



SoilSmart — A Comprehensive Research of Soil Analysis Techniques and Crop Recommendation Systems

Vaibhav Pandey

Computer Science and Engineering

Shri Ramswaroop Memorial College of Engineering and Management

Lucknow, India

vaibhavpandeyy88@gmail.com

Utkarsh Mishra

Computer Science and Engineering

Shri Ramswaroop Memorial College of Engineering and Management

Lucknow, India

utkmishra2026@gmail.com

Mahesh Bahadur Singh

Computer Science and Engineering

Shri Ramswaroop Memorial College of Engineering and Management

Lucknow, India

maheshcolab@gmail.com

Mekhla Rai

Computer Science and Engineering

Shri Ramswaroop Memorial College of Engineering and Management

Lucknow, India

mekhla.cs@srmcem.ac.in

Abstract— India constitutes a part of the top three producers of several different crops worldwide and serves as a well-known agricultural center. While Indian farmers play a crucial role in the agriculture industry, a large proportion of them are still at the lower end of the socioeconomic scale. Even with a few technological fixes, farmers still struggle to recognize the most lucrative and viable crops for their soil given the diversity of soil types in different parts of the world. This investigation introduces a crop recommendation system that forecasts the best crop derived from a thorough examination of several criteria, such as geography, soil type, yield, selling price, and more. It does this by using both a Convolutional Neural Network (CNN) architecture and a Random Forest Model. It is anticipated that the CNN design would provide an accuracy rate of 95.21%, while the Random Forest Algorithm may produce an accuracy rate of 75%.

Keywords— *Random Forest, Image classification, Deep learning, Convolutional Neural Network, MobileNet v2.*

I. INTRODUCTION

Two-thirds of India's people directly rely on agriculture for their livelihoods, rendering it a traditional and ongoing pillar of the country's economy. Interestingly, it accounts for 20% of India's GDP. At the heart of the agriculture industry is the farmer, our country's Anna Datta (Food Provider), who faces a number of difficulties these days:

1. Since there are so many different kinds of soil within the nation, farmers often struggle to choose the best crop for their particular soil, climate, and area, which may result in large losses.
2. As weather patterns are unpredictable, farmers have significant difficulties in calculating profitability and forecasting yields for certain planting seasons.
3. The 'farm to market' method, which involves several middlemen who eat a sizeable percentage of income via the transportation and selling of products, is the cause of the meagre profits obtained by farmers for their goods.

Artificial intelligence and machine learning technologies have several uses in the contemporary agriculture sector. For example, crop recommender systems and precision agriculture can forecast yields, identify plant pests, improve overall quality of harvests, and solve nutrient deficits in farms. AI system integration has the capacity to revitalise the struggling agriculture industry.

There exist significant apprehensions regarding the future of the agriculture industry in India given the present state of affairs. The agriculture industry confronts difficulties even though it employs about two thirds of the workers and contributes 20% of the GDP. Because they lack collateral, almost all Indian farmers—about 85% of them—operate on fewer than five acres of land, assuming significant product and market risks every season. Due to poor literacy rates, smallholding farmers—who account for half of agricultural output and 46% of cultivated land—often find themselves shut out of contemporary market structures like contract farming and direct purchasing.

II. LITERATURE SURVEY

The research on crop prediction and soil classification in agriculture is compiled and analyzed in this part.

Using data mining methods, Pudumalar et al. [1] created a recommender system for precision agriculture. Their strategy included the CHAID, Random Tree, Naive Bayes, and K-Nearest Neighbor models in an ensemble technique with majority voting. The advice received support from studies on crop yield, soil properties, and kinds.

An ensemble model that integrates Support Vector Machine and Artificial Neural Network, Random Tree, and Naive Bayes models with a majority voting method was proposed by Rajak et al. [2]. The dataset included several soil properties, such as pH and water density, gathered from university and soil testing facilities.

A two-step strategy for crop recommendation and soil categorization was presented by Reddy et al. [3]. During the subsequent phase, consideration was given to characteristics like soil series, temperature, humidity, rainfall, and pH; the initial phase used chemical properties of the soil to forecast the soil series or type. For crop suggestions, classification techniques including Support Vector Machine, K-Nearest Neighbors, and Bagging were proposed.

Venugopal et al. [4] used the Random Forest, Naive Bayes, and Logistic Regression algorithms to forecast yield values and suggest appropriate crops. To train their models, historical data on temperature, weather, and other variables were gathered; the Random Forest Algorithm produced the best accuracy.

A Crop Selection Method (CSM) was presented by Kumar et al. [5] aiming to increase crop net yield rates. The strategy suggested a sequence of crops across a season, considering various factors such as crop type, weather, soil type, and water density. The significance of crop selection was emphasized, along with its influencing variables, encompassing market price, government policy, and production rate.

Data mining methods were utilized by Ahamed et al. [6] to estimate the crop production of cereal crops in Bangladesh. Their technique used KNN, linear regression, and artificial neural networks in the quick miner tool for categorization and clustering to create district clusters.

The approach presented by Mahendra et al. [7] forecasts the ideal crop based on soil properties such as pH, moisture content, and rainfall. The rainfall was predicted using the Support Vector Machine (SVM) technique, while the prediction of crop yield was conducted using the Decision Tree technique.

III. SYSTEM ARCHITECTURE

1) CNN, or Convolutional Neural Network:

i) Collection of Datasets:

Convolutional Neural Network (CNN) training dataset is carefully selected and includes photos of several soil types, including clay, alluvial, black, and red soil. The dataset includes 150–200 photos for each kind of soil in the training set and an extra 50 images in the test set to guarantee robust model training. The Soil Classification Image Dataset on Kaggle and other relevant web sources were carefully gathered to create this dataset.

ii) Pre-processing Data and Implementing Algorithms:

Preparing the data is an important phase in the process because of the variations in sizes and dimensions of the gathered photographs. Pictures are downsized to a common 300x300x3 size, which takes into consideration the RGB format's three channels (Red, Green, and Blue). Afterwards, the machine learning model converts these pictures into numerical arrays. This numerical data is then sent into the algorithm, which is tailored to a particular use case, for further processing and training, making sure the CNN is properly trained to categorize different kinds of soil.

MobileNetV2:

An architecture for deep learning called MobileNetV2 is intended for embedded and mobile devices. It uses inverted residual blocks and depthwise separable convolutions to perform efficient picture recognition. Below is an explanation of the algorithm and how it works:

Algorithm:

4. Input picture: An input image is used to start the process.
5. Fully convolutional layer: To extract the first features from the picture, it is processed through a fully convolutional layer.
6. Inverted residual blocks: After that, a number of inverted residual blocks are run through the features. After applying depth-wise filters and expanding the input channels, each block shrinks the channels to the intended output dimension. The network's effective information flow is guaranteed by the shortcut connection.
7. Bottleneck residual blocks: These blocks find application in downsampling. Before the 3x3 depthwise convolution, these blocks incorporate an extra 1x1 pointwise convolution to further minimize the number of channels.
8. Average pooling layer: An average pooling layer is used to minimize the dimensions of the final feature map.
9. Classification layer: To predict the category of the input picture, the pooled features are subsequently fed into a classification layer.

2) Random Forest:

The Random Forest method is a supervised learning technique that enhances overall accuracy and manages overfitting by aggregating predictions from numerous decision trees.

Algorithm:**Data preparation:**

- The training set is split up into many replacement subsamples.
- A random selection of characteristics is chosen for every subsample.

Decision tree construction:

- On each subsample, a decision tree is constructed using the chosen characteristics. The tree continues to grow until a certain depth or other condition is satisfied.
- The optimal split is selected at each node according to an impurity measure.

Prediction:

- For a new data point, each decision tree in the forest makes a prediction.
- For classification, the majority vote of the trees determines the final prediction. For regression, the average prediction of the trees is used.

IV. CONCLUSION

In conclusion, the study report emphasizes how cutting-edge technologies have transformed India's agricultural environment. Convolutional neural network (CNN) and random forest models are integrated into the suggested crop recommendation system, which is predicted to have accuracy rates of 95.21% and 75%, respectively. These models, considering a variety of variables including soil types and local circumstances, are crucial for enhancing crop choices. The study examines the Random Forest Algorithm's strong performance in crop result prediction as well as the effectiveness of MobileNetV2 in real-time applications.

In keeping with the goals, the study highlights how small farmers may easily access cutting-edge technology via user-friendly channels like email and SMS. By serving as a direct marketplace, the suggested online platform has the ability to cut out middlemen, simplify the supply chain, and give farmers more control. The benefits of utilizing machine learning in crop recommendation, including precision, customization, scalability, and flexibility, emphasize the potential of these technologies to completely change conventional agricultural methods.

Looking forward, the survey's conclusion emphasizes the need for conducting additional research on creative fixes for India's agricultural ecosystem to make it more robust and sustainable. With its innovative technology and farmer-focused approach, the suggested crop selection system is a big step toward reviving the agriculture industry and providing hope for a better future.

V. Future Vision and Next Steps

We aim to deliver a fully functional, data-driven crop recommendation system that empowers farmers with scientific insights accessible through an intuitive interface. The system will bridge the gap between complex agricultural science and practical farming decisions, making precision agriculture accessible to smallholder farmers. By combining cutting-edge technology with farmer-centric design, this system has the potential to revolutionize agricultural decision-making and contribute to sustainable food production.

REFERENCES

- [1] S. Pudumalar and colleagues, "Crop recommendation system for precision agriculture," in Proceedings of the Eighth International Conference on Applied Computing, 2016, pp. 32–36, doi:10.1109/ICoAC.2017.7951470
- [2] R. K. Rajak et al., "A Machine Learning-Based Crop Recommendation System to Maximize Crop Yield," International Journal of Engineering and Technology, issue 12, vol. 4, pp. 950-953, December 2017.
- [3] A.K.M. Reddy, S. Chithra, H.M. Hemashree, and Thanu K., "Soil Classification and Crop Suggestion using Machine Learning," International Journal for Research in Applied Science and Engineering Technology, issue 7, vol. 8, pp. 1625-1628, July 2020.
- [4] A. Venugopal, S. Aparna, J. Mani, R. Matthew, and V. Williams, "Crop Yield Prediction using Machine Learning Algorithms," Int. J. of Eng. Res. and Technol., issue 13, volume 9, pp. 87-91, Aug 2021.
- [5] A. T. M. S. Ahamed et al., "Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh," in Proc. 2015 IEEE/ACIS 16th Int. Conf. Soft. Eng., Artif. Intell., Netw. and Parallel/Distrib. Comput., pp. 1-6, doi: 10.1109/SNPD.2015.7176185.
- [6] Mahendra N., Dhanush V., Nischitha K., Ashwini, and Manjuraju M. R., "Crop Prediction using Machine Learning Approaches," International Journal of Engineering Research and Technology, issue 8, volume 9, pp. 23–26, Aug 2020.
- [7] "Soil-Type-Classification for Crops Suggestion," Accessed: Feb 2021. [Online]. Available: <https://www.kaggle.com/omkargurav/soil-classification-image-data>.
- [8] "Crop Production in India," Accessed: Feb 2021. [Online]. Available: <https://www.kaggle.com/abhinand05/crop-production-in-india>
- [9] A. Barvin, P.; Sampradeepraj, T. "Crop Recommendation Systems Based on Soil and Environmental Factors Using Graph Convolution Neural Network: A Systematic Literature Review." Engineering Proceedings, 58(1), 2023.
- [10] Saki, Mahdi; Keshavarz, Rasool; Franklin, Daniel; Abolhasan, Mehran; Lipman, Justin; Shariati, Negin. "Precision Soil Quality Analysis Using Transformer-based Data Fusion Strategies: A Systematic Review." arXiv preprint (2024).
- [11] Sahu, Ritik; Lowanshi, Vinay; Khare, Nitya. "Soil Health Assessment and Crop Recommendation Using Deep Neural Networks: A Data-Driven Approach for Sustainable Agriculture." Research & Review: Machine Learning and Cloud Computing (e-ISSN: 2583-4835).
- [12] Gaikwad, Swaraj; Redekar, Sudhanshu; Mahamuni, Yash; Dhupal, Ritesh; Pukale, Rohan; Badodekar, Hemant. "A Review on Smart Crop Selection using Machine Learning and Soil Analytics." Journal of Electronics and Telecommunication System Engineering.
- [13] Sudianto, Image; Cahyadi, Eko Fajar. "A Soil Nutrient Assessment for Crop Recommendation Using Ensemble Learning and Remote Sensing." International Journal of Intelligent Systems and Applications (IJISA), Vol. 17 No. 3, 2025.
- [14] Yuliyanto, Supriadi Sahibu, Taufik Imran, Andriansyah Oktafiandi Arisha AMIK, Munawirah. "Crop Recommendation Based on Soil and Weather Conditions Using the K-Nearest Neighbors Algorithm." Journal of System and Computer Engineering.
- [15] Ghani, M.A.; Ul Haq, Syed Rizwan; Mohib Khan, M.D. "Agriculture Soil Analysis, Classification and Crop Suitability Recommendation Using Machine Learning." International Journal of Engineering and Science Research, Vol 14 No.3 (2024).
- [16] Swetha P., Harshitha L., Jyoti S V, Karuna M N, Deeksha S. "Soil Based Crop Recommendation System Using Machine Learning." International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), 2025.
- [17] "Artificial Neural Networks Based Integrated Crop Recommendation System Using Soil and Climatic Parameters." Indian Journal of Science and Technology.
- [18] Pandey, Das, Biswas. "AgroSense: An Integrated Deep Learning System for Crop Recommendation via Soil Image Analysis and Nutrient Profiling." arXiv preprint, 2025.
- [19] Dey, B., Ferdous, J., & Ahmed, R. (2024). Machine learning based recommendation of agricultural and horticultural crop farming in India under the regime of NPK, soil pH and three climatic variables. Heliyon, 10(3), e25112.
- [20] Phanikanth Chintamaneni & Subrahmanyam Kodukula (2025). Integrating Machine Learning and IoT Sensors for Enhanced Soil Nutrient Monitoring and Crop Recommendation Systems. Journal of Information Systems Engineering and Management, 10(44s), article 8617.
- [21] S. Babu, "A software model for precision agriculture for small and marginal farmers," in Proc. 2013 IEEE Global Humanitarian Technology Conference: South Asia Satellite, pp. 352–355, doi: 10.1109/GHTC-SAS.2013.6629944.
- [22] A. Venugopal, S. Aparna, J. Mani, R. Matthew, and V. Williams, "Crop Yield Prediction using Machine Learning Algorithms," Int. J. of Eng. Res. and Technol., issue 13, vol. 9, pp. 87–91, Aug 2021.

- [23] Mahendra N., Dhanush V., Nischitha K., Ashwini and Manjuraju M.R, "Crop Prediction using Machine Learning Approaches," *Int. J. of Eng. Res and Technol.*, issue 8, vol. 9, pp. 23–26, Aug 2020.
- [24] M. Champaneri, D. Chachpara, C. Chandvidkar and M. Rathod, "Crop Yield Prediction using Machine Learning," *Int. J. of Sci. and Res.*, issue 1, vol. 10, pp. 01–03, Apr 2018.
- [25] Kaggle, "Crop Production in India," Accessed: Feb 2021. [Online]. Available: <https://www.kaggle.com/abhinand05/crop-production-in-india>
- [26] A. Krizhevsky, I. Sutskever and G.E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 25th Int. Conf. on Neural Information Processing Systems*, 2012, pp. 1097–1105, doi: 10.1145/3065386.
- [27] Bharath K.R., Balakrishna K., Bency C.A., Siddesha M. and Sushmitha R., "Crop Recommendation System for Precision Agriculture," *Int. J. of Comput. Sci. and Eng.*, issue 5, vol. 7, pp. 1277–1282, May 2019.
- [28] A.E. Khedr, M. Kadry and G. Walid, "Proposed Framework for Implementing Data Mining Techniques to Enhance Decisions in Agriculture Sector," *Procedia Comput. Sci.*, issue 25, vol. 65, pp. 633–642, 2015.
- [29] A. T. M.S. Ahamed, "Applying data mining techniques to predict annual yield of major crops," in *Proc. 2015 IEEE/ACIS 16th Int. Conf Soft. Eng., Artif. Intell., Netw. and Parallel/Distrib. Comput.*, pp. 1–6, doi: 10.1109/SNPD.2015.7176185.
- [30] A. Savla, "Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture," in *Proc. 2015 Int. Conf Innov. in Inf., Embedded and Commun. Syst.*, pp. 1–7, doi: 10.1109/ICIIECS.2015.7193120.

