Machine Learning Approaches for Soil Quality Classification in Precision Agriculture

Dr. Ayesha Banu[1*], Goli Navyasri[2], Gone Nikhil[2], Gonela Hemanth Sai[2], Jangam Abhiram[2]

[1]Professor & Head, [2]UG Student, [1,2]Department of Computer Science and Engineering (Data Science), Vaagdevi College of Engineering, Bollikunta, Warangal, Telangana.

*Corresponding Email: ayeshabanuvce@gmail.com

ABSTRACT

Soil type classification plays a pivotal role in precision agriculture by optimizing crop management and maximizing productivity. Traditional manual methods, which involve physical sample collection, laboratory analysis, and expert interpretation, are time-consuming, costly, and prone to human error—leading to delays and inconsistent outcomes in decision-making. To address these challenges, this research aims to automate soil classification using advanced artificial intelligence techniques, integrating classical machine learning models with deep learning approaches in a unified, user-friendly system. The proposed solution streamlines the classification process by leveraging image processing to convert soil images into numerical data. This data is used to train models such as the Extra Trees Classifier and a Convolutional Neural Network (CNN). The system is deployed through a graphical user interface (GUI) built with Tkinter, enabling seamless interaction for users without technical expertise. Additionally, the system includes visualization tools such as confusion matrices and training history graphs, allowing real-time monitoring of model performance and accuracy. By automating soil classification, this research aims to modernize current practices, reducing reliance on scarce expert resources while ensuring timely and accurate results. The system supports rapid and reliable soil type identification, leading to improved resource allocation and more informed decision-making in agricultural operations. Ultimately, the implementation of this AI-driven tool has the potential to significantly enhance the efficiency and effectiveness of precision agriculture, contributing to sustainable farming practices and higher crop yields.

Keywords: Soil type classification, Precision agriculture, Predictive analytics, Machine Learning, Deep CNN.

1. INTRODUCTION

Soil classification is used to categorize soils with similar engineering properties based on a shared set of properties or characteristics. The behavior of soil is unlike other engineering materials such as steel or concrete due to its natural formation, geological history, and particulate nature. The determination of soil properties and the explanation of soil behavior form the basis of geotechnical engineering. Soil classification usually depends on certain soil properties, such as grain-size distribution, liquid limit, and plasticity index. Soil classification is important to classify soils with similar properties and to facilitate the dissociation of soils. With soil classification, soil is grouped according to similar properties that it shows when it is exposed to load. Soil classification is a must-do before a foundation design. It is very important to define and classify the soil for the research and design stages of geotechnical engineering processes. Thus, a soil survey should be conducted to determine the soil properties. The correct classification of soils is essential from an engineering point of view. Soil classification is performed based on an analysis of soil properties. One of these is sieve analysis, which determines soil particle size. Based on the sieve openings and the amount of material remaining on the sieve, soil classification

doi: 10.48047/ijiee.2025.15.4.12

is achieved. Plasticity is another relevant parameter that ought to be taken under consideration for the soil classification procedure.

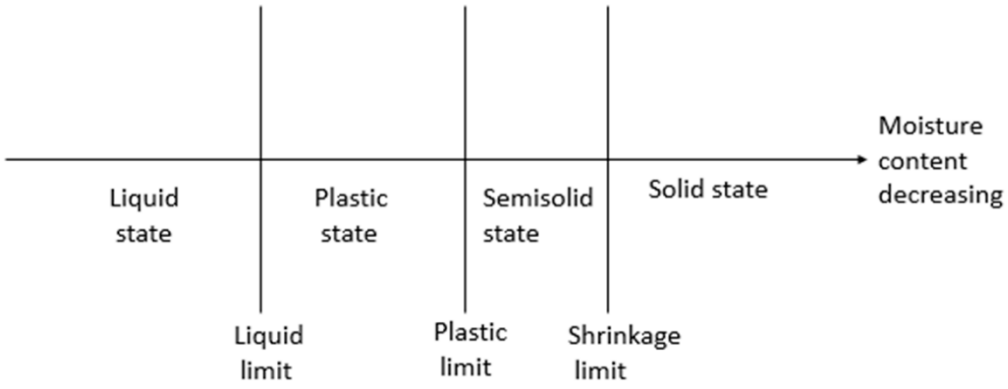

Fig. 1: Consistency of cohesive soils.

The limits used to describe the state of the soil are the liquid limit (LL), plastic limit (PL), and shrinkage limit (SL). As shown in Figure 1, with drying, the soil sample changes to a semisolid state and a solid state. The liquid limit (LL) and the plasticity index (PI) are important parameters in defining and classifying soil. The moisture content from a plastic to a liquid state is the liquid limit, and the moisture content at the point of transition from a semisolid to a plastic state is the plastic limit. Casagrande determined the liquid limit experimentally. The Casagrande cup and fall cone methods can be used to determine the liquid limit. The plastic limit test is simple and is performed by repeated rolling of an elliptic soil mass by hand on a ground glass plate. The numerical difference between the liquid limit and the plastic limit is defined as the plasticity index. Soil classification can be performed by using the liquid limit and plasticity index values on the Casagrande plasticity card.

2. LITERATURE SURVEY

Artificial intelligence (AI) technologies have predicted the behavior of nonlinear systems and have contributed to controlling variables to improve system-operating conditions. A recent analysis highlights the emergence of artificial intelligence as part of the solutions for enhanced farm productivity. Sharma et al. [1] suggested that solar-powered IoT sensor nodes monitor and operate the agricultural sectors. Operations such as crop management, crop harvesting, water supply control, control of animals, distribution of pesticide, moisture, and temperature measuring technologies will also be monitored and controlled in agriculture. Suchithra [2] suggested that sensors can detect field conditions such as temperature, humidity, humidity, and farm fertility. The value of sensing is authenticated and then transmitted to the Wi-Fi, and the verified data from the Wi-Fi module is transmitted via the cloud to the mobile or laptop of the farmer. If the field requires care, farmers are also informed by SMS. An algorithm with temperature, humidity, and fertility thresholds is created that can be configured to manage water quantity in an MCU node. From anywhere in the world, farmers control the engine. Joshi [3] described the construction of the wireless agricultural environmental sensor nodes to monitor climatic conditions and deduct the optimum external conditions for high crop yields in a specific agricultural field. This research focuses on the literature on the construction of wireless agricultural environmental sensor nodes to monitor climatic conditions and deduct the optimum external conditions for high crop yields in a specific agricultural field. Agriculture and food production is a sector that has recently remitted its concentration to WSN, which seeks to raise its production and the agricultural yield benchmark using these cost-effective modern technologies. In recent years, wireless sensor networks (WSNs) have been attracting great attention.

doi: 10.48047/ijiee.2025.15.4.12

Mekonnen [5] discussed that the present analysis is a comprehensive evaluation of the implementation in sensor data analytics within the agroecosystem of different machine learning algorithms. It covers a case study on integrated food, energy, and water (FEW) systems based on IoT-driven smart farm prototypes. Sangeeta et al. [4] suggested that machine learning approach is intended to forecast the best crop yield in a certain area through the analysis of several climatic parameters, such as precipitation, temperature, and dampness, soil pH, and soil type, and previous plant crop records. Ghadge [6] suggested that farmers monitor soil fertility based on data extraction analyses. The method, therefore, focuses on the monitoring of soil quality to determine the crop fit for production by type of soil and to maximize crop production using the right fertilizer recommended. Sujawat [7] discussed that the enormous uses of artificial intelligence are in many domains. Artificial intelligence can be of tremendous help in addressing agricultural illnesses due to its ability to understand the problems and develop the right reasons for them and find ideal solutions for them. The study gives a quick introduction of artificial intelligence application in agriculture, its available farming practices, and the numerous ways available to detect disease in plants. Kshirsagar and Akojwar [8–11] elaborate on the use of artificial intelligence for different classification and prediction problems and furthermore explained the use of hybrid artificial intelligence for feature extraction, classification, and prediction along with modeling with different algorithms and optimization techniques. Significant demonstration in the domains of artificial intelligence, case-based reasoning, multiagent optimization, scheduling, data mining, web crawlers, comprehending and translating natural languages, and virtual vision reality [12–14].

3. PROPOSED METHODOLOGY

The proposed system leverages a Convolutional Neural Network (CNN) for the classification of soil images. CNNs are specialized for image analysis as they use convolutional layers to automatically extract spatial features, such as edges, textures, and patterns, that distinguish different soil types.

Step-1: Soil Dataset

The initial step is to collect and organize a diverse dataset of soil images, specifically focusing on four distinct types of soil: Alluvial Soil, Black Soil, Clay Soil, and Red Soil. Alluvial Soil is typically light-colored with a fine texture, found in river basins. Black Soil is dark and fertile, commonly associated with cotton farming. Clay Soil has a compact texture and retains water well, whereas Red Soil is reddish due to its iron oxide content and is more porous. The dataset should include high-quality images taken under consistent lighting conditions, ensuring variations in soil properties such as moisture content or texture are accurately represented. These images serve as the foundational input for the classification system.

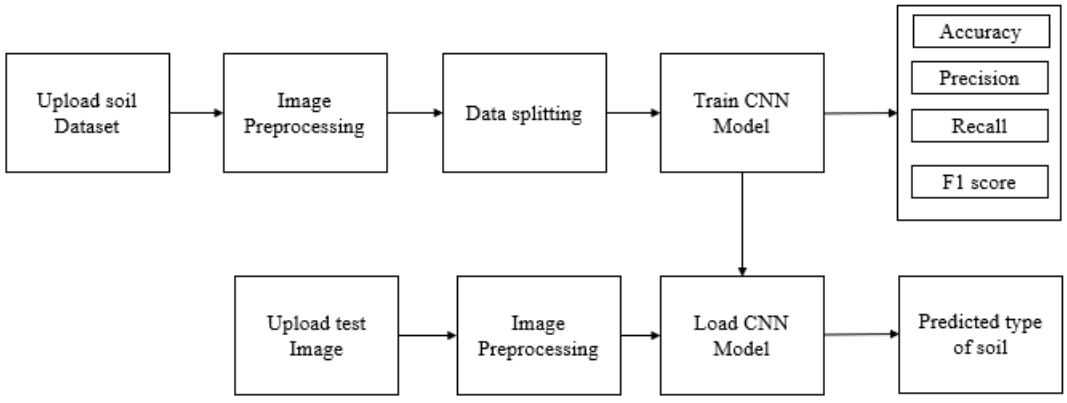

Fig.1: Block Diagram.

doi: 10.48047/ijiee.2025.15.4.12

Step-2: Dataset Preprocessing

Dataset preprocessing is crucial to enhance the quality of the soil images and ensure consistency across the dataset. Images are resized to a fixed resolution, such as 224x224 pixels, to standardize input dimensions for the machine learning models. Normalization of pixel values to a range between 0 and 1 ensures faster convergence during model training. Techniques like image augmentation (e.g., flipping, rotating, and adding noise) are applied to expand the dataset and make the model robust against variations in orientation or lighting. Furthermore, noise reduction methods, such as Gaussian filtering, are employed to remove any unwanted artifacts from the images, improving the clarity of soil features.

Step-3: Existing Algorithm (Extra Trees Classifier)

The Extra Trees Classifier is used as the baseline algorithm for soil type classification. This ensemble learning method aggregates predictions from multiple randomized decision trees, providing high efficiency and stability in handling structured datasets. Each tree in the ensemble is trained on a randomly selected subset of features and samples, which helps reduce overfitting. The classifier's ability to rank feature importance is particularly useful in identifying the visual properties most relevant to soil classification. Although the Extra Trees Classifier is computationally efficient and performs well with tabular data, it lacks the capability to deeply analyze complex spatial patterns in images, making it a suitable comparison point for the proposed deep learning model.

Step-4: Proposed Algorithm (Convolutional Neural Network)

The architecture includes multiple convolutional layers followed by pooling layers to progressively reduce the spatial dimensions and focus on the most critical features. Fully connected layers at the end of the network map these features to the corresponding soil types. The CNN's ability to learn hierarchical feature representations makes it highly effective for the nuanced task of soil classification, outperforming traditional algorithms in recognizing subtle differences between soil types.

Step-5: Performance Comparison

The final step involves comparing the performance of the Extra Trees Classifier and the CNN model using quantitative metrics. Metrics like accuracy indicate the overall correctness of predictions, while precision and recall measure the model's reliability and sensitivity in detecting each soil type. The F1-score, a harmonic mean of precision and recall, is also considered for balanced evaluation. Additionally, computational efficiency (e.g., training and prediction times) be analyzed to assess the practicality of each approach. The comparison highlights the CNN's ability to achieve higher classification accuracy by effectively capturing intricate features from soil images, thus validating its superiority as the proposed solution.

3.1 Model Building

Building a Machine Learning (ML) model involves several key steps. First, select the appropriate algorithm based on the problem type (classification, regression, clustering) and dataset characteristics. Preprocess the dataset by cleaning and transforming data, scaling features, and splitting it into training and testing sets. Initialize the model using a library such as scikit-learn or TensorFlow, and configure hyperparameters as needed. Train the model on the training data by fitting it to identify patterns and relationships. Evaluate the trained model on the testing data using metrics like accuracy, precision, recall, or RMSE, depending on the problem type. If necessary, optimize the model by fine-tuning hyperparameters or experimenting with different algorithms. Once the performance is satisfactory, save the model using tools like joblib or pickle for deployment. Finally, integrate the model into a pipeline or application to make predictions on new data.

doi: 10.48047/ijiee.2025.15.4.12

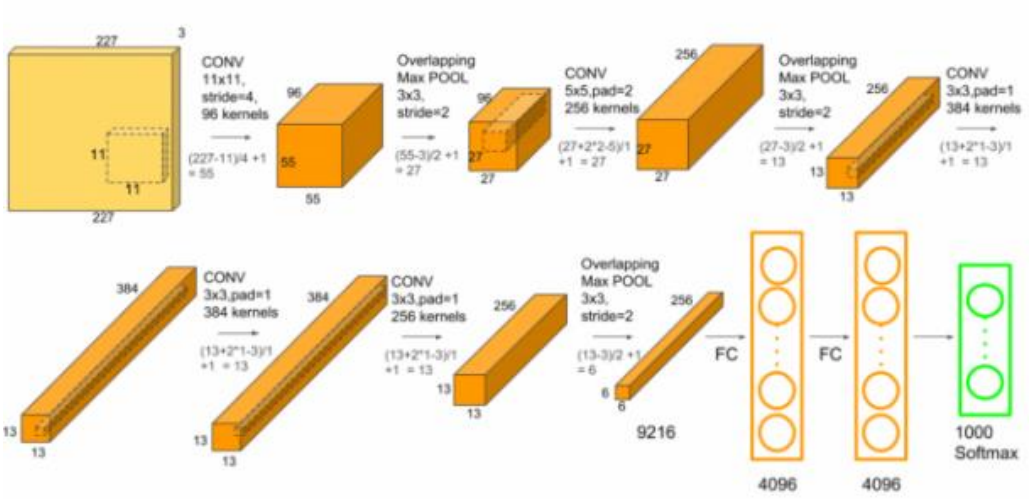

Fig. 3: Block diagram of CNN model.

A Convolutional Neural Network (CNN) is a deep learning model designed to process and analyze visual data such as images and videos. It excels in tasks like image classification, object detection, and pattern recognition by automatically learning spatial hierarchies of features, eliminating the need for manual feature extraction. Inspired by the human visual system, CNNs utilize layers such as convolutional, pooling, and fully connected layers to progressively extract and learn image features. The architecture begins with an input layer that takes pixel values of an image (e.g., 64x64x3 for an RGB image), followed by convolutional layers that apply filters to detect edges, textures, and shapes, producing feature maps. These maps are passed through an activation function like ReLU to introduce non-linearity, and then through pooling layers (like MaxPooling) to reduce dimensionality and computational complexity. The resulting feature maps are flattened into a one-dimensional vector and fed into fully connected layers, where the model learns to associate features with specific classes. Finally, a softmax output layer assigns probabilities to different categories for classification. CNNs process images by sliding filters across the input to extract relevant features layer by layer, gradually building a deeper understanding of the image. This efficient and powerful structure makes CNNs ideal for various computer vision applications.

4. RESULTS AND DISCUSSION

4.1 Dataset description

The dataset comprises 1,215 images representing four distinct soil types, designed for use in soil classification tasks to support precision agriculture. The soil types include Alluvial Soil (523 images), known for its high fertility and presence in river valleys and floodplains; Black Soil (228 images), also called Regur soil, rich in clay content and moisture-retaining properties, commonly found in volcanic regions and ideal for crops like cotton; Clay Soil (197 images), characterized by its fine, compact particles and poor drainage, which makes it challenging to manage despite its fertility; and Red Soil (267 images), identified by its reddish hue due to iron oxide content, typically found in warm, dry climates, and requiring organic matter to enhance fertility. This dataset is intended for image-based classification of soil types, aiding in the development of AI models that can automatically identify and categorize soils. Such classification supports more effective agricultural decision-making by optimizing practices like irrigation, fertilization, and crop selection, ultimately contributing to enhanced productivity and resource efficiency in farming.

4.2 Results description

doi: 10.48047/ijiee.2025.15.4.12

Fig. 4 illustrates the classification performance for each soil type based on precision and recall metrics. Alluvial Soil and Black Soil both demonstrate high precision and recall, indicating that the model performs well in correctly identifying and classifying these soil types with minimal false positives and false negatives. Clay Soil, on the other hand, shows low precision but high recall, suggesting that while the model successfully identifies most actual Clay Soil samples, it frequently misclassifies other soil types as Clay Soil. Red Soil presents a significant challenge for the model, with both precision and recall at zero. This indicates that the model either completely fails to predict Red Soil or misclassifies all Red Soil samples, highlighting a serious performance issue that may require additional data, feature enhancement, or model tuning to address.

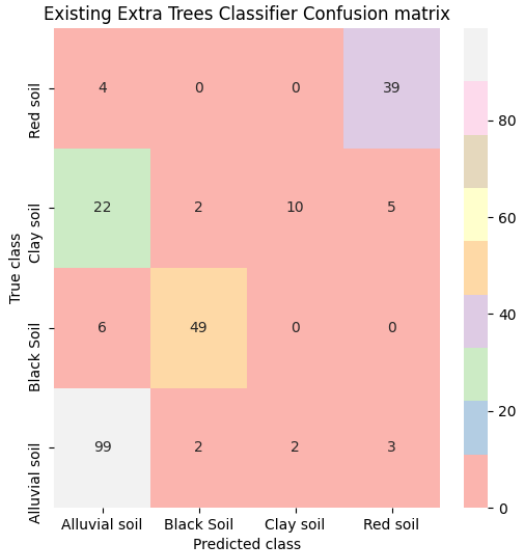

Fig. 4: Confusion matrix of Extra Tree Classifier.

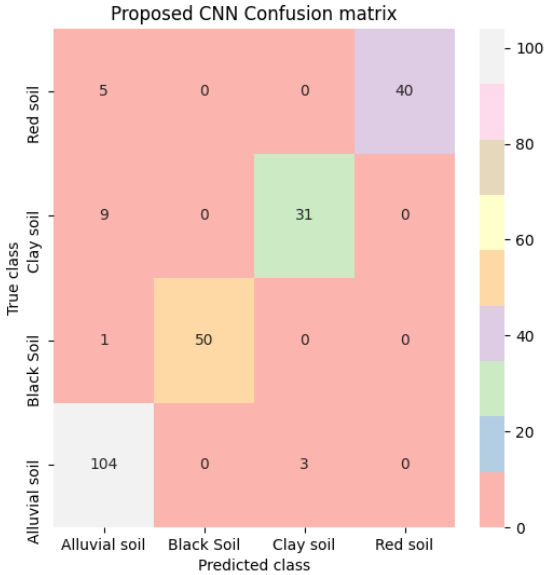

Fig. 5: Confusion matrix CNN.

Fig. 5 highlights the comparative performance of the CNN and Extra Trees Classifier (ETC) models across the four soil classes. For Alluvial and Black Soils, the CNN demonstrates a slight improvement

doi: 10.48047/ijiee.2025.15.4.12

in both precision and recall over the ETC, indicating enhanced classification accuracy. Notably, for Clay Soil, the CNN achieves a significant boost in precision while maintaining high recall, suggesting that it not only identifies more true Clay Soil instances but also reduces false positives. However, both models continue to struggle with Red Soil, exhibiting zero precision and recall, meaning the class is either not predicted or consistently misclassified. Overall, the confusion matrix for the CNN model reflects a better alignment between predicted and actual values, showcasing its superior performance in capturing true classifications compared to the Extra Trees Classifier.

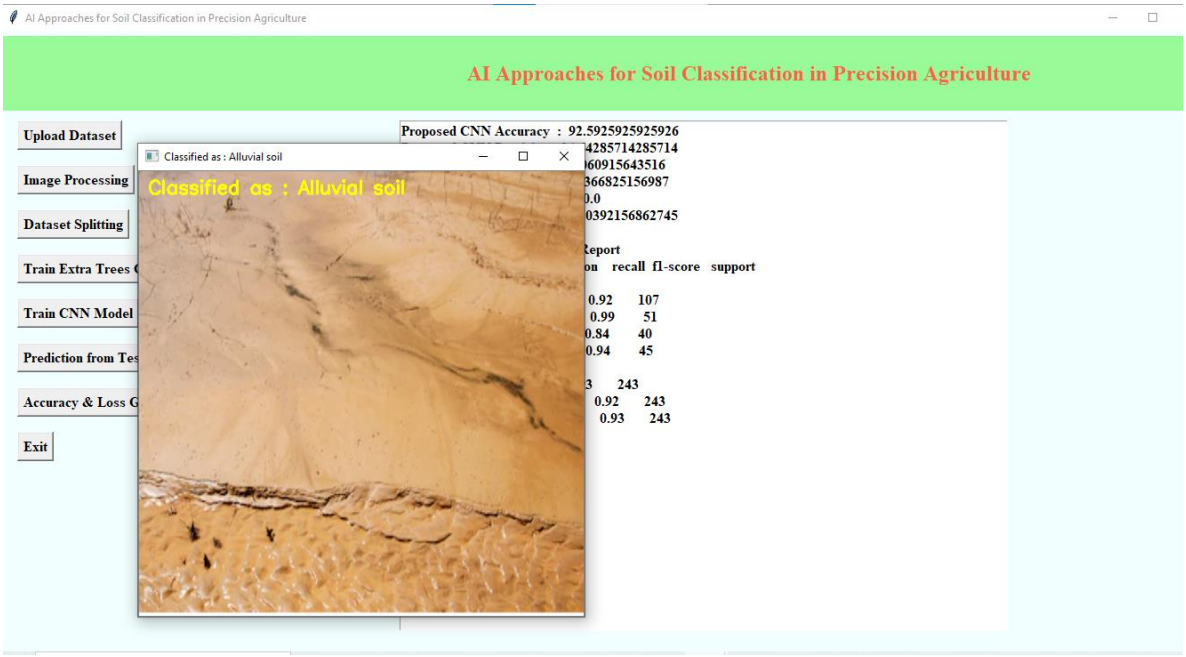

Fig. 6: Predicted output of soil i.e. alluvial Soil.

Table.1: Comparative Analysis: Existing Extra Trees Classifier vs Proposed CNN Algorithm

Metric	ETC model	Proposed Deep CNN model
Accuracy	81.07%	92.59%
Precision	83.58%	94.64%
Recall	74.71%	90.41%
F1-Score	75.04%	92.24%
Sensitivity	98.02%	100%
Specificity	89.09%	98.04%

The comparative analysis between the Existing Extra Trees Classifier (ETC) and the Proposed CNN Algorithm reveals that the CNN significantly outperforms the ETC in all key metrics. The CNN achieves an accuracy of 92.59%, much higher than the ETC's 81.07%. Additionally, the CNN has a superior precision of 94.64% compared to ETC's 83.58%, and its recall of 90.41% far exceeds ETC's 74.71%. The CNN also outperforms the ETC in F1-score (92.24% vs. 75.04%), sensitivity (100% vs. 98.02%), and specificity (98.04% vs. 89.09%). The CNN shows exceptional class-wise performance, particularly with perfect precision and recall for certain classes like Black Soil and Red Soil,

doi: 10.48047/ijiee.2025.15.4.12

demonstrating its ability to accurately classify soil types with fewer errors. This indicates that the proposed CNN model offers a more robust and reliable solution for soil classification tasks.

5. CONCLUSION

The comparison between the Extra Trees Classifier (ETC) and the proposed CNN model reveals that the CNN significantly outperforms the existing model in terms of accuracy, precision, recall, and overall classification performance. The ETC model achieved an accuracy of 81.07%, with strong precision (83.58%) and specificity (89.09%), but it struggled with recall (74.71%) and F-score (75.04%), indicating that while it is effective in predicting positive instances, it fails to capture all relevant cases. Additionally, the confusion matrix shows that the ETC model performs well for Alluvial and Black Soil but struggles significantly with Red Soil, with 0 precision and recall. In contrast, the CNN model demonstrates superior classification abilities, achieving an impressive accuracy of 92.59%, with a high precision of 94.64%, recall of 90.41%, and an F-score of 92.24%, ensuring a well-balanced performance. Its sensitivity of 100% indicates that it perfectly identifies positive cases, while the specificity of 98.04% highlights its robustness in minimizing false positives. The CNN model also shows improved performance in classifying Clay Soil, resolving the low precision issue observed in the ETC model. However, both models continue to struggle with the classification of Red Soil, suggesting that further fine-tuning or the inclusion of additional features required to improve performance in this category. The overall improvements in the confusion matrix demonstrate that CNN provides a more reliable and efficient approach to soil classification. Given the CNN's superior accuracy and balanced precision-recall trade-off, it is evident that the proposed model is a more suitable choice for soil classification tasks, offering better reliability and generalization. Therefore, leveraging deep learning models such as CNN in soil classification can lead to substantial improvements in predictive performance, enhancing agricultural decision-making and land management practices.

REFERENCES

[1] H. Sharma, A. Haque, and Z. A. Jaffery, "Smart agriculture monitoring using energy harvesting Internet of Things (EH-IoT)," An International Scientific Journal, vol. 121, pp. 22–26, 2019.

[2] M. Suchithra, "Sensor data validation," International Journal of Pure and Applied Mathematics, vol. 119, no. 12, pp. 14327–14335, 2018.

[3] P. Joshi, "Wireless sensor network and monitoring of crop field," IOSR Journal of Electronics and Communication Engineering (IOSR-JECE), vol. 12, no. 1, pp. 23–28, 2017.

[4] S. G. Sangeeta, "Design and implementation of crop yield prediction model in agriculture," International Journal of Scientific & Technology Research, vol. 8, no. 1, 2020.

[5] Y. Mekonnen, "Review—machine learning techniques in wireless sensor network based precision agriculture," Journal of the Electrochemical Society, vol. 167, no. 3, article 037522, 2020

[6] R. Ghadge, "Prediction of crop yield using machine learning," International Research Journal of Engineering and Technology, vol. 5, no. 2, pp. 2237–2239, 2018.

[7] G. S. Sujawat, "Application of artificial intelligence in detection of diseases in plants: a survey," Turkish Journal of Computer and Mathematics Education, vol. 12, 2021.

[8] P. Kshirsagar and S. Akojwar, "Optimization of BPNN parameters using PSO for EEG signals," in Proceedings of the International Conference on Communication and Signal Processing, pp. 385–394, India, 2016.

doi: 10.48047/ijiee.2025.15.4.12

[9] S. Oza, A. Ambre, S. Kanole et al., "IoT: the future for quality of services," ICCCE 2020, Springer, Singapore, pp. 291–301.

[10] P. K. Kollu, K. Kumar, P. R. Kshirsagar et al., "Development of advanced artificial intelligence and IoT automation in the crisis of COVID-19 Detection," Journal of Healthcare Engineering, vol. 2022, Article ID 1987917, 2022.

[11] P. R. Kshirsagar, P. P. Chippalkatti, and S. M. Karve, "Performance optimization of neural network using GA incorporated PSO," Journal of Advanced Research in Dynamical and Control Systems, vol. 10, no. 4, pp. 156–169, 2018.

[12] N. Ahmed, D. De, and I. Hussain, "Internet of Things (IoT) for smart precision agriculture and farming in rural areas," IEEE Internet Things Journal, vol. 5, no. 6, pp. 4890–4899, 2018.

[13] S. Sundaramurthy, C. Saravanabhavan, and P. Kshirsagar, "Prediction and classification of rheumatoid arthritis using ensemble machine learning approaches," in 2020 International Conference on Decision Aid Sciences and Application (DASA), pp. 17–21, India, 2020.

[14] M. Padmaja, S. Shitharth, K. Prasuna, A. Chaturvedi, P. R. Kshirsagar, and A. Vani, "Grow of artificial intelligence to challenge security in IoT application," Wireless Personal Communications, 2021.

doi: 10.48047/ijiee.2025.15.4.12