2023). Classification can be used for the prediction of soil nutrient status (Oumnia Ennaji *et al,* 2023). The main classification algorithms identified for soil nutrient prediction are k-Nearest Neighbours (KNN), Decision Trees (DT), Random Forest (RF), Support Vector Machine (SVM) and Naive Bayes (NB) (Oumnia Ennaji *et al,* 2023; Sharma *et al,* 2020).

MATERIALS AND METHODS

The dataset was collected from the soil testing laboratory at Krishi Vigyan Kendra in Jabalpur. It includes readings of 12 parameters, pH (soil pH value), EC (electrical conductivity), OC (organic carbon), N (nitrogen), P (phosphorus), K (potassium), S (sulphur), Zn (zinc), B (boron), Fe (iron), Mn (manganese), and Cu (copper) and corresponding labels indicating nutrient status categories, low, medium, high. After data collection, pre-processing was performed to eliminate duplicates, handle outliers, standardize attribute values, and substitute missing data. The dataset of 1000 soil samples has been divided into training and testing data in the ratio 70: 30. The proposed methodology is presented in Fig 1.

Classification algorithms suitable for soil nutrient status prediction are follows (Folorunso Olusegun *et al,* 2023; Oumnia Ennaji *et al,* 2023).

K-Nearest Neighbours (KNN)

The KNN algorithm is a flexible and easy-to-understand method that can be applied to both classification and regression problems. It works by selecting a parameter, k, that represents the number of nearest neighbours to consider. For classification tasks, KNN assigns a new data point to the category most common among its k-nearest neighbours, while a larger k tends to enhance stability. As a non-parametric algorithm, KNN uses the training data to group new points without

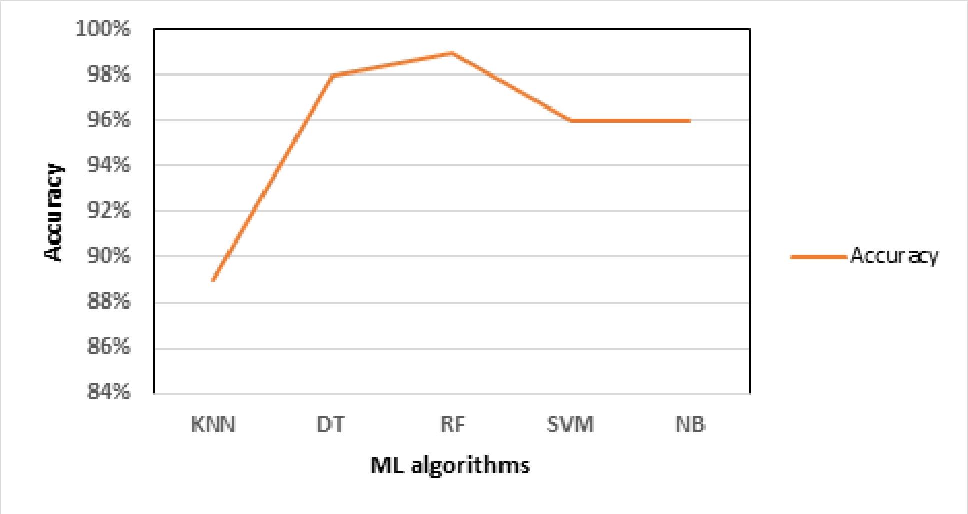


Fig 2: Evaluation for finding better accuracy

assuming any specific data distribution. It is effective for capturing both linear and non-linear patterns, adjusting to local data variations, and handling outliers well, making it suitable for predicting soil nutrient levels based on nearby data characteristics.

Decision Trees (DT)

DTs are versatile algorithms suitable for both classification and regression tasks, with a preference for classification problems. In a tree-structured classifier, internal nodes represent dataset features, branches symbolize decision rules, and leaf nodes signify outcomes. Decision Trees offer excellent clarity and interpretability, highlighting feature importance and handling non-linear relationships effectively. They work well with mixed data types (categorical and numerical), have low computational complexity, and manage outliers efficiently, making them ideal for predicting soil nutrient status based on various soil parameters.

Random Forest (RF)

RF is a powerful and very easy-to-use supervised learning algorithm known for its high accuracy. It creates a forest of decision trees, on various subsets of the given dataset. Even without fine-tuning, it often produces good results. RF is effective in dealing with outliers and noisy data, works well with large numbers of categories, highlights important variables, and avoids overfitting, making it a great option for predicting soil nutrient levels.

Support Vector Machines (SVM)

SVMs are powerful and flexible supervised learning algorithms used for both classification and regression tasks. They are popular because they can work with both categorical and continuous variables. SVMs separate data by creating a decision boundary, or hyperplane, between two classes. This hyperplane is adjusted to minimize errors and maximize the margin, which helps improve generalization. SVMs perform well in high-dimensional spaces, are less likely to over fit, and can use different kernel functions. They also work effectively with small to medium datasets and are not affected by irrelevant features, making them a good choice for predicting soil nutrient levels.

Naive Bayes (NB)

NB is a simple but powerful classification algorithm that builds fast machine learning models for quick predictions. It works as a probabilistic classifier, making predictions based on the likelihood of an object, using Bayes' theorem and assuming that features are independent. The algorithm estimates class labels by calculating the probability for each instance and requires only a small amount of training data. NBs is especially useful for datasets with multiple classes. It efficiently handles large datasets, scales well, and manages irrelevant features. It also works with limited training data and is easy to interpret, making it a good option for predicting soil nutrient levels.

RESULTS AND DISCUSSION

A comparative analysis was conducted to determine the most accurate technique among machine learning algorithms for prediction of soil nutrient status. Machine learning models were implemented using Python's scikit-learn library. The performance of each model was evaluated on pre-processed soil datasets with feature scaling applied where necessary. Fig 2 provides a summary of the evaluation results of each algorithm. Various machine learning methods were employed to determine the most accurate model based on soil data sets. The following machine learning methods were employed: K-Nearest Neighbour, Decision Tree, Random Forest, Support Vector Machine and Naive Bayes. The models were trained using soil nutrient data, with the data split into training and testing sets in a way that made sure all categories were fairly represented.

Classification tasks and evaluation performance metrics were used to test the accuracy of each machine learning model, yielding the following results: 89% for KNN, 98% for decision tree, 99% for random forest, 96% for support vector machine, and 96% for Naive Bayes. The random forest method achieved the highest accuracy. The random forest technique outperformed all other machine learning methods in predicting soil nutrients status, enabling farmers to make more informed decisions to increase productivity.

CONCLUSION

This study demonstrates the effectiveness of machine learning algorithms in predicting soil nutrient status. By utilizing a dataset of soil samples with 12 parameters were used to classify soil nutrients as low, medium and high and implementing supervised learning algorithms using Python, the research achieved a high accuracy rate of 99% with the random forest model. Accurate soil nutrient prediction can significantly aid farmers in making informed decisions regarding crop selection, land preparation, and fertilizer application, ultimately enhancing agricultural productivity and resource management. The success of the random forest algorithm underscores its robustness and reliability, making it a valuable tool for optimizing soil nutrient management practices.

REFERENCES

Awasthi N and Bansal A (2017). Application of data mining classification techniques on soil data using R. *Int J Adv in Electron and Comput Sci* **4** 33-37.

Abraham, Gigi Annee and Singh, A K (2024). Understanding the Applications of Artificial Intelligence and Drones in Agriculture. *J Krishi Vigyan* **12** (2): 424-428.

Folorunso Olusegun, Oyewole Adebayo, Adebayo Olaniyi and Fatimah Opeyemi (2023). Exploring Machine Learning Models for Soil Nutrient Properties Prediction: A Systematic Review. *Big Data and Cognitive Computing* **7**(2): 113.

Kumar Maneesh (2020). Soil Test Based Fertilizer's Recommendation on Yield, Soil Fertility Status and Economics of Rice Crop (*Oryza sativa L.*). *J Krishi Vigyan* **9**(1): 316-321

Oumnia Ennaji, Leonardus Vergütz and Achraf El Allali (2023). Machine learning in nutrient management: A review, *Artificial Intelligence in Agri* **9**: 1-11.

Pandith V, Kour H, Singh S, Manhas J and Sharma V (2020). Performance evaluation of machine learning techniques for mustard crop yield prediction from soil analysis. *J Scienti Res* **64(2)**: 394-398.

Sarker I H (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *S N Comput Sci* **2:** 160.

Sharma A, Jain A, Gupta P and Vinay Chowdary V (2020). Machine learning applications for precision agriculture: A comprehensive review. *IEEE Access* **9**: 4843-4873.

Received on 15/6/2024 Accepted on 22/82024