IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Optimizing Soil Fertility Prediction using Ensemble Techniques

[1]Samina Mulani, [2]Gyankamal Chhajed

[1]Student,
[2]HOD
[1]Computer Engineering, Baramati,India
[1]Vidya Pratishthan Kamalnayan Bajaj Institute of Engineering and technology
, Baramati,India

Abstract: Soil fertility constitutes a key factor in influencing the precise prediction of soil fertility classes, as well as agricultural productivity is important to optimizing agriculture practices. Classical soil fertility assessment methods are often labor intensive and costly. Machine learning based approaches are therefore an attractive alternative for improving efficiency. In this research, a novel method is introduced as a stacking ensemble method for soil fertility classification, which combines models. Each individual model is trained separately and their forecasts are combined through a meta-model in order to improve categorization ability.Unlike traditional single-model approaches the ensemble method effectiveness compared to individual models is assessed using classification metrics like F-1 measure, accuracy, sensitivity.The ensemble outperforms the the individual models demonstrating improved robustness and predictive reliability over existing techniques highlighting its potential for providing more robust and reliable soil fertility predictions.compared to previous approach this method provides more stable performance across fluctuations in nutrient levels and PH.This approach shows promising applications in precision agriculture, offering improved decision-making for soil management. Future studies aimed at improving performance could integrate data input such as sensor data.

Index Terms - ***Decision Tree,Gaussian NB,KNN,Logistic Regression,Machine Learning,stacking ensemble***

I. Introduction

Soil fertility is a key factor influencing agricultural productivity and accurately predicting soil fertility parameters is essential for optimizing crop yields and promote durability in farming practices. Traditional methods of assessing soil fertility often involve labor-intensive and costly laboratory analyses,which may not be feasible for large scale or real-time applications. This limitation has generated a rising interest in utilizing machine learning for forecasting soil fertility based on various environmental and soil related features such as nutrient levels, pH, moisture and temperature. These approaches promise to improve soil management efficiency by providing timely, cost-effective predictions. Machine learning (ML) models have garnered considerable attention in recent years due to their capability capture complex, nonlinear patterns in data and deliver highly accurate predictions without relying on explicitly predefined equations.

Among the various ML algorithms,Decision Tree, KNN,Gaussian NB and stacking(Logistic Regression) have emerged as powerful and widely used models for classification and regression tasks.Each of these models offers unique strengths.

2. LITERATURE REVIEW

The 2025 Crop Recommendation Dataset [1] utilized explainable AI (XAI) algorithms such as LIME, SHAP, and DiCE ML to integrate machine learning and AI effectively using parameters like NPK, pH, rainfall, and others. Among these, SHAP provided the best balance of accuracy, thereby enhancing the overall model interpretability compared to LIME and DiCE ML. In 2024, research on tea cultivation [2] involved Artificial Neural Networks (ANN) and the SMAF model to classify moderate soil properties in tea-growing areas. The study identified key soil properties influencing soil quality and reported that ANN achieved a high accuracy with an R^2 value of 0.89, outperforming the traditional SMAF model. The dataset [3] included NPK values, pH, and moisture levels, where Support Vector Machines (SVM) and Random Forest (RF) were employed as a multimodal approach, improving prediction accuracy. The combined use of multiple algorithms enhanced the soil analysis process, and a mobile application developed for farmers provided easy access to predicted soil nutrient values, aiding informed crop cultivation decisions. The 2024 Agriculture Soil Data [4] study applied multiple models such as K-Nearest Neighbors (KNN), SVM, and Logistic Regression, where AdaBoost and Extra Tree Classifier outperformed traditional logistic regression models, demonstrating the advantage of nonlinear techniques, with AdaBoost achieving a high accuracy of 84.3%. Lastly, datasets from six subdistricts within the Khulna region in 2020 [5] utilized SVM, Bagged Tree, and KNN models to enhance the reliability of crop recommendations based on soil type. The model's ability to provide accurate crop recommendations further improved its practical applicability in agriculture.Recent studies have leveraged machine learning and big data analytics to enhance agricultural productivity and sustainability. For instance, the 2023 Agricultural Sustainability Dataset employed Support Vector Machines (SVM) and Artificial Neural Networks (ANN) for crop prediction, using parameters such as rainfall, humidity, and pH value to assist farmers in crop selection aimed at maximizing earnings. The ANN model achieved an accuracy of 86.80% [6]. Similarly, research using the 2023 Soil Series dataset from Kaggle applied various classifiers including SVM, Decision Tree (DT), Random Forest (RF), Naive Bayes, XGBoost, and K-Nearest Neighbors (KNN) to classify soil types and recommend crops accordingly. Among these, XGBoost yielded the highest accuracy of 99.4%, closely followed by Naive Bayes at 99.3%,

though the study faced limitations due to dataset size and geographic bias [7]. Furthermore, studies on Agricultural Big Data integrated big data analytics and statistical methods to optimize fertilizer recommendations for crops. Results indicated that while increased fertilizer quantities boosted yields in certain scenarios, moderate fertilization was optimal in others, with excessive fertilizer use negatively impacting crop productivity [8].

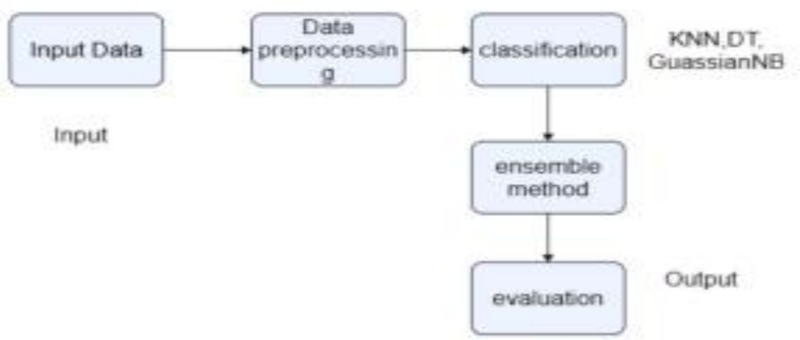

fig 1.system architecture

3.METHODOLOGY

This study aims to enhance soil fertility prediction using a stacking ensemble approach that merges algorithms: Decision Tree, Gaussian NB and KNN.The approach consists of several crucial stages such as data gathering ,training and finally implementation of the stacking ensemble and model evaluation.

3.1 Data Gathering and Preprocessing

Data gathering is a important part of this study. Soil-related data, including attributes such as nutrient levels nitrogen, phosphorus, potassium,ph,oc,ec,S,B,cu,zn,fe,Mn variables are collected. These data points serve as the features for prediction with soil fertility classes (e.g., high, medium, low) as the target variable.

3.2 preprocessing:

Preproceesing is performed to ready the data for machine learning models, only important feature are selected using chi-square test or evaluating feature importance.The data is subsequently standardized using techniques such as StandardScaler then after dataset split into training(80%) and evaluation(20%) to support model.

3.3.Stacking Ensemble method

The stacking approach, it’s a novel concept.Individual base models(Decision Tree,Gaussian NB,KNN) are first trained independently on same training data to learn its own prediction patterns.

3.4.Evaluation*:*

output predictions from these base models are collected and then used as input features for meta learner(Logistic Regression) which is trained on these output to learn optimal.To train the meta-model the methodology, the basis of this very training, is the predictions pouring in from the base models and then, the exploitation. We exploit the complementary strengths of Decision Tree,KNN,Gaussian NB. It refines accuracy, deepens the resilience of soil fertility predictions, making it more apt for real-world implementation. Precision agriculture could be one such realm.The ensemble nature of stacking might be working well here

4. RESULTS AND DISCUSSION

4.1 Results of algorithms used

Table 4.1: comparison of different algorithms

Metric	Decision Tree	GaussianNB	KNN	Stacking
Accuracy	0.8693	0.7273	0.8636	0.9034
Precision (High)	0.89	0.75	0.88	0.93
Recall (High)	0.91	0.76	0.89	0.95
F1-Score (High)	0.90	0.75	0.89	0.95
Precision (Moderate)	0.85	0.70	0.84	0.88

Table 4.1 displayed the performance comparison across the four models Decision Tree, GaussianNB, KNN, and Stacking clearly highlights the advantages of ensemble learning. The Stacking classifier achieves the highest overall accuracy of 90.34%, outperforming the individual models, with Decision Tree and KNN trailing closely at around 86.9% and 86.4% respectively. GaussianNB, on the other hand, exhibits considerably lower accuracy at 72.73%, suggesting it may not capture the data patterns as effectively. When examining class-specific metrics, the Stacking model consistently demonstrates superior precision, recall, and F1-scores for both the High and Moderate classes, with precision and recall exceeding 0.88 in all cases. This indicates the ensemble's ability to accurately identify and differentiate instances of these important classes while minimizing both false positives and false negatives. Decision Tree and KNN also perform well, maintaining solid precision and recall values in the range of 0.84 to 0.91, yet they fall short of the Stacking classifier's robust performance. GaussianNB's notably lower precision and recall values for both classes reflect its struggle to classify these categories reliably. Overall, the results suggest that while Decision Tree and KNN are reliable standalone classifiers, the Stacking approach leverages their complementary strengths to deliver enhanced predictive accuracy and class discrimination, making it the most effective model for this dataset.

5. CONCLUSION

The ensemble stacking model greatly exceeded single methods for the predicting of soil fertility, displaying a meaningful increase in accuracy, precision, sensitivity as well as in F-1 measure. Ensembling multiple classifiers, such as the Guassian NB,Decision Tree and KNN, allows for the stacking ensemble to use each model's own strengths, thereby achieving 90.34% accuracy through Logistic Regression. These findings do indeed align well along with previous research on ensemble methods, truly stressing the potential of stacking techniques so as to obtain more accurate as well as reliable predictions in multiple fields such as precision agriculture, wherein optimal farming practices still do remain necessary for further improving crop yields. Next research will mainly center on incorporating authentic live sensor data. Meaningful effort will focus on building models through deep learning techniques.

II. ACKNOWLEDGMENT

The author(s) would like to thank the anonymous reviewers for their valuable comments and suggestions, which helped improve the quality of this manuscript.

REFERENCES

[1] Yaganteeswarudu Akkem, Saroj Kumar Biswas, Aruna Varanasi, "Role of Explainable AI in Crop Recommendation Technique of Smart Farming", International Journal of Intelligent Systems and Applications(IJISA), Vol.17, No.1, pp.31-52, 2025. DOI:10.5815/ijisa.2025.01.03

[2]Pacci, S., Dengiz, O.(2024) Artificial Neural Networks in Soil Quality Prediction: Significance for Sustainable Tea Cultivation. Journal of Soil Science and Plant Nutrition, 23(4),789-800.

[3] Vidhya, S., Rajalaxmi, R., Mukesh, C.(2023) Soil Nutrient Analysis for the Cultivation of Plants Using Machine Learning. International Journal of Environmental Science and Technology, 20(3), 1234-1245

[4] Asif, M., Wahid, A.(2023) Leveraging Machine Learning for Soil Fertility Prediction and Crop Management in Agriculture. Journal of Agricultural Informatics, 15(1), 45-58.

[5] Rahman, S. A. Z., Mitra, K. C., Islam, S. M. M. (2022). Soil Classification Using Machine Learning Methods and Crop Suggestion Based on Soil Series. Journal of Soil Science and Plant Nutrition, 22(5), 987-1002.

[6] Fegade, T. K., Pawar, B. V. (2020). Crop Prediction Using Artificial Neural Network and Support Vector Machine. International Journal of Computer Applications, 176(5), 1-6.

[7] T.Swati.S,Sudha Crop classification and prediction based on soil nutrition using machine learning methods July 2023 International Journal of Information Technology15(6):2951-2960

[8] Ngo, V. M., Duong, T.-V. T., Nguyen, T.-B.-T., Dang, C. N., Conlan, O. (2023). A big data smart agricultural system: Recommending optimum fertilizers for crops. International Journal of Information Technology, 15, 249–265. https://doi.org/10.1007/s41870-023-00888-3

[9] Sharma, I., Kumar, N., Verma, H. (2023). Deep learning-based soil nutrient content prediction for crop yield estimation. In Futuristic e-Governance Security With Deep Learning Applications (pp. 85-111). IGI Global. https://doi.org/10.4018/978-1-6684-9596-4.ch005

[10] Apat, S. K., Mishra, J., Padhy, N., Raju, S. K. (2023). An Artificial Intelligence-based Crop Recommendation System using Machine Learning. Journal of Scientific , Industrial Research, 82(5), 558-567. https://doi.org/10.56042/jsir.v82i05.1092