Applying machine learning to soil analysis for accurate farming

L Venkateswara Reddy[1,*] , *D.*Ganesh[2], *M.* Sunil Kumar[2], Sreenivasulu Gogula[3], *M.* Rekha[4], *Archana* Sehgal[5]

[1]Department of CSE – Data Science, K G Reddy College of Engineering and Technology,Hyderabad, Telengana, India.
[2]Department of CSE,Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering College(Autonomous), Tirupati, Andhra Pradesh , India.
[3]Professor , CSE – Data Science, Vardhaman College of Engineering, Shamshabad, Hyderabad, Telengana , India
[4]Department of IT, GRIET, Hyderabad, Telangana, India
[5]Lovely Professional University, Phagwara, Punjab, India.

Abstract. A crucial component of agriculture is soil. Soil analysis is critical for optimizing agricultural practices and ensuring sustainable crop production. Traditional methods are often time-consuming and labor intensive, limiting their scalability and real-time applicability. The application of machine learning techniques in soil nutrient analysis has emerged as a promising solution. There is a lot of complicated soil data, but algorithms that use machine learning can handle it all, enabling accurate prediction and assessment of soil nutrient content. Integration with remote sensing technologies enhances the capabilities of soil nutrient analysis, allowing for rapid assessment at different scales. Machine learning facilitates personalized recommendations for fertilizer application, irrigation strategies, and soil amendments, tailored to the specific needs of individual fields or crops. The continuous learning and adaptive capabilities of machine learning models ensure up-to-date nutrient management recommendations. Challenges include the availability of representative training data and the interpretability of models. Nevertheless, the integration of machine learning in soil nutrient analysis offers improved resource utilization, enhanced crop productivity, and sustainable soil management practices. Ongoing research and collaboration with domain experts will further advance the application of machine learning in this field.

Keywords. Soil type, Crop type, Pesticides usages, Number of doses, Number of weeks, Number of weeks quit, Season, Yield damage.

* Corresponding author: logeshwarisairam@gmail.com

1 Introduction

As a major source of both finished goods and food, agriculture is an essential part of any economy. However, increasing global population and changing climate conditions pose a significant challenge to agricultural productivity. To address this challenge, precision farming techniques have been developed to optimize crop production and reduce environmental impact. Precision farming involves using technology[2-5] and data-driven approaches to optimize crop production by ensuring that each plant receives the right amount of nutrients, water, and other resources. Soil analysis plays a critical role in precision farming as it provides information on the soil's characteristics, including its nutrient content, pH levels, and water holding capacity. Decisions on fertilisation, irrigation, and other aspects of crop management are based on this data. Soil testing the old-fashioned way is labor-intensive, costly, and takes a long time. Furthermore, due to the small sample size of the soil, the results are not necessarily reliable. Machine learning algorithms may analyze massive volumes of soil and environmental information to circumvent these restrictions to accurately predict the appropriate crop for a particular spatial land. The use of machine learning algorithms can lead to more precise recommendations for crop management practices, resulting in higher yields and improved efficiency[6-9]. In this project, we propose the development of a prediction model for soil analysis in precision farming using machine learning. The prediction model will be based on decision tree and random forest algorithms, which will be trained on environmental and soil data. The model will accurately predict the appropriate crop for a particular spatial land, leading to more precise recommendations for crop management practices. The proposed system has the potential to significantly improve crop yields, reduce costs, and promote sustainability in agriculture. Soil analysis plays a crucial role in precision farming, as It offers helpful information on soil properties that can improve farming methods. Traditional soil analysis methods involve manual collection of soil samples and subsequent laboratory testing, which can be time-consuming and labor-intensive. However, with the advancements in machine learning, it is now possible to leverage this technology to streamline and enhance soil analysis processes.

Machine learning algorithms can sift through mountains of data in search of patterns and draw conclusions. Machine learning has the potential to revolutionize soil research by sifting through mountains of data on physical and chemical qualities and revealing trends, patterns, and correlations that human analysts would miss. By harnessing the power of machine learning[10-11], precision farming can achieve a more comprehensive and efficient soil analysis approach. This paper's goal is to investigate how precision farming might benefit from soil analysis using machine learning approaches. By using machine learning algorithms, we aim to develop models that can accurately predict soil characteristics and provide actionable insights for farmers. This can help optimize crop yield, improve resource management, and minimize environmental impact. In this paper, we will first discuss the importance of soil analysis in precision farming and the challenges associated with traditional methods. We will then delve into the fundamentals of machine learning and highlight its potential in revolutionizing soil analysis. Next, we will explore various machine learning techniques commonly used in soil analysis, such as supervised learning, unsupervised learning, and machine learning. We will discuss how these techniques can be applied to soil data and the advantages they offer over traditional methodsThere are two main applications of cipher transformations in bio-cryptosystems. The first is the security and protection of biometric templates. The key sequence utilized as a cipher in traditional cryptographic techniques is sometimes generated using distinctive biometric features. Several transformation activities have been studied in different situations to enhance system performance and dependability [15] addressed the root causes of the system's vulnerability to attacks by encrypting biometric templates using a methodology that generates two chaotic maps.

Furthermore, we will discuss the data requirements for training machine learning models, including the types of soil data collected, preprocessing techniques, and feature selection. We will also explore the potential sources of soil data, such as remote sensing, soil sensors, and historical databases. Additionally, we will address the challenges and limitations of using machine learning in soil analysis, such as data quality, model interpretability, and scalability. Finally, we will present case studies and examples of successful implementations of machine learning in soil analysis for precision farming. These real-world applications will demonstrate the effectiveness and benefits of using machine learning models in improving soil management practices and agricultural productivity. We will conclude with a discussion of future directions and potential advancements in machine learning techniques that can further enhance soil analysis and contribute to sustainable agriculture.

2 Literature Survey

In this work of Soil Classification Utilising Machine Learning Techniques and Crop Suggestion Using a Soil Series. For farmers, soil is an essential component. Soil comes in a wide variety of types [12]. Soil types affect crop yields because certain soils are more suited to growing certain crops. To determine which crops thrive in certain soil types, we need to understand the characteristics and functions of various types of soil. It has made significant development in the last few years. Machine learning is still a challenging and developing topic of research in statistics about agriculture analysis. Our research suggests a version that may predict soil series according on land type and, if accurate, can recommend crops that would thrive in those conditions. There are a number of device mastering techniques employed, including weighted Okay-Nearest Neighbor (okay-NN), Bagged trees, and Gaussian kernel-based algorithms for soil category[13].

In the work of Comparison of machine mastering algorithms for soil kind type For automated soil type, a set of rules that the machine studies may be used[14]. This study analyses several system learning methods for identifying different types of soil. For this category, algorithms using the assist vector machine (SVM), the neural network selection tree, and naive Bayesian are suggested and evaluated. The soil dataset is derived from the real data. Rapid Miner Studio is used to execute the simulation. The correctness is demonstrated by the performance. As a consequence, SVM with a linear feature kernel performs better than alternative methods. The SVM's fine accuracy is 82.35 percent. In the work of improved prediction of clay soil growth using device gaining knowledge of The calamity caused by soil swelling is regarded as one of the most destructive geo-hazards in recent memory. Soil expansion potential must be properly dedicated in order to achieve a secure and safe footing for infrastructures. As a result, this work has provided a unique and intelligent method that makes use of kernelized machines to provide an enhanced assessment of swelling. More than one linear regressor (LR), Bayesian linear regression[15], Bayes point machine (BPM), support vector machine (SVM), and tree-based methods, such as choice forest. Additionally, meta-heuristic classifiers including the vote casting (VE) and stacking (SE) procedures have been used for the first time. Investigated are various, unbiased combinations of explanatory functions that impact soil behavior during swelling. Last but not least, it is encouraged that the concepts put forth here be applied throughout the earliest stages of a geotechnical and geological web site classification with the help of the best functioning meta-heuristic models by their background coding assistance.

Soil Attribute Prediction and Soil Analysis of Data Using Classifier Techniques has been useful for agricultural research from technological advancements, automation, and data mining. Although there are many off-the-shelf and domain-specific statistics mining machine products and services available today, and statistics mining is employed in many other fields, agricultural soil datasets represent a relatively new study topic[16]. The enormous volumes of data that can currently be plainly obtained from the side of plants need to be analyzed and utilised fully. This study uses facts mining techniques to analyze a soil dataset. It specializes in classifying soil using a variety of algorithms. Predicting unknown properties through the use of regression analysis and automated soil sample categorization is another crucial goal.

3 Methodology

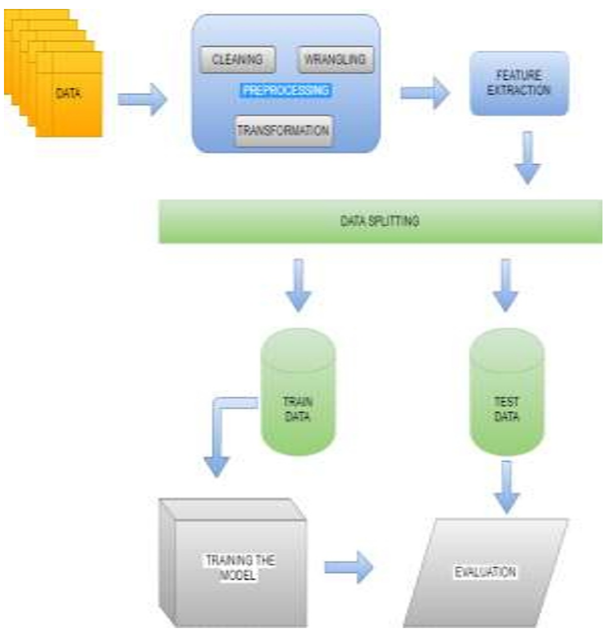

Fig. 1. System Architecture

The architecture for soil analysis using machine learning typically involves several components and steps. Here's a high-level overview of possible system architecture. A probabilistic technique called Support Vector Classifier is based on the idea that characteristics are independent of one another. Despite being straightforward, Support Vector Classifier has produced results that are competitive in Soil recommendation systems, averaging an accuracy of 89%. This method offers quick predictions and is especially effective at handling massive datasets.

3.1 Data Collection

Collect soil samples from different locations, along with their corresponding soil type. Additionally, measure other relevant features such as Soil type, Crop type, Pesticides usages, Number of doses, Number of weeks, Number of weeks quit, Season, Yield damage.

3.2 Data Preprocessing

Clean the records by way of doing away with any outliers or mistakes. Handle missing values through strategies like imputation or deletion. Normalize or standardize the enter capabilities to make sure they're on a comparable scale.

3.3 Feature Selection

Assess the significance of every characteristic in predicting soil series. You can use strategies like correlation analysis, characteristic significance from tree-primarily based totally models, or domain understanding to choose out the maximum applicable capabilities.

3.4 Model Selection

Choose the ideal system learning set of rules for regression obligations. For this situation, we'll select a random wooded area regression set of rules, regarded for its capability to deal with complex relationships and offer function importance rankings.

3.5 Model Training

Divide the data set into two parts: instructional and experimental. Use the education set to teach the random wooded area regression version at the enter capabilities and corresponding nutrient levels. The model will study the styles and relationships between the functions and nutrient values.

3.6 Model Evaluation

Evaluate the overall performance of the trained model the usage of the trying out set. Calculate metrics which includes imply squared blunders (MSE) or R-squared to evaluate how properly the version predicts the soil nutrient.

3.7 Prediction and Visualization

Use the trained model to make predictions on new, unseen soil samples. Compare the predicted pH values with the actual pH values. You can visualize the results by creating a scatter plot with the actual pH values on one axis and the predicted pH values on the other axis. This will help you understand how well the model performs and identify any patterns or discrepancies. Remember that this example focuses on predicting soil Value mapping shows that which value is belongs to Good Soil and Soil Affect some reasons then Soil Affect (pesticides). It helps in easy reading the predicted value.

The trained models will be integrated into a user-friendly software application or an online platform. Users will be able to input their soil sample data, and the system will provide comprehensive analysis results. This may include predictions on soil properties such as Soil type, Crop type, Pesticides usages, Number of doses, Number of weeks, Number of weeks quit, Season, Yield damage. The system can also generate visualizations or reports to aid in data interpretation and decision-making.

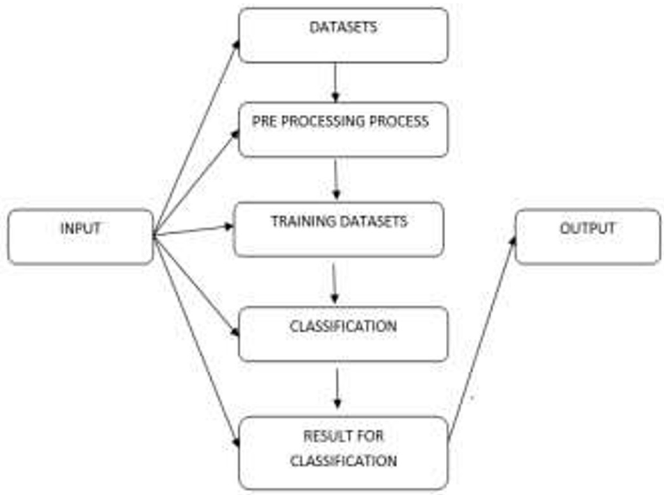

Fig. 2. Proposed System Data Flow

4 Dataset Collection

To Implement a System for Soil Analysis Using Machine Learning it will required datasets, the dataset can contain more accuracy to predict the most suitable Soil for the crop. Furthermore, we will discuss the data requirements for training machine learning models, including the types of soil data collected, preprocessing techniques, and feature selection. We will also explore the potential sources of soil data, such as remote sensing, soil sensors, and historical databases. Additionally, we will address the challenges and limitations of using machine learning in soil analysis, such as data quality, model interpretability, and scalability.

4.1 Soil Dataset

The Soil dataset contains a variety of parameters, including soil type, crop type, pesticide usage, the number of doses, number of a few weeks, number of weeks stopped, season, and yield damage. The data sets can be utilised to build models using machine learning that can advise farmers on the best crops to grow on their property, taking into account these parameters.

4.2 Data Pre-processing for dataset

In this, the most important and time-consuming step in any machine learning project. Pre-processing involves data cleansing, categorical data categorization, and variable correlation testing. in order to maximize accuracy. Missing values are filled using methods converting values in a given range, and cleaning the data. Using the NumPy and learn packages, the obtained datasets were pooled and cleaned to eliminate any duplicate columns or NA values.

5 Conclusion

In conclusion, soil analysis using machine learning algorithms offers valuable insights and predictive capabilities for understanding soil characteristics and properties. By leveraging machine learning techniques, we can extract patterns and relationships from soil data, allowing us to make accurate predictions and gain a deeper understanding of soil behavior. We can create reliable models that can predict a variety of soil variables, including soil type, crop type, pesticide usage, number of dosages (fertilisers), season, yield damage, and more through the processes of data collection, processing, selecting features, model training, and assessment. These models can offer useful knowledge for environmental research, land management, and agricultural practices. Support vector machines and random forest regression are two machine learning techniques that can handle the complicated correlations seen in soil data. The models gain knowledge from the input characteristics and the associated soil qualities, which enables them to make precise predictions on unobserved data. Graphs and charts may be used to visualise the outcomes of soil analysis applying machine learning, providing a clear comparison between projected values and actual data. These visualizations offer a thorough analysis of the model's performance as well as can support decisions about soil management and the improvement of agricultural practices. Overall, soil analysis using machine learning presents a powerful approach to enhance our understanding of soil properties, enabling us to make informed decisions for sustainable agriculture, environmental conservation, and land use planning. By leveraging the capabilities of machine learning, we can unlock the potential of soil data and drive advancements in soil science and related fields.

Where MSE denotes the mean squared error.
Based on separate parametric measurements such as brightness (μ) and contrast (σ), the structurally similarity index (SSIM) is employed to investigate the structural similarity of images in the following way:

References

1. J. Gholap, A. Ingole, J. Gohil, S. Gargade, and V. Attar, (2012). Soil data analysis and attribute prediction utilising classification techniques. arXiv:1206.1557 is an arXiv preprint.
2. Sofianita Mutalib, S-N-Fadhlun Jamian, Shuzlina AbdulRahman, and Azlinah Mohamed, all in (2010). Soil Classification Using a Self-Organizing Map and k-means
3. Dr. D.V. Padole, Prachi Sharma, (2017). Soil Analyzer Design and Implementation Using IoT.
4. S.Pudumalar, E.Ramanujam, R.Harine Rajashree, C.Kavya, T.Kiruthika, and J.Nisha. Precision Agriculture Crop Recommendation System.
5. Avinash Kumar, Sobhangi Sarkar, and Chittaranjan Pradhan. Crop Identification and Pest Control Technique Recommendation System in Agriculture.
6. Nidhi H Kulkarni, G N Srinivasan, B M Sagar, and N K Cauvery (2018). Improving Crop Productivity Using an Ensembling Technique for Crop Recommendation.
7. Meiqin Zhang, Shanqin Wang, Shuo Li, Jing Yi, and Peng Fu. Prediction and Map-making of Soil Organic Matter of Soil Profile Based on Imaging Spectroscopy: A Case in Hubei China.

8. Parin M. Shah (2012). Support Vector Machine is used to detect faces in images.
9. Hortle, James O., (2018). Alzheimer's Disease and Support Vector Machines: A Machine Learning Overview.
10. Heejong Suh, Daehyon Kim, and Changsoo Jang (2018).
11. Rimona Palas and Amos Baranes, (2019). Prediction of Earning Movement Using Machine Learning-Support Vector Machines (svm).
12. Ganesh, D., et al. "Improving Security in Edge Computing by using Cognitive Trust Management Model." *2022 International Conference on Edge Computing and Applications (ICECAA)*. IEEE, (2022).
13. Davanam, Ganesh, T. Pavan Kumar, and M. Sunil Kumar. "Novel Defense Framework for Cross-layer Attacks in Cognitive Radio Networks." *International Conference on Intelligent and Smart Computing in Data Analytics: ISCDA 2020*. Springer Singapore, (2021).
14. Ganesh, D., Kumar, T. P., & Kumar, M. S. (2021). Optimised Levenshtein centroid cross-layer defence for multi-hop cognitive radio networks. *IET Communications*, *15*(2), 245-256.
15. Davanam, Ganesh, T. Pavan Kumar, and M. Sunil Kumar. "Efficient energy management for reducing cross layer attacks in cognitive radio networks." *Journal of Green Engineering* 11 (2021): 1412-1426.
16. Venkateswara Reddy, L., Davanam, G., Pavan Kumar, T., Sunil Kumar, M., Narendar, M. (2023). Bio-Inspired Firefly Algorithm for Polygonal Approximation on Various
17. Shapes. In: Rao, B.N.K., Balasubramanian, R., Wang, SJ., Nayak, R. (eds) Intelligent Computing and Applications. Smart Innovation, Systems and Technologies, vol **315**. Springer, Singapore