Soil Analysis Using Deep Learning for Precision Agriculture

Thierry Forsack Fotabong
Department of Computer Science and Applications
Sharda School of Engineering and Technology
Sharda University
Greater Noida, India
thierryforsack@outlook.com

Abdullahi Muhammad
Department of Computer Science and Applications
Sharda School of Engineering and Technology
Sharda University
Greater Noida, India
Muhammadabdullahi7976@gmail.com

Murphy Keisham
Department of Computer Science and Applications
Sharda School of Engineering and Technology
Sharda University
Greater Noida, India
k9keisham@gmail.com

Tushar Mehrotra
Department of Computer Science and Engineering
School of Computing Science and Engineering
Galgotias University
Greater Noida, India
Tusharmehrotra9@gmail.com

Rajneesh Kumar Singh
Department of Computer Science and Applications
Sharda School of Engineering and Technology
Sharda University
Greater Noida, India
kunwar.rajneesh@gmail.com

S. Pratap Singh
Symbiosis Institute of Technology Nagpur Campus
(Deemed University)
Pune, India
profdrsprataps@gmail.com

Arun Prakash Agrawal
Department of Computer Science and Applications
Sharda School of Engineering and Technology
Sharda University
Greater Noida, India
arunpragrawal@gmail.com

Abstract— This research explores the application of deep learning techniques for soil analysis in precision agriculture, focusing on the classification of soil types using convolutional neural networks (CNNs). The study addresses the limitations of traditional soil classification methods, which are often labor-intensive and reliant on expert judgment. The proposed approach involves collecting soil image data, preprocessing it using techniques such as Gabor filtering, and classifying the images with a CNN model. The methodology includes data acquisition, feature extraction, and model training, aiming to accurately predict soil parameters using diverse input data sources like sensor readings and satellite imagery. The potential benefits of this approach include real-time monitoring, optimal fertilizer application, and enhanced crop management, which can contribute to sustainable agricultural practices and food security.

Keywords: Soil Analysis, Image classification, Convolutional Neural Network, Deep Learning, Precision Agriculture.

I. Introduction

The use of deep learning methods for soil analysis in precision agriculture is covered in this research study. It emphasizes how crucial it is to comprehend soil characteristics to manage crops and produce as much as possible [1]. Although deep learning algorithms are capable of extracting complicated patterns from big datasets with efficiency, traditional soil analysis approaches are labor-intensive and imprecise. We described the approach, which included gathering data, preprocessing, creating the model, and validating it. The capacity of convolutional neural networks to handle spatial and temporal input is being investigated [2]. The goal of the models is to forecast soil parameters with high accuracy using a variety of input data, including sensor data, satellite imaging, and historical records.

The potential advantages of deep learning in soil analysis are highlighted in this research, including real-time monitoring, erosion avoidance, and optimal fertilizer application. This will entail comparing several machine learning algorithms and conventional soil analysis methods across a variety of soil

attributes and geographical locations [3]. The performance trade-offs between model complexity, accuracy, and computing efficiency will also be looked at in this research. It also covers issues like ethical concerns, interpretability, and data quality. The study also examines potential collaborations with other cutting-edge technologies including blockchain, IoT, and UAVs [4].

Furthermore, the research paper examines the economic and policy implications, as well as the importance of tailoring solutions to local conditions and stakeholder needs. It highlights the need for knowledge transfer, capacity building, and inclusive approaches to ensure equitable access to these technologies [5].

Deep learning is portrayed in the paper as a revolutionary method for precision agriculture that may solve issues with food security and the environment while facilitating data-driven decision-making, sustainable practices, and higher production [6].

II. Literature Review

Given that different soil types have an impact on crop yields, the study on soil classification and crop recommendation emphasizes the importance of soil in agriculture. It is suggested to use deep learning algorithms to forecast soil data and suggest appropriate crops [7]. Soil classification involves the use of several techniques, such as Gaussian kernel-based algorithms, Bagged Trees, and Weighted K-Nearest Neighbour (K-NN). Machine-learned criteria are used for automatic soil typing in another study comparing deep learning algorithms; Support Vector Machine (SVM) demonstrates good accuracy [8]. Determining which crops do best in a given set of soil conditions requires an understanding of the characteristics of

979-8-3503-6501-6/24/$31.00 ©2024

various soil types. In this sense, machine learning techniques have become useful tools that provide forecasts and insights to support agricultural decision-making. Furthermore, research focuses on applying machine learning approaches to anticipate the growth of clay soil.

Regression analysis is used to forecast soil attributes, categorize soil, and analyze soil information using data mining approaches. Statistical models, image processing methods, and hyperspectral data are used in other research to investigate soil categorization. The algorithm and data mining are used to get accurate results. The algorithm will suggest crops that are profitable and beneficial to the farmer's kind of soil. The model's accuracy while using the Naive technique for prediction was 75%. [9]. The significance of comprehending soil properties for forecasting agricultural yields is highlighted, and machine learning techniques like SVM, Random Forest, and Naive Bayes are utilized to classify crops and soil. Data mining techniques are often used to suggest acceptable crop recommendations based on soil parameters [10]. Large volumes of data are transformed into innovations and meaningful insights that farmers may use thanks in large part to data mining. Farmers and other decision-makers may benefit from the readily available wealth of data by using it to get insightful knowledge that will help them make timely and efficient decisions. In addition, this study looks at how crops are affected by climate change and uses statistical models to forecast how crops will react to different weather patterns [11]. These models help stakeholders and policymakers create adaptive strategies for agricultural resilience by illuminating the intricate relationships between climate factors and crop performance.

III. PROBLEM STATEMENT

Lower agricultural yields are the consequence of the inaccuracy and inefficiency of traditional soil analysis techniques. To properly evaluate soil qualities, a system that smoothly combines image processing and deep learning approaches must be developed. The ultimate goal is to provide farmers with practical knowledge so they can make decisions that will increase crop yield and agricultural sustainability.

Through the application of cutting-edge technologies like deep learning and image processing, researchers want to build a solid platform that offers suggestions in real time that are customized to the unique requirements of farmers. With this method, farmers can monitor soil health more effectively and carry out focused interventions to maximize agricultural production while reducing environmental impact. It also expedites the process of soil analysis.

IV. PROPOSED ARCHITECTURE

Deep neural networks will be used in the proposed system to assess different soil types and offer farmers ideas on how to increase crop output. A soil organization framework, represented by a square chart, is presented, the main objective of which is to collect particular kinds of pictures from soil tests. This framework takes into account variables including soil properties, user requirements, data accessibility, and budgetary and temporal restrictions.

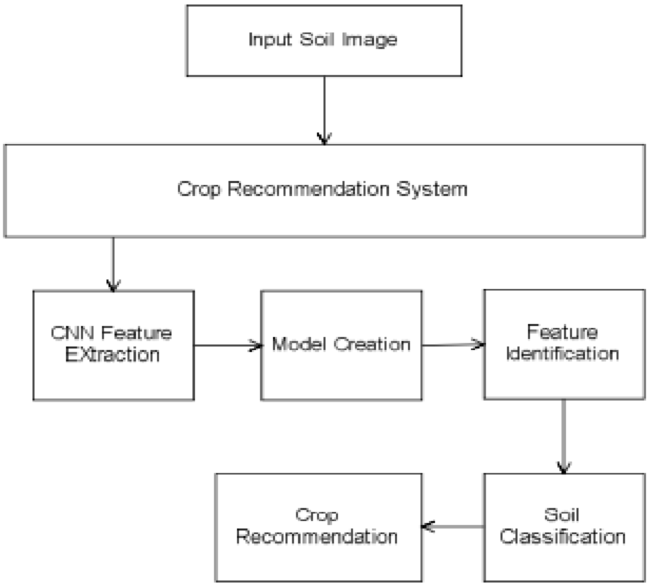

Fig. 1. *Proposed Architecture*

Convolutional Neural Networks (CNNs) require a sufficient number of training samples for image classification; these samples are usually collected by fieldwork, taking into account variables like spatial resolution and data complexity. Support vector machines (SVM) are used to generate agricultural recommendations after successful categorization. This helps farmers sell products online and make educated selections, which is especially helpful during a pandemic.

The Soil and Crop datasets are used in the training and testing phases of the technique. The four parts that make up the CNN model architecture are input, capture, output, and classification. The three processing components are CNN Learning and Visualisation, sample test selection, and configurations and initialization. While CNN Learning and Visualisation investigate picture attributes across layers to efficiently interpret model performance, configurations, and initialization determine fundamental model properties.

V. METHODOLOGY

This technique captures a well-designed framework that supports our soil analysis research activities. This detailed blueprint takes the reader through the complex terrain of soil analysis, covering a wide range of topics from the beginning of image collecting to the end of model validation and assessment. Every aspect of this approach has been carefully considered to provide practitioners and scholars with the knowledge and resources they need to understand the mysteries surrounding soil analysis. This approach develops in three separate but related phases:

Stage A: Image acquisition and pre-processing comprise.

This preparatory phase establishes the framework for the study that follows. It involves gathering a representative and varied collection of soil photographs from various platforms and pre-existing soil databases. The obtained photos are subjected to a rigorous pre-processing routine that includes a variety of methods for enhancement, noise reduction, standardization, and normalization. Strongly discriminative characteristics may be extracted from the soil dataset using deep learning techniques such as transfer learning or fine-tuning pre-trained models. The preparation of training data completes Stage A. For supervised learning, segmented soil particles must be labeled with the proper class labels (soil type, texture, composition). To guarantee appropriate model assessment, the labeled dataset is divided into subgroups for training, validation, and testing. This ensures that each subset contains representative samples. Class imbalance problems that might negatively impact performance are mitigated by balancing class distributions by oversampling or data augmentation. Normalization or feature scaling guarantees uniform ranges for various feature kinds. To ensure optimal accuracy and relevance in the subsequent phases of the workflow, these pre-processing processes are crucial in setting up the raw pictures for downstream analysis.

Stage B: Segmentation and Feature Extraction Techniques

Using advanced algorithms to separate soil regions of interest from background components, Stage B builds on the pre-processed pictures by focusing on segmentation and feature extraction. A deep learning architecture should be selected based on the available computing resources and the difficulty of the task. Convolutional neural networks, or CNNs, operate well for soil analysis and are frequently employed for image classification applications. Segmentation techniques such as thresholding, edge detection, and Gabor filter are used to isolate the salient features indicative of soil composition and texture. Subsequent feature extraction endeavors draw upon a rich repertoire of methodologies ranging from traditional color histograms and texture descriptors to cutting-edge techniques such as Gabor filters and Law's Texture Energy Measures. Once the architecture is selected, implement the deep learning model using frameworks like TensorFlow, and Keras. Define appropriate loss functions (cross-entropy for classification) and optimization algorithms (Adam, SGD). Train the model on the training dataset while monitoring validation set performance to prevent overfitting. Techniques like early stopping, learning rate scheduling, or checkpointing enhance convergence and generalization. After training, evaluate the model's performance on the test set using metrics like accuracy, precision, recall, F1- score, and confusion matrix. Visualize and interpret predictions, activations, and feature importances for deeper insights into decision-making. Model refinement and ensemble methods improve performance. Analyze misclassified samples to identify reasons for errors like ambiguous features, class imbalance, or data quality issues. These extracted features serve as the cornerstone for subsequent classification endeavors, encapsulating the intrinsic characteristics pivotal for discerning between different soil types.

Stage C: Model Training and Evaluation

The final stage of the methodology unfolds in Stage C, where the preparatory groundwork laid in the preceding stages coalesces into a cohesive framework for model training and evaluation. This stage encompasses the preparation of training data, the selection and fine-tuning of appropriate deep-learning model architectures, and the iterative cycles of model training and validation. Leveraging state-of-the-art deep learning frameworks and methodologies, we embark on a journey of discovery, fine-tuning model parameters, optimizing hyperparameters, and evaluating model performance against stringent validation criteria. Analyze and interpret results, drawing insights and conclusions relevant to soil analysis and management practices, advancing the field. These learned features are then progressively abstracted and combined across successive layers of the network, ultimately culminating in high-level representations that facilitate accurate classification decisions. Moreover, CNNs are adept at capturing both local and global spatial dependencies within the input image, enabling them to effectively differentiate between objects, shapes, and textures across varying scales and orientations. The iterative refinement process iterates until a model of requisite accuracy and generalization prowess is attained.

VI. Algorithm

1. Load training and test data directories
2. Load necessary libraries and modules
3. Mount Google Drive
4. Create image datasets for training and validation
5. Data preprocessing tools (image data generators)
6. Load training and testing data into Python
7. Define model architecture parameters (number of convolutional layers, dense layers, layer size, number of epochs)
8. Initialize sequential model
9. Add layers with activation and max pooling
10. Add a flattened layer to reduce dimensionality
11. Add dense layers with activation and dropout
12. Add output layer with softmax activation
13. Compile the model with appropriate loss function, optimizer, and metrics
14. Train model using prepared training, validation data
15. Save the trained model
16. Plot training and validation loss over epochs
17. Define the function to make predictions on a single image:
    a. Load the image
    b. Preprocess the image
    c. Make a prediction using the trained model
    d. Extract the true label from the image file path
    e. Output the predicted label, true label, and whether the prediction is correct
18. Check if the test image file exists
19. Load the test image

20. Call the prediction function with the test image

21. Define a function to split images into smaller patches:

a. Iterate over image directories

b. Read each image

c. Split the image into smaller patches of specified size

d. Save the patches as separate image files

22. Create a directory to divide images (if it doesn't exist)

23. Load the trained model

24. Define a function to classify images:

a. Load the image

b. Resize the image

c. Iterate over smaller patches of the image

d. Classify each patch using the trained model

e. Count the occurrences of each class

f. Calculate the proportion of each class

g. Return the proportions

25. Function to display the classification percentages:

a. Call the classify_images function with the given image

b. Load and display the image

c. Print the percentage of each class and the time taken for classification

26. Call the classify_percentage function with the test image

VII. RESULT AND DISCUSSION

In the proposed model, the classification process involved a total of 628 samples of different soil images, among which 70% of the samples were used for training 20% samples were used for training, and 10% for validation. Processing is done on the soil images, and features are extracted using the Gabor filter and these features are classified using our convolutional neural network (CNN). We trained our model and stored the training history to find how much percentage accuracy we have. We visualized the validation loss vs the epoch, which helped us understand the model's performance and find potential overfitting or underfitting issues and finetuned them appropriately. The model takes the dataset as an input, preprocess it, and outputs the classification that the model predicted and the true soil type with an accuracy of 100% when the image is used as an input to the model for validation.

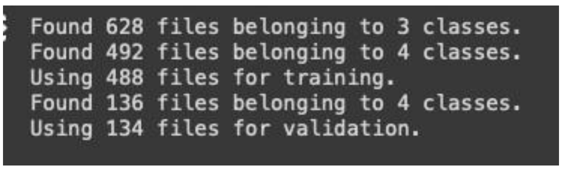

Fig. 2. *Data processed*

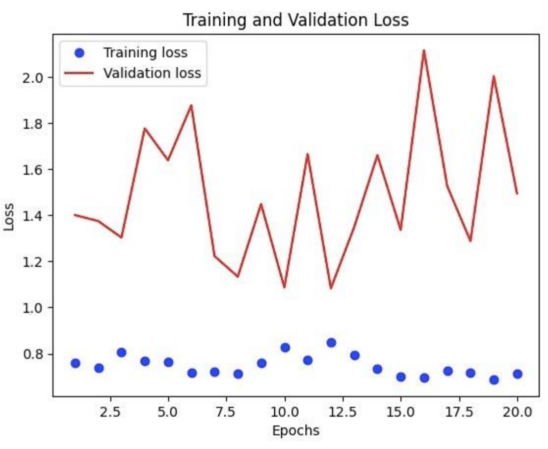

Fig. 3. *Graph showing Validation loss vs Epoch.*

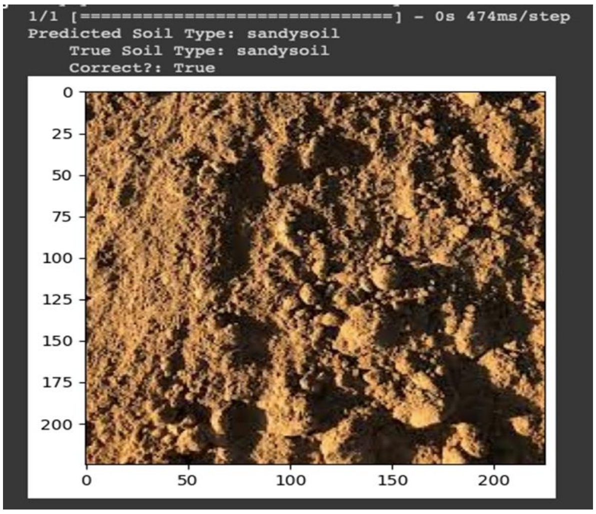

Fig. 4. *Accuracy for the validation.*

VIII. CONCLUSION

In conclusion, Soil classification is a growing research area in the current era. Many studies have been performed and some of them have provided similar techniques and some provided different techniques to deal with the issues, including rule-based, statistical, and traditional learning. However, Convolutional Neural Networks (CNNs) are utilized to produce efficient soil categorization. This work explores the use of several collections of soil pictures to automate the job of soil categorization. The procedure entails obtaining these photos, preparing them using methods such as color moments, and then using Gabor filters to extract characteristics. Using a CNN classifier is the last stage in achieving precise classification. When the CNN classifier's output is compared to other techniques like artificial neural networks (ANN) and linear support vector machines (SVM), it shows a notably higher degree of precision.

IX. FUTURE SCOPE

A promising direction for soil analysis research in the future is provided by the suggested sophisticated method of classifying soil using feature extraction and image preprocessing. This approach may be improved further to increase classification efficiency and accuracy. Future work should look at more image pre-processing methods, combine multi-modal data fusion for thorough soil characterization, and integrate hybrid machine learning models like CNNs with other algorithms. Furthermore, looking at deep learning architectures other than CNNs might offer fresh insights for enhancing classification techniques. Model performance is heavily dependent on parameters like layer_size, which affect training speed and accuracy. Fractional predictions are produced via iterative classification, providing information about the distribution of soil composition. These sophisticated methods might identify geographical and temporal correlations in soil data, producing more complex classification results.

REFERENCE

[1] G. K. Harikaran, "Machine Learning Approaches for Predicting Soil Properties of Vemagal Hobli, Kolar District, Karnataka," Ph.D. dissertation, Univ. Agric. Sci., GKVK, Bangalore, India, 2022

[2] W. Zhang, C. Wu, H. Zhong, Y. Li, and L. Wang, "Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization," Geoscience Frontiers, vol. 12, no. 1, pp. 469-477, 2021.

[3] L. Benos, A. C. Tagarakis, G. Dolias, R. Berruto, D. Kateris, and D. Bochtis, "Machine learning in agriculture: A comprehensive updated review," Sensors, vol. 21, no. 11, p. 3758, 2021

[4] J. Padarian, B. Minasny, and A. B. McBratney, "Machine learning and soil sciences: A review aided by machine learning tools," Soil, 2021.

[5] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," Frontiers in Plant Science, vol. 7, p. 215232, 2023.

[6] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: A review," Sensors, vol. 18, no. 8, p. 2674, 2018.

[7] N. Pavón-Pulido, J. A. López-Riquelme, R. Torres, R. Morais, and J. A. Pastor, "New trends in precision agriculture: A novel cloud-based system for enabling data storage and agricultural task planning and automation," Precision Agriculture, vol. 18, pp. 1038-1068, 2017.

[8] L. Benos, A. C. Tagarakis, G. Dolias, R. Berruto, D. Kateris, and D. Bochtis, "Machine learning in agriculture: A comprehensive updated review," Sensors, vol. 21, no. 11, p. 3758, 2021.

[9] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: A review," Sensors, vol. 18, no. 8, p. 2674, 2018.

[10] G. K. Harikaran, "Machine Learning Approaches for Predicting Soil Properties of Vemagal Hobli, Kolar District, Karnataka," Ph.D. dissertation, Univ. Agric. Sci., GKVK, Bangalore, India, 2022

[11] N. H. Batjes, E. Ribeiro, and A. Van Oostrum, "Standardized soil profile data to support global mapping and modeling (WoSIS snapshot 2019)," Earth System Science Data, vol. 12, no. 1, pp. 299-320, 2020.

[12] A. V. Deorankar and A. A. Rohankar, "An analytical approach for soil and land classification system using image processing," in Proc. 2020 5th Int. Conf. Commun. Electron. Syst. (ICCES), 2020, pp. 1416-1420.

[13] F. M. Riese and S. Keller, "Soil texture classification with 1D convolutional neural networks based on hyperspectral data," arXiv preprint arXiv:1901.04846, 2019.

[14] M. P. K. Devi, U. Anthiyur, and M. S. Shenbagavadivu, "Enhanced crop yield prediction and soil data analysis using data mining," Int. J. Mod. Comput. Sci., vol. 4, no. 6, 2016.

[15] A. Plowcha, et al., "Online Soil Classification Using a UAS Sensor Emplacement System," in Proc. Springer, 2020, pp. 174-184.

[16] L. Benos, A. C. Tagarakis, G. Dolias, R. Berruto, D. Kateris, and D. Bochtis, "Machine learning in agriculture: A comprehensive updated review," Sensors, vol. 21, no. 11, p. 3758, 2021.