Semiconductor Optoelectronics
ISSN: 1001-5868

MACHINE LEARNING-ENABLED SOIL ANALYSIS FOR CROP SELECTION AND YIELD PREDICTION

Shreeshayana R[1], Raghavendra L[2], Kishore K V[3] and Anil D[4]
[1] Department of Electrical and Electronics Engineering,
ATME College of Engineering, Mysuru, India.
[2] Department of Electrical and Electronics Engineering,
ATME College of Engineering, Mysuru, India.
[3] Department of Mathematics
CMR Institute of Technology, Bengaluru, India
[4] Department of Computer Science & Engineering,
CMR Institute of Technology, Bengaluru, India.

Abstract— The foundation of a productive farm is healthy soil. One of the most crucial farm resources is the soil. It serves as a vital reservoir of water and nutrients for crops. Each type of soil has unique characteristics. Farmers can always improve soil quality by controlling nitrogen levels and soil pH even when soil texture cannot be changed. Regular soil analysis is one of the key elements in regulating soil quality. This article explores the efficacy of three machine learning algorithms—Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), and Random Forest—in agricultural decision-making by focusing on crop prediction, fertilizer estimation, and yield forecasting based on soil attributes. Notably, all three algorithms demonstrate outstanding accuracy in crop prediction, with each achieving a perfect score, affirming their reliability in recommending crops aligned with specific soil characteristics. In the domain of fertilizer estimation, SVM emerges as the standout performer with flawless accuracy, while Random Forest closely follows with remarkable precision. For yield prediction, Random Forest leads with an impressive accuracy rate, followed by K-NN, both offering valuable insights into anticipated harvests.

Keywords — Agriculture, Crop Prediction, Fertilizer Estimation, Machine Learning, Soil Analysis, Yield Forecasting.

I. INTRODUCTION

India, often characterized as an agriculture-based nation, has historically relied on farming as a cornerstone of its economy, with approximately 50% of its workforce actively engaged in agricultural activities. This sector holds a pivotal role in the country's economic framework, contributing a substantial 7.68% to the global agricultural output, exceeding the world average of 6.1%. One of the persistent issues in Indian agriculture is the reliance on age-old, traditional farming methods. A considerable portion of Indian farmers continue to cultivate crops without adequate knowledge of the content and quality of their soil. This historical practice has had significant repercussions, leading to crop failures and economic losses for many farmers.

Consequently, India's traditional farming sector still grapples with some of the lowest per capita productivity and farmer incomes in comparison to its potential. Furthermore, traditional farming methods often demand a substantial human workforce to perform various tasks such as manual watering, cultivation, and the application of pesticides [1]. Understanding the soil's fertility and characteristics is paramount in predicting which crops are best suited for a particular type of soil. This knowledge can be the difference between a bountiful harvest and crop failure. Soil, therefore, emerges as a crucial determinant in agricultural success [2].

In this research paper, we propose a model that leverages the power of machine learning to tackle these agricultural challenges head-on. Our model utilizes key soil parameters, including Nitrogen (N), Phosphorus (P), Potassium (K), pH levels, moisture content, and temperature, to provide valuable insights into soil quality. Specifically, it suggests suitable crops based on soil characteristics, estimates the required number of fertilizers, and predicts crop yields using a range of machine learning algorithms, including Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Random Forest. In doing so, we aspire to enhance crop yields, increase farmer incomes, and contribute to sustainable growth of India's agricultural sector [3]. The main aim of our system is to revolutionize current manual soil testing procedure by introducing partial automation. In our proposed system, we put forward an innovative approach that encompasses crop suggestion, fertilizer estimation, and yield prediction. These predictions are based on crucial soil attributes, including nitrogen (N), phosphorous (P), potassium (K), pH levels, moisture content, and temperature [4]. The amount of soil organic matter, the desired yield, and the soil test value of the nutrient in question all play a significant role in determining a crop's fertilizer needs. The three essential plant nutrients nitrogen, phosphorus, and potassium are included in the majority of fertilizers that are frequently used in agriculture. Certain "micronutrients," including zinc and other metals, that are essential for plant growth are also present in some fertilizers.
To accomplish this, our software model will employ a robust machine learning framework.

II. LITERATURE REVIEW

Singh et al., [5] explains the utilization of Machine Learning techniques to enhance crop yield prediction and nitrogen management in precision agriculture. Their research underscores the potential of Machine Learning in optimizing both crop yield and resource allocation, offering valuable insights for the advancement of modern farming practices.
Kwaghtyo et al., [6] provides an extensive and holistic examination of the application of Machine Learning and Data Mining techniques in precision agriculture. Encompassing diverse facets such as soil analysis, crop prediction, and yield forecasting, this review furnishes a comprehensive understanding of how these advanced technologies are shaping and enhancing modern agricultural practices.
Oikonomidis et al., [7] conducts a thorough exploration of diverse Machine Learning algorithms employed in the context of crop yield prediction. This study not only emphasizes the significance of precise yield forecasting but also delves into the evolving applications and trends within this field, showcasing the dynamic role of Machine Learning in revolutionizing agricultural practices.
Singh et al., [8] a microcontroller-based soil testing technique is proposed. An electronic device

is used to measure the N(nitrogen), P(Phosphorous), K(potassium) and pH (potential of Hydrogen) values to estimate the fertility of the soil in the field of agriculture. This helps in predicting a suitable crop and type of fertilizer to be used. Sensors are used to send the ionic particles and its output is processed by signal conditioning unit. Microcontroller compares the pre-stored values with the actual values and displays the output on the LCD. Some methods use one or two prominent features as input to estimate other parameters on the basis of which actual output is predicted.

Durai et al., [9] location is the prominent feature. The estimated soil parameters using location and the previous crop data base are used to suggest the crops. Garg et al., [10] throws light on how information technology technique cloud computing could be deployed in farming for better data management and accessibility. Machine learning technique includes examination of enormous dataset which helps in acquiring precise results. In this manner, it is additionally being utilized in the field of agribusiness. The Types of Soil in India, with its land coverage, Soil nutrients richness, pH value and commonly grown crops is shown in Figure 1.

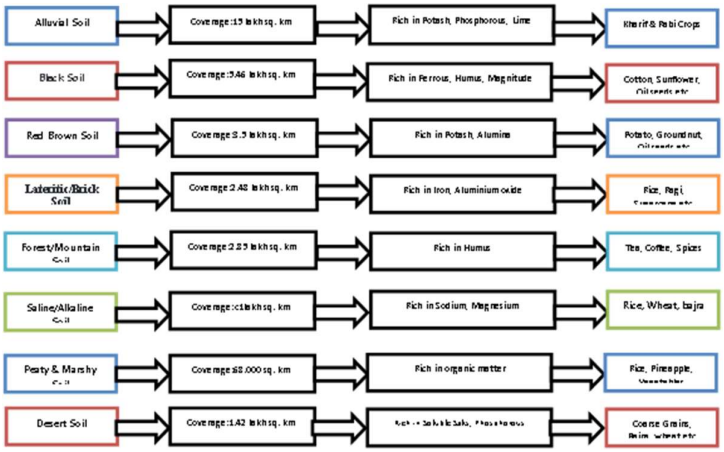

Figure-1: Types of Soil in India and its details

III. SYSTEM ARCHITECTURE

The system architecture for the "Soil Classification using Machine Learning Methods and Crop Suggestion" web application encompasses the high-level structure and components that enable its functionality [11]. This architecture is designed to provide a user-friendly and efficient platform for farmers and agriculture enthusiasts. At its core, the system utilizes machine learning techniques to analyse soil properties, including Nitrogen (N), Phosphorus (P), Potassium (K), pH levels, moisture content, and temperature values. These attributes serve as input data for the machine learning algorithms, which are responsible for making predictions and recommendations.

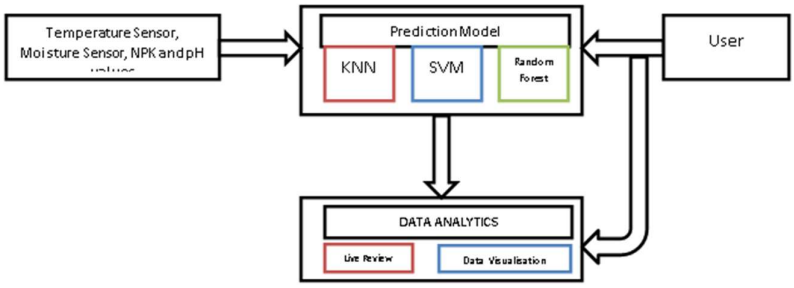

Figure-2: System Architecture

The system's hypermedia structure defines the organization of web pages, content, and user interactions. It ensures that users can seamlessly navigate through the application, access various features, and input their soil data for analysis. User interactions are managed through components that handle registration, login, and data input. Internal processing tasks involve the application's ability to process and analyze the input soil data. K-NN helps in suggesting suitable crops based on the similarity of soil properties, while SVM assists in classifying soil conditions and crop suitability as shown in Figure-2 [12].

The web application's architecture also addresses content presentation as shown in Figure-3 and Figure-4. It ensures that the results of the soil analysis are presented to users in an easily comprehensible format. Users can view the recommended crop for their soil, receive estimates of crop yields, and access information about the approximate amount of fertilizer needed for successful cultivation. Overall, the system architecture is designed using Visual Basic to streamline the process of soil analysis, crop selection, and yield prediction [13]. By harnessing the power of machine learning and providing users with actionable insights, the web application aims to empower individuals in the agricultural sector to make informed decisions, ultimately contributing to improved crop productivity and agricultural sustainability.

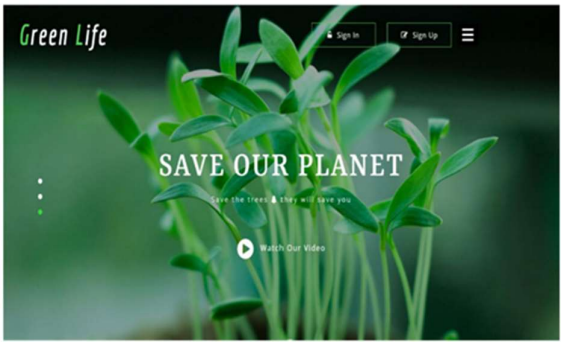

Figure-3: Web Application's Architecture

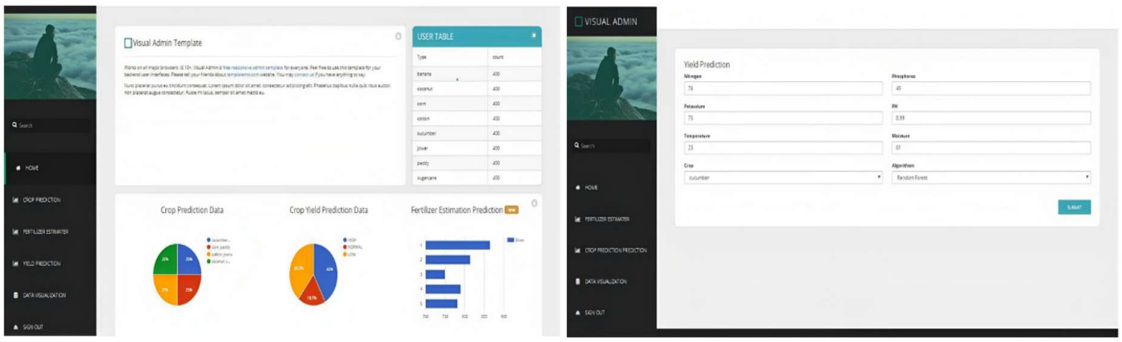

Figure-4: Window to enter Data Fields and View Results

IV. IMPLEMENTATION

User first registers and then login using the register data. The register data is stored in data base. It checks from the data base whether the user is registered or not if not, it says user does not exist. User uploads the soil data and prediction of crop, fertilizer and yield is done by using three algorithms by using the data set from csv file. Data preprocessing is done before the data set is used for prediction. K-Nearest Neighbor (K-NN) stands as one of the most straightforward Machine Learning algorithms, rooted in the realm of Supervised Learning techniques. The K-NN algorithm operates under the assumption of similarity between new data

or cases and existing data, classifying the new case into the category that is most similar to the available categories. K-NN can be used for both Regression and Classification tasks, although it is primarily employed for solving Classification problems. The functioning of the K-NN algorithm can be explained as follows: Initially, the appropriate value for K must be chosen. Ideally, K should be an odd number, and a larger K generally leads to higher accuracy. The algorithm calculates the Euclidean distance between the test data point and all other data points. For instance, if there are two points, A (x1, y1) and B(x2,y2), the Euclidean distance between them is in a two-dimensional space, where (x1, y1) and (x2, y2), are the coordinates of the two points. The Euclidean distance serves as a fundamental metric in K-NN, as it measures the proximity between data points, guiding the algorithm in classifying the new case based on the "K" nearest neighbors. Random Forest is a crucial component in our system, serving the purpose of accurate prediction of crop yields and fertilizer requirements based on soil attributes. In our problem context, Random Forest is utilized as follows:

Training the Model: A Random Forest model is trained using historical data that includes soil properties, such as Nitrogen ((N)), Phosphorus ((P)), Potassium ((K)), pH, moisture, and temperature, along with corresponding crop yields. This historical dataset serves as the training data for the model [14]. Predicting Crop Yields: Once the model is trained, it can predict crop yields for new soil samples. Given the input soil attributes, the Random Forest model estimates the yield using the following equation:

Yield = RF_Model (N, P, K pH, moisture, temperature) _____Eqn.1

Here, RF_Model represents the trained Random Forest model, and N, P, K, pH, moisture and temperature are the input soil attributes.

Fertilizer Estimation: Indeed, the Random Forest model's ability to estimate the required amount of fertilizer based on soil properties and yield predictions is a valuable asset for farmers. Fertilizer management is a critical aspect of modern agriculture, as it directly impacts crop health, yield, and resource efficiency.

Support Vector Machine (SVM) plays a pivotal role in our system, primarily responsible for classifying soil conditions and providing recommendations for suitable crop choices. SVM, a robust classification algorithm, excels in its ability to identify the optimal hyperplane that effectively segregates data into distinct classes. This capability makes SVM a valuable tool for accurately categorizing soil attributes and assisting farmers in making informed decisions regarding crop selection based on their specific soil conditions. In our problem context, SVM is applied as follows:

Crop = SVM_Model (N, P, K, pH, moisture, temperature) _____Eqn.2

Training the Model: An SVM model is trained using labeled data that includes soil properties and the corresponding suitable crops. The SVM model learns to classify soil conditions based on these attributes.

Crop Classification: Given a set of soil properties, the SVM model classifies the soil into one of the suitable crop categories [15]. This classification is based on maximizing the margin between different classes and is represented as:

Here SVM Model represents the trained SVM model, and N, P, K, pH, moisture and temperature are the input soil attributes. By integrating Random Forest for yield prediction and SVM for crop suggestion into the system, farmers and agricultural enthusiasts can make informed decisions about crop selection, fertilizer application, and yield expectations based on

soil properties, contributing to more efficient and sustainable farming practices.

V. SYSTEM IMPLEMENTATION

a. Data Pre-processing: Data pre-processing is a crucial data mining technique employed to convert raw data into a format that can be effectively utilized by the system. Raw data, often collected from the real world, is typically incomplete and may contain errors.

b. Model Fitting: The system employs three machine learning algorithms: Naïve Bayes, Random Forest, and K-Nearest Neighbors (KNN).

c. Crop Prediction: using attributes such as Nitrogen (N), Phosphorus (P), Potassium (K), moisture, temperature, and pH levels, the three machine learning algorithms (Naïve Bayes, Random Forest, and KNN) are applied to predict live crop data based on the training dataset. The choice of algorithm depends on its accuracy in making predictions.

d. Fertilizer Estimation Prediction: Similar to crop prediction, this step uses attributes including NPK levels, moisture, temperature, pH, and crop information to predict the fertilizer estimation. Again, the three machine learning algorithms are applied, and the choice of algorithm is determined by its accuracy. Accurate fertilizer estimation is crucial for optimizing crop growth and yield.

e. Yield Prediction: Utilizing NPK levels, moisture, temperature, pH, and crop attributes, the system employs the same three machine learning algorithms to predict crop yields. The algorithm selection is based on accuracy. Yield prediction is essential for farmers as it allows them to anticipate and plan for their crop harvests.

f. Comparative Analysis: In the final step, all three algorithms (Naïve Bayes, Random Forest, and KNN) are used for crop prediction, fertilizer estimation, and yield prediction.

In summary, the system's implementation involves user registration, data pre-processing to clean and prepare the data, and the application of machine learning algorithms (Naïve Bayes, Random Forest, and KNN) to predict crop information, fertilizer estimates, and crop yields. Comparative analysis helps determine the most accurate algorithm for each prediction task, ensuring the system's reliability and usefulness to users in the field of agriculture.

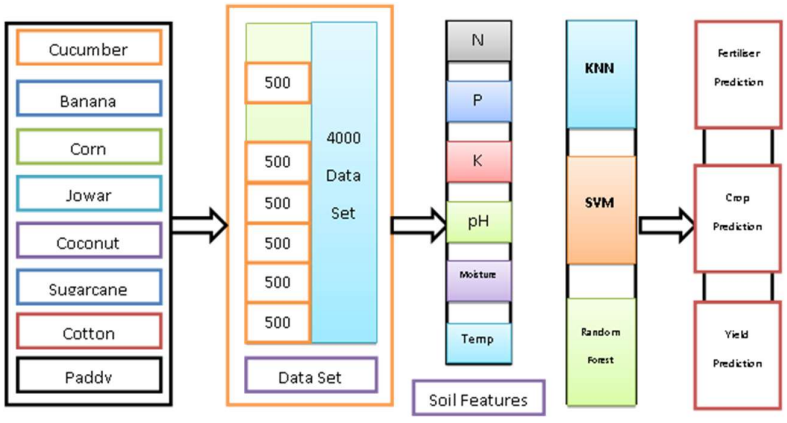

Figure-5: Implementation Model of the Proposed System

VI. RESULTS AND DISCUSSIONS

This section presents the outcomes derived from the practical implementation of the system and conducts a comparative analysis of the performance of three prominent machine learning algorithms: Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), and Random

Forest. These algorithms are employed to tackle critical agricultural decision-making tasks, namely predicting optimal crop selection, estimating fertilizer requirements, and forecasting crop yields.

For any given values of N, P, K, pH, Temperature and Moisture content, two suitable crops, approximate yield and fertiliser is predicted with the assistance of any one chosen machine learning algorithms. The Results is exhibited in Table 1. Accuracy of Algorithms in Crop Prediction, Fertilizer Estimation, and Yield Prediction is tabulated in Table 2.

TABLE 1: Results Summary of Machine Learning Algorithm for different input data

N	P	K	pH	Temperature	Crop Predicted (K-NN)	Crop Chosen	Fertilizer Estimation (SVM)	Crop Chosen	Yield Prediction
73	50	74	5.76	24	Cucumber Banana	Jowar	2-Moderate	Paddy	Low
49	47	75	7.11	13	Corn Paddy	Corn	1-Least	Corn	High
38	41	76	8.16	29	Cotton Jowar	Sugarcane	4-Slightly Moderate	Cucumber	Low
71	109	75	8.28	33	Coconut Sugarcane	Banana	5-Large	Sugarcane	Moderate
58	57	73	6.76	20	Corn Paddy	Coconut	1-Least	Jowar	High

TABLE 2: Performance Analysis of Algorithms in Crop Prediction, Fertilizer Estimation, and Yield

Algorithm	Crop Prediction (%)	Fertilizer Prediction (%)	Yield Prediction (%)
SVM	100	100	91.5
KNN	100	70	94.4
Random Forest	100	98	97.73

Figure-6 also shows the comparisons of all three algorithms with graph for the same. It also visualizes the accuracy of three different algorithms (SVM, KNN, and Random Forest) in three distinct prediction tasks (Crop Prediction, Fertilizer Estimation, and Yield Prediction). It plots lines for each algorithm, with markers denoting data points, and different line styles for each prediction type. The x-axis displays the algorithm names, while the y-axis represents accuracy percentages. The resulting chart helps compare and visualize the performance of these algorithms across the various prediction tasks, providing a clear overview of their relative accuracies.

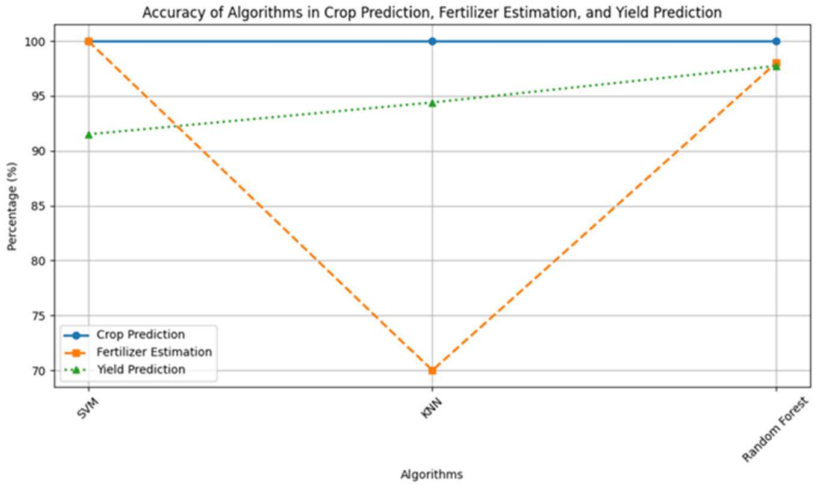

Figure-6: Comparisons of accuracies for Algorithms

For crop prediction, all three algorithms (SVM, K-NN, Random Forest) exhibit exceptionally high accuracy, with each achieving 100%. This indicates that any of these algorithms can be relied upon to accurately suggest suitable crops based on soil attributes. SVM demonstrates the highest accuracy of 100% in fertilizer estimation. This implies that SVM is the most reliable algorithm for estimating the approximate amount of fertilizer needed based on soil properties. K-NN, while having an accuracy of 70%, shows that it is still a viable option for fertilizer estimation but may not be as accurate as SVM. Random Forest achieves an accuracy of 98%, indicating a very high level of accuracy in estimating fertilizer requirements. For Yield Prediction, Random Forest stands out with the highest accuracy of 97.73% in yield prediction. This suggests that Random Forest is particularly effective at predicting crop yields under specified conditions. K-NN follows closely with an accuracy of 94.4%, indicating its suitability for yield prediction, although it falls slightly short of Random Forest. SVM, with an accuracy of 91.5%, also performs well in yield prediction but is slightly less accurate than both Random Forest and K-NN. The results reveal that all three algorithms excel in crop prediction, ensuring that users receive accurate recommendations for suitable crops based on soil attributes. For fertilizer estimation, SVM outperforms the other algorithms, making it the top choice when precise fertilizer recommendations are crucial. In yield prediction, Random Forest demonstrates the highest accuracy, making it the preferred algorithm when anticipating crop yields under specific conditions. The choice of algorithm should be tailored to the specific task. For example, if a user requires precise fertilizer estimation, SVM is recommended. For accurate crop yield prediction, Random Forest is the preferred option.

VII. CONCLUSION

In summary, the presented work showcases the exceptional performance of three machine learning algorithms, namely Support Vector Machine (SVM), K-Nearest Neighbours (K-NN), and Random Forest, in the context of agricultural decision-making. Across critical domains of crop prediction, fertilizer estimation, and yield forecasting based on soil attributes, these algorithms deliver highly accurate results. Crop prediction, a fundamental aspect of farming, enjoys flawless accuracy from all three algorithms, ensuring farmers receive precise

recommendations for optimal crop selection aligned with their soil characteristics. In the realm of fertilizer estimation, SVM emerges as the most reliable choice with a perfect accuracy score, while Random Forest follows closely with a remarkable 98% accuracy, both offering valuable insights into efficient resource management. Additionally, yield prediction benefits significantly from Random Forest, achieving an impressive 97.73% accuracy, providing farmers with robust forecasts to plan their harvests effectively. Collectively, these findings highlight the pivotal role of algorithm selection in modernizing agriculture, empowering farmers with data-driven tools to enhance productivity and sustainability.

REFERENCES

[1] Karmakar, S. & Bhunia, S. Energy Conservation in Farm Operations for Climate- Smart Agriculture. Handbook Of Energy Management in Agriculture. pp. 1-23 (2023)

[2] Jong, M., Guan, K., Wang, S., Huang, Y. & Peng, B. Improving field boundary delineation in ResUNets via adversarial deep learning. International Journal of Applied Earth Observation and Geoinformation. 112 pp. 102877 (2022)

[3] Cheng, M., Penuelas, J., McCabe, M., Atzberger, C., Jiao, X., Wu, W. & Jin, X. Combining multi-indicators with machine-learning algorithms for maize yield early prediction at the county-level in China. Agricultural And Forest Meteorology. 323 pp. 109057 (2022)

[4] Bhuyan, B., Tomar, R., Singh, T. & Cherif, A. Crop Type Prediction: A Statistical and Machine Learning Approach. Sustainability. 15, 481 (2022)

[5] Singh, A., Janu, N., Trivedi, S. & Jain, M. Precision agriculture and machine learning. 2022 IEEE World Conference On Applied Intelligence And Computing (AIC). pp. 659-664 (2022)

[6] Kwaghtyo, D. & Eke, C. Smart farming prediction models for precision agriculture: a comprehensive survey. Artificial Intelligence Review. 56, 5729-5772 (2023)

[7] Oikonomidis, A., Catal, C. & Kassahun, A. Deep learning for crop yield prediction: a systematic literature review. New Zealand Journal Of Crop And Horticultural Science. 51, 1-26 (2023)

[8] Singh, H., Halder, N., Singh, B., Singh, J., Sharma, S. & Shacham-Diamand, Y. Smart Farming Revolution: Portable and Real-Time Soil Nitrogen and Phosphorus Monitoring for Sustainable Agriculture. Sensors. 23, 5914 (2023)

[9] Durai, S. & Shamili, M. Smart farming using machine learning and deep learning techniques. Decision Analytics Journal. 3 pp. 100041 (2022)

[10] Garg, D. & Alam, M. A Sensor Data Acquisition System for Smart Agriculture. SN Computer Science. 4, 667 (2023)

[11] Shivakoti, M. & Others A Comparative Study on Prediction of Soil Health and Crop Recommendation using Machine Learning Models. International Journal Of Computing And Digital Systems. 14, 1-xx (2023)

[12] Raja, S., Sawicka, B., Stamenkovic, Z. & Mariammal, G. Crop prediction based on characteristics of the agricultural environment using various feature selection techniques and classifiers. IEEE Access. 10 pp. 23625-23641 (2022)

[13] Devan, K., Swetha, B., Sruthi, P. & Varshini, S. Crop Yield Prediction and Fertilizer Recommendation System Using Hybrid Machine Learning Algorithms. 2023 IEEE 12th International Conference On Communication Systems And Network Technologies (CSNT). pp. 171-175 (2023)

[14] Senapaty, M., Ray, A. & Padhy, N. IoT-Enabled Soil Nutrient Analysis and Crop Recommendation Model for Precision Agriculture. Computers. 12, 61 (2023)
[15] Bhat, S., Hussain, I. & Huang, N. Soil suitability classification for crop selection in precision agriculture using GBRT-based hybrid DNN surrogate models. Ecological Informatics. 75 pp. 102109 (2023)