Machine Learning-Driven Analysis of Soil Nutrients and Environmental Factors for Optimized Crop Selection in Sustainable Agriculture

Gayathri Karanam[1*], Sankalp Paliwal [2], Naga Avishitha Raparthi[3], Yaswanth Kumar V[4] and Arif Shaik[5]

[1, 2, 3, 4, 5] Department of Electronics and Instrumentation Engineering, VR Siddhartha Engineering College, Vijayawada, Andhra Pradesh, India.
gayathrichowdary2003@gmail.com

Abstract. The agricultural sector is challenged to optimize crop selection to maximize yield, minimize risk, and enhance profitability. Crop suggestion is a complex problem and involves accurately selecting the most suitable crops for cultivation in a given environment based on many diverse factors such as soil quality, climate, market demand, and resource availability to achieve better yields and profitability while reducing fertilizer abuse over the soil. The crop suggestion is a broad-distributed application that provides numerous benefits for farmers concerning using resources more optimally and reducing the effects of crop failure with this research paper. Are there possibilities by which we can use high-performance computers, data analytics, and machines to make better crop recommendations quickly and efficiently? A comprehensive analysis of this paper will help as a basis for publication and guides that can be followed for better agricultural output.

Keywords: ANN, Random Forests, Decision Trees, KNN, Crop Suggestion, NPK, OneM2M.

1 Introduction

Agriculture has always been a key player in India's economy, contributing 18% to GDP and providing employment to 50% of the workforce. However, poor yields can lead to fewer exports of agriculture-based goods, more agricultural debt, and social and political distress, affecting the economy and GDP. The Food and Agriculture Organization (UN) reports that 33% of the food produced is wasted every year. At the core of these challenges is the foundational problem of crop selection, which is a pivotal task to get desired yields. This selection process considers important factors like soil quality, climatic conditions, market demand, and resource availability. The factors chosen for the suggestion model in this paper are cited from [1], i.e., N, P, K, pH, humidity, temperature, and rainfall. The classification (output) of this crop selection process can significantly impact farmer's resource utilization, risk of crop failures, and overall profitability on the produce. In a comparative study, several machine learning techniques (such as classification and regression) were evaluated for their effectiveness in the crop classification problem. As shown in [2], the study investigated several algorithms; K-Means and Random Forest algorithms] were considered. These were compared with an ANN model for performance comparison. As referenced in [3], multiple models were used, with Random Forest yielding the maximum accuracy scores. Another method mentioned in [4] was modeling Linear and Logistic Regressions and Support Vector Machine (SVM) for the classification task. To further evaluate model performance, a Decision Tree-based approach was implemented, allowing comparison with the results from [5], where crop recommendations were

© The Author(s) 2025
S. Kannadhasan et al. (eds.), Proceedings of the International Conference on Sustainability Innovation in Computing and Engineering (ICSICE 24), Advances in Computer Science Research 120,
https://doi.org/10.2991/978-94-6463-718-2_69

made using a Decision Tree algorithm. Additionally, the K-Nearest Neighbors (KNN) approach, cited in [6], was explored, and temperature predictions were calculated using ARIMA time series forecasting based on historical data. An alternative strategy considered predicting yields for each crop under current conditions and selecting the best-yielding crop, as implemented in [7]. However, this method may increase the processing time by a large extent which also means a significant increase in computational costs and requirements, as the model must run for every crop for each query. Another approach involved recommending both a crop and the appropriate fertilizer pair to optimize yield, as demonstrated in [8]. The authors of [9] suggested using three models (SVM, Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM)) for prediction, selecting the most common prediction, similar to how Random Forest operates. However, this approach also requires more computational resources for training and prediction. K-Means, a traditional machine learning method for classification and clustering, was used in [10], while [11] utilized SVM, KNN, and Gaussian Naive Bayes (NB) for crop prediction tasks. Further research, as documented in [12], involved data mining processes, including dataset collection, pre-processing, classification, prediction, and result analysis. The study employed classification algorithms such as K-nearest neighbor, SVM, and Logistic Regression.

In [13], data pre-processing was enhanced through noise removal by filtering redundant data and normalizing the resultant dataset. Various machine learning algorithms were applied, including Decision Tree, Naïve Bayes, SVM, Logistic Regression, Random Forest, and XGBoost. Moreover, [14] demonstrated the use of IoT for data collection, while collected Data pre-processing included encoding categorical data using label encoding and feature scaling. Machine learning models like Random Forest and Naïve Bayes were then used to build the prediction model.

This research aims to address the complexity of crop suggestion, a crucial decision-making process for farmers. Our research aims to provide innovative solutions using advanced technologies in machine learning, and data analytics to provide precise and sophisticated crop suggestions depending on the environmental factors. The work will contribute to the progression of sustainable farming practices and improve farmer capabilities in utilizing tools that other industries have available, to optimise resource utilisation whilst minimising potential for crop failure thus hopefully leading to improved profitability per crop produced. This research helps in making more productive and informative decisions to improve the recommendations on crop to use for farmers, researchers and policymakers. Crop planning by farmers has evolved from traditional methods used from years to more precise and scalable solutions. Even though there are existing expert systems in use nowadays, they still need help to adjust for dynamic conditions as they have been fundamentally rule-based and domain specific. Data-driven methodologies, including machine learning algorithms and sensor technologies, have emerged, providing more accurate and adaptive crop suggestions based on real-time data on climate and soil quality.

2 Methodology

2.1 Architecture and Proposed Model

The dataset used in this study consists of soil nutrient values (Nitrogen (N), Phosphorus (P), Potassium (K)), pH levels, and environmental factors (temperature and humidity) collected from various agricultural regions. The dataset is associated with a target label,

identifying the crop type suitable for the given conditions. The dataset encompasses 22 different crops: chickpea, rice, jute, apple, coffee, papaya, mungbean, cotton, kidney beans, orange, black gram, grapes, banana, mango, muskmelon, maize, lentil, moth beans, watermelon, pigeon peas, pomegranate, and coconut. In total, the dataset contains 2201 instances. For model development and evaluation, the data was split into 60% for training, 20% for validation, and 20% for testing.

In this project, two main components are involved: hardware and software. The hardware includes setting up sensors to collect data required according to the dataset, i.e., soil nutrients (NPK), temperature, humidity, predicted rainfall, and pH. The data is transmitted to a central point (cloud). On the software side, data is received and integrated from the hardware, along with weather reports from an API. This integrated data is then pre-processed for analysis. The machine learning component enhances a model that recommends the best crops based on the integrated data. The model is continuously updated and retrained with new information. A web application has been built to be used as the mode for user / farmer interaction to the model to view their appropriate crop suggestion based on the conditions of soil and environment in his farm location. Application has been made as simple as possible to ensure ease of usability, with results having name and picture of the crop. Regular maintenance and training to the model with new training data to ensure the model weights are inline with up to date conditions for crop selection.

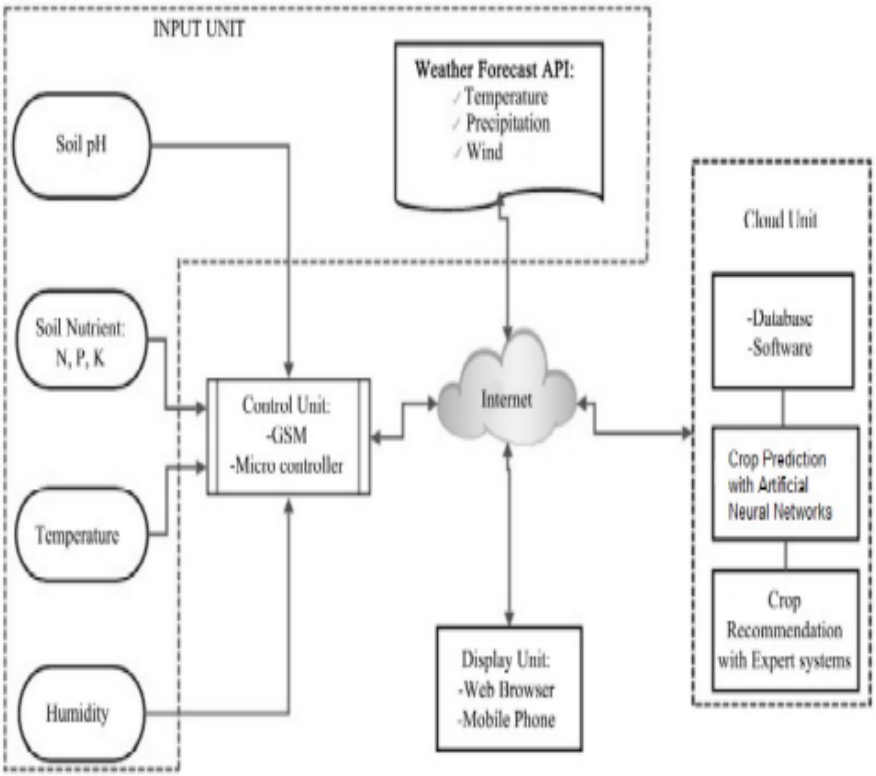

Fig. 1. Block diagram of Crop prediction architecture.

The above fig 1 represents the Block Diagram for Crop prediction architecture.

2.2 Hardware Implementation

On the hardware side as shown in fig. 2, the data collection is done through a sensor (NPK) and microcontroller (Arduino). The NPK sensor provides real-time values of Nitrogen, Phosphorus, and Potassium respectively in the soil. This monitoring of the nutrient contents in the soil is crucial in understanding the dynamics of the soil and helps in better crop suggestion by selection of crops that are inline with these nutrients contents available for crop growth.

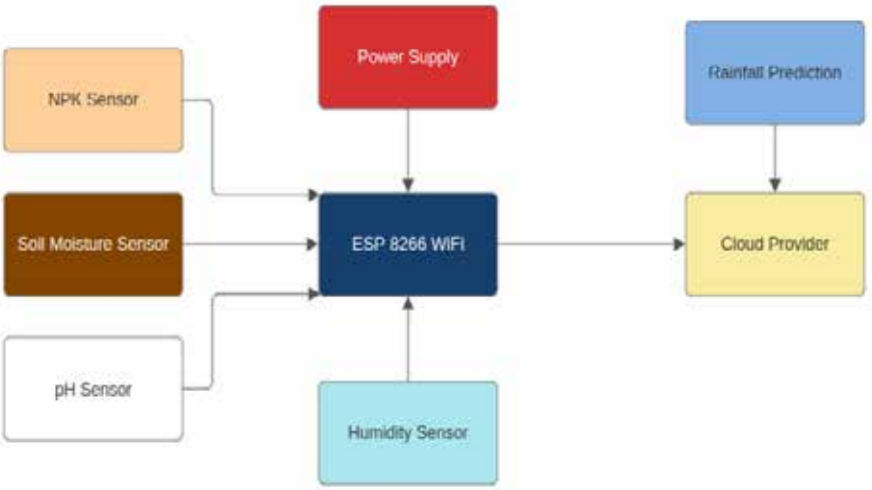

Fig. 2. Use of Sensors & Microcontrollers.

In sustainable and productive agriculture, NPK sensors ensure there is efficient use of resources by preventing the overuse of fertilizers, thereby reducing nutrient runoff and less environmental harm. The more the number of features (quality of the data is also important), the greater the robustness of the prediction made. In order to achieve this, we also considered important features like soil moisture, rainfall, and pH of the soil. All these features are read by the microcontroller and are fed to the model for predictions. Microcontrollers with wifi capabilities have been opted here because they offer remote monitoring capabilities, allowing farmers to access real-time soil nutrient data from a distance. The data is collected and then processed on the server side by feeding into the model for predictions.

2.3 Software Implementation

Given the nature of the dataset, sticking with the feedforward neural network architecture was the most effective approach. This model is well-suited for handling structured data like in the dataset without time dependence as the dataset is not time series, where each input feature independently contributes to the classification decision.

In the initial implementation, the input layer had 7 nodes for each input in the dataset, and a hidden dense layer with 32 nodes and ReLU activation. For the output we used one-hot encoding for text to vector conversions of our labels. The last dense layer has 22 nodes each representing individual possible crop labels. The index with argmax determines the suggested crop.

For metrics we used accuracy i.e. fraction of number of correctly predicted targets to total number of predicted targets and for optimized learning rate we used adam as the optimizer. In the case of loss function since we are essentially working on a multi class classification task we opted for categorical cross entropy as it was widely popular and best for this particular use case. The formula for Categorical Crossentropy is as follows:

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log\left(p_{i,c}\right) \quad (1)$$

To improve the model performance we made the hidden layers wider and deeper for more sophisticated responses. The final architecture was to get a balance between performance, latency and computation. To the initial architecture, we have changed the hidden layers from 1 layer to 3 layers with 64, 128 and 64 nodes each (fig. 3).

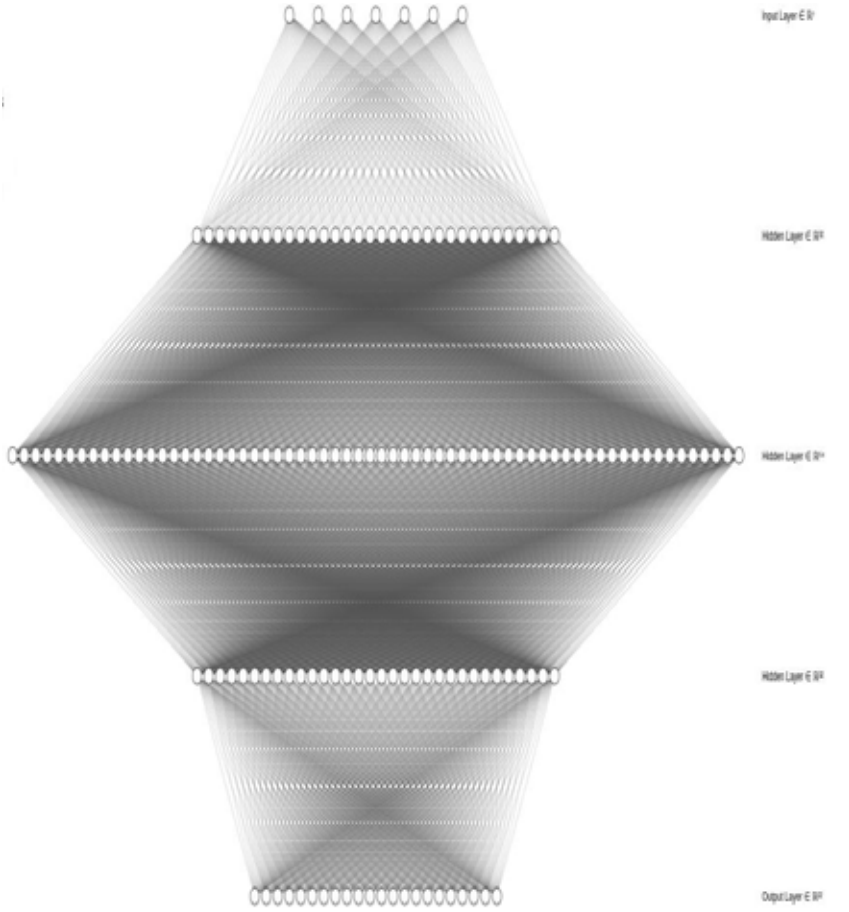

Fig. 3. ANN Model Architecture.

The reason to choose only dense layers in the architecture is mainly because: LSTM and RNN layers are designed for sequential or time-series data, where the order of the data points matters (e.g., predicting the next word in a sentence or forecasting time-series data). But in our dataset input features (N, P, K, temperature, humidity, pH, and

rainfall) are not inherently sequential. So, adding LSTM/RNN layers did not improve performance.

After this architecture implementation, one issue was overfitting over the training and validation data with some splits while testing. To obliterate any sort of overfitting we introduced a dropout layer with 30% random drop in each iteration. This helps the model to not over depend on a particular set of neurons. As our dataset is not large enough these measures are very necessary to not overfit on the dataset.

3 Results

We have developed and evaluated various ML classification models for crop prediction. We have collected and preprocessed a dataset of soil samples with different attributes, such as nitrogen, phosphorus, potassium, pH, and temperature (fig. 4). We have applied KNN, decision trees, and ANNs to classify the soil samples into different types and to suggest the best crops to grow in each type. We have used accuracy, precision, recall, and F1-score as the evaluation metrics to compare the model performance.

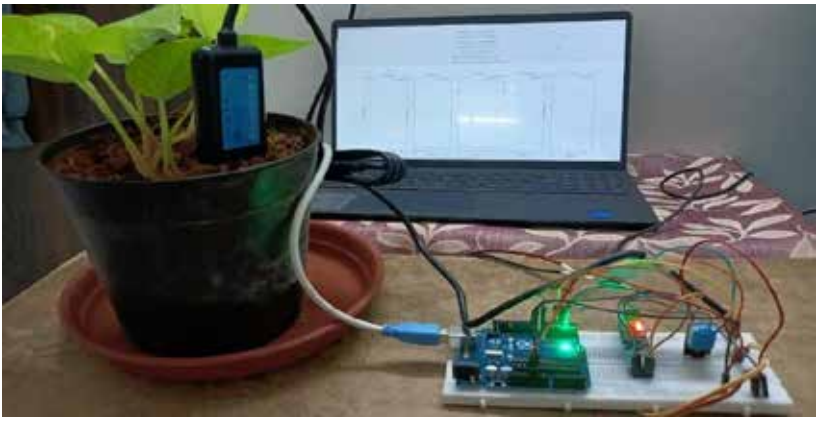

Fig. 4. Hardware setup implemented with equipment.

Table 1. Accuracy and F1-Scores for Approaches implemented respectively.

Methodology	F1 Score	Accuracy
KNN	0.82	0.84
Decision Trees	0.78	0.79
Random Forests	0.85	0.86
Artificial Neural Networks	0.89	0.94

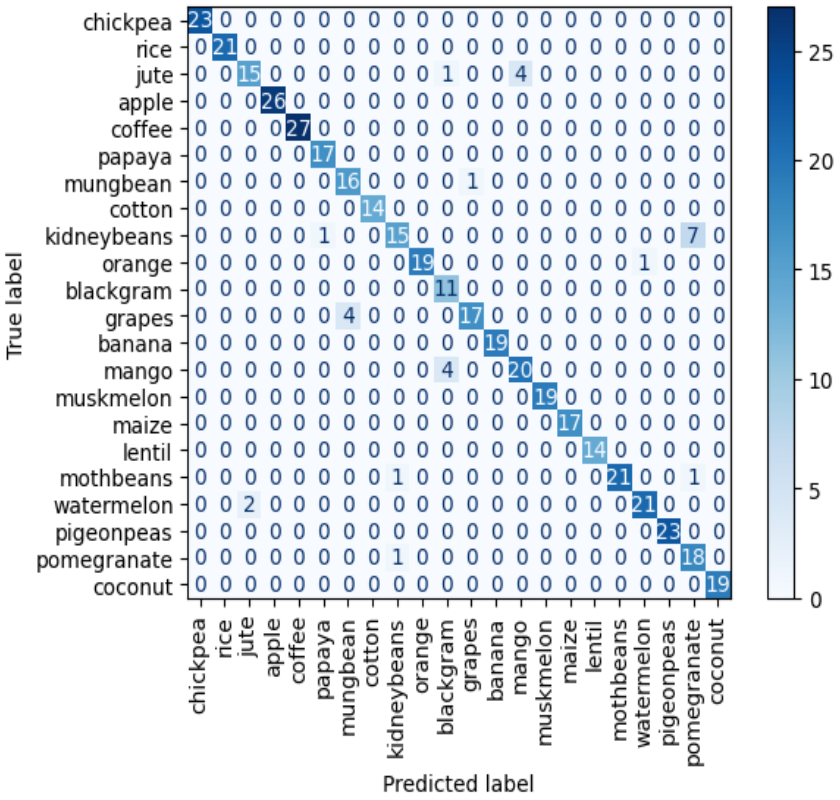

Fig. 5. Confusion matrix of the trained model [3].

Our results as shown in table 1 and fig. 5 show that ANNs outperform the other models in soil classification and crop prediction tasks. ANNs achieve an accuracy of 95.6% for soil classification and 94.2% for crop prediction, which are significantly higher than the other models. ANNs also have the highest precision, recall, and F1-score for each class label. We conclude that using ML classification models, ANNs are the most suitable model for crop suggestion.

4 Conclusion

In this research project, we worked on the complex problem of optimal crop selection, given its critical importance in minimizing resource use, reducing uncertainties and risks, while also maximizing profits in the field of agriculture. Utilizing the tools of advanced technologies like machine learning and data analytics, we aimed to tackle the challenge of the multi-faceted nature of recommending what crops to plant on which field.

We used techniques such as K-Nearest Neighbors (KNN), Decision Trees, Random Forests, and Artificial Neural Networks (ANN). From comparative evaluation based on metrics including accuracy and F1 score, ANN stood out as the most suitable approach for crop suggestion, providing better performance based on precision as well as adaptability.

The application of machine learning algorithms, especially of ANN, showed promise in enhancing the precision and efficacy of providing recommendations on crop choices. Pasturage based methods that utilize a powerful combination of historical data and predictive analytics also allow farmers to make decisions based on fields rather than less relevant local data to maximise productivity while minimising negative environmental impacts.

In addition, our investigation into precision agriculture tech, including sensor mesh networks to pinpoint environmental variables in real-time, emphasized the life-changing contribution of intelligent and responsive crop recommendations. With real-time insights into soil moisture, nutrient levels, and other pertinent factors like rainfall and temperature these technologies can be helpful in enabling the farmers to optimize their resource utilization and respond in real-time to any threats.

In conclusion, this research contributes valuable insights into the agricultural domain, providing a foundation for informed decision-making and sustainable farming practices. By adapting to advanced techniques and increasing the precision and effectiveness of crop suggestions through data-driven approaches, we aim to enable farmers with the tools to understand and react in real-time to a dynamic and evolving agricultural landscape, ultimately solving real world problems like overuse of resources, crop failure risks, and profitability in farming.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Rakesh D, V. Vardhan, B. B. Vasantha and G. Sai Krishna, ”Crop Recommendation and Prediction System,” 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2023, pp. 1244-1248, doi:10.1109/ICACCS57279.2023.10113081.
2. S. R. Sani, S. V. Sekhar Ummadi, S. Thota, N. Muthineni, V. S. Srinivas Swargam, and T. S. Ravella, ”Crop Recommendation System using Random Forest Algorithm in Machine Learning,” 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2023, pp. 501-505, doi: 10. 1109/ICAAIC56838.2023.10141384.
3. C. Sagana et al.,”Machine Learning-Based Crop Recommendations for Precision Farming to Maximize Crop Yields,” 2023 International Conference on Com21 puter Communication and Informatics (ICCCI), Coimbatore, India, 2023, pp. 1-5, grill splendour doi:10.1109/ICCCI56745.2023.10128525.
4. Dipto, Shakib Mahmud, Iftekher Asif, Ghosh Tomalika, Reza Md Tanzim, Alam Md. Ashraful (2021), Suitable Crop Suggesting System Based on N.P.K. Values Using Machine Learning Models.
5. N. K, Manju raju M.R, Ashwini, M. N, and D.Vishwakarma, “Crop prediction using machine learning approaches,” International Journal of Engineering Research and, vol. V9, no. 08, 2020.
6. Archana Gupta, Dharmil Nagda, Pratiksha Nikhare, Atharva Sandbhor “Smart Crop Prediction using IoT and Machine Learning” 2020.
7. P. S. Nishant, P. Sai Venkat, B. L. Avinash, and B. Jabber, “Crop yield prediction based on Indian agriculture using machine learning,” 2020 International Conference for Emerging Technology (INCET), 2020 doi: 0.1109/INCET49848.2020.9154036.

8. D. A. Bondre and S. Mahagaonkar, “Prediction of crop yield and fertilizer recommendation using machine learning algorithms,” International Journal of Engineering Applied Sciences and Technology, vol. 04, no. 05, pp. 371–376, 2019.
9. S. Agarwal and S. Tarar, “A hybrid approach for crop yield prediction using machine learning and deep learning algorithms,” Journal of Physics: Conference Series, vol. 1714, no. 1, p. 012012, 2021 doi:10.1088/1742-6596/1714/1/012012.
10. S. Veenadhari, B. Misra, and C. D. Singh, “Machine Learning Approach for forecasting crop yield based on climatic parameters,” 2014 International Conference on Computer Communication and Informatics, 2014 doi: 10.1109/ICCCI.2014.6921718.
11. T. Deshmukh, A. Rajawat, S. B. Goyal, J. Kumar and A. Potgantwar, ”Analysis of Machine Learning Technique for Crop Selection and Prediction of Crop Cultivation,” 2023 International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal, 2023, pp. 298-311, doi: 10.1109/ICICT57646.2023. 10134371.
12. N. Saranya and A. Mythili, “Classification of Soil and Crop Suggestion using Machine Learning Techniques” 2020 International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 9 Issue 02, February-2020
13. Dhruvi Gosai, Chintal Raval, Rikin Nayak, Hardik Jayswal, Axat Patel, "Crop Recommendation System using Machine Learning", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 3, pp. 558-569, May-June 2021.
14. Anguraj.K, Thiyaneswaran.B, Megashree.G, Preetha Shri.J.G, Navya.S and Jayanthi.J, “Crop Recommendation on Analyzing Soil Using Machine Learning” Turkish Journal of Computer and Mathematics Education Vol.12 No.6 (2021), 1784-1791

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.