Smart Crop Recommendation System with Plant Disease Identification

Ravikant Diwakar
School of Computer Science Engineering
Kalinga Institute of Industrial Technology
Bhubaneswar, India
diwakarr956@gmail.com

Amit Kumar
School of Computer Science Engineering
Kalinga Institute of Industrial Technology
Bhubaneswar, India
21051535@kiit.ac.in

Gaurav Kumar
School of Computer Science Engineering
Kalinga Institute of Industrial Technology
Bhubaneswar, India
21051649@kiit.ac.in

Aditya Choudhary
School of Computer Science Engineering
Kalinga Institute of Industrial Technology
Bhubaneswar, India
21051534@kiit.ac.in

Jaya Vashistha
School of Computer Science Engineering
Kalinga Institute of Industrial Technology
Bhubaneswar, India
21051672@kiit.ac.in

Abhaya Kumar Sahoo
School of Computer Science Engineering
Kalinga Institute of Industrial Technology
Bhubaneswar, India
abhaya.sahoofcs@kiit.ac.in

2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT) | 979-8-3503-7809-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICEECT61758.2024.10738975

Abstract—**Agriculture is the backbone of many countries, including India, and provides livelihoods to millions of people facing challenges such as climate change and plant disease outbreaks. Through research, a web application has been developed that provides real-time recommendations for crop selection based on various factors such as soil nutrients, temperature, humidity, pH levels, and rainfall. Recent advances in machine learning and artificial intelligence offer promising solutions to these problems, enabling accurate, data-driven decision-making in agriculture. These technologies have the potential to transform how we predict crop yields and detect plant diseases, thus improving agricultural practices. To achieve this, we trained and examined seven machine learning models, Decision tree, Naive Bayes, SVM, Logistic Regression, Random Forest, XGBoost, and KNN. Among these, Random forest gives the highest accuracy, making it the best choice for crop forecasting. In addition to crop recommendation, the web application also integrates a Plant Disease Identification system using Convolutional Neural Network (CNN). By analyzing leaf images, CNN detects and accurately classifies plant diseases, allowing farmers to intervene early to prevent crop losses. This study aims to empower farmers with accessible technology to make informed decisions, improve crop selection, and effectively cure plant diseases. Combining crop recommendation with disease detection, intelligent crop recommendation systems with plant disease detection contribute to sustainable agriculture, economic stability, and food security in India and beyond.**

Keywords— ***cnn, crop recommendation, machine learning, plant disease identification, random forest***

I. INTRODUCTION

The integration of machine learning and artificial intelligence (AI) technologies has ushered in a new era across various sectors, with agriculture emerging as a prominent beneficiary. Among the diverse applications of these technologies in agriculture, smart crop recommendations and plant disease identification stand out as pivotal advancements. Nowadays, the realm of machine learning and data science have revolutionized agricultural practices and one such innovation is the development of crop recommendation systems, which leverage predictive models to offer tailored guidance to farmers regarding the selection of crops best suited for their specific soil and environmental conditions and crop disease detection using image classification techniques. In this paper, we present a comprehensive Crop Recommendation System implemented using Python, incorporating various machine learning algorithms for classification tasks. The Crop Recommendation System aims to assist farmers in making informed decisions about crop selection by analyzing soil characteristics, climatic data, and geographical parameters [8]. The system allows users to input essential information such as soil nutrient levels (nitrogen, phosphorus, potassium), rainfall patterns, and geographical area. Leveraging a large dataset and employing multiple classification models, our system predicts the most suitable crops for cultivation in a given location. Furthermore, our Crop Recommendation System provides an innovative feature for crop disease detection using image classification techniques. By allowing users to upload images of leaves of the crop, showing symptoms of disease, the system identifies the likely ailment, enabling timely intervention and management practices. The Crop Recommendation System is complemented by a user-friendly web interface developed using Streamlit, facilitating seamless interaction and accessibility for farmers and agricultural stakeholders. Through this interface, users can effortlessly input data, view recommendations, and receive insights regarding crop health.

This paper is organized as follows : section 2 describes the literature review. Section 3 tells about proposed methodology. Section 4 explain the result analysis. Section 5 concludes the paper along with future scope.

II. LITERATURE REVIEW

The rise of smart farming technologies has completely transformed how we approach agriculture. These innovations offer fresh solutions to boost crop production, make better use of resources, and effectively tackle plant diseases. Let's dive into the core concepts and research behind smart crop recommendation systems and plant disease identification, giving you a thorough rundown of what's happening in this field. This early detection helps farmers intervene quickly, stopping the spread of disease and minimizing crop losses [11].

A. Machine Learning Algorithms

The integration of machine learning has truly transformed agriculture, offering advanced solutions for crop prediction and management. In this section, we'll delve into the various machine learning algorithms that have been successfully utilized to predict crop yields and determine the most suitable crops for cultivation, thereby enhancing agricultural productivity and sustainability [10].

1. Decision Trees

Decision trees are a foundational class of machine learning algorithms renowned for their effectiveness in classification and regression tasks [3]. In crop prediction, decision trees are invaluable tools for modeling the relationship between diverse soil and environmental factors and the potential yield of different crops. By partitioning the data into subsets based on input features like soil pH, moisture levels, and temperature, decision trees facilitate easy interpretation and visualization of the decision-making process. This simplicity makes them highly beneficial for farmers and agronomists seeking actionable insights.

2. Random Forests

Expanding upon the principles of decision trees, random forests employ an ensemble learning approach by combining multiple decision trees. This method enhances predictive accuracy and mitigates overfitting by aggregating the predictions of numerous trees [6]. Random forests excel in handling large datasets with numerous input variables, common in agricultural datasets encompassing various factors influencing crop growth and health. By reducing model variance, random forests yield more robust and reliable predictions, contributing significantly to improved crop yield forecasts.

3. Convolutional Neural Networks (CNN)

Inspired by the structure of the visual cortex in the human brain, Convolutional Neural Networks (CNNs) are a specialized type of neural network designed for processing structured grid data, such as images. In the context of agriculture, CNNs are widely used for plant disease identification. By analyzing images of crops, CNNs can learn to detect patterns indicative of various diseases, enabling early and accurate diagnosis. CNNs are especially effective for this task due to their ability to automatically extract relevant features from images, making them well-suited for analyzing plant images captured in the field or through remote sensing technologies.

B. Optimization of Web based Crop Recommendation Systems

In the realm of web-based smart crop recommendation systems, optimizing the performance and user experience of

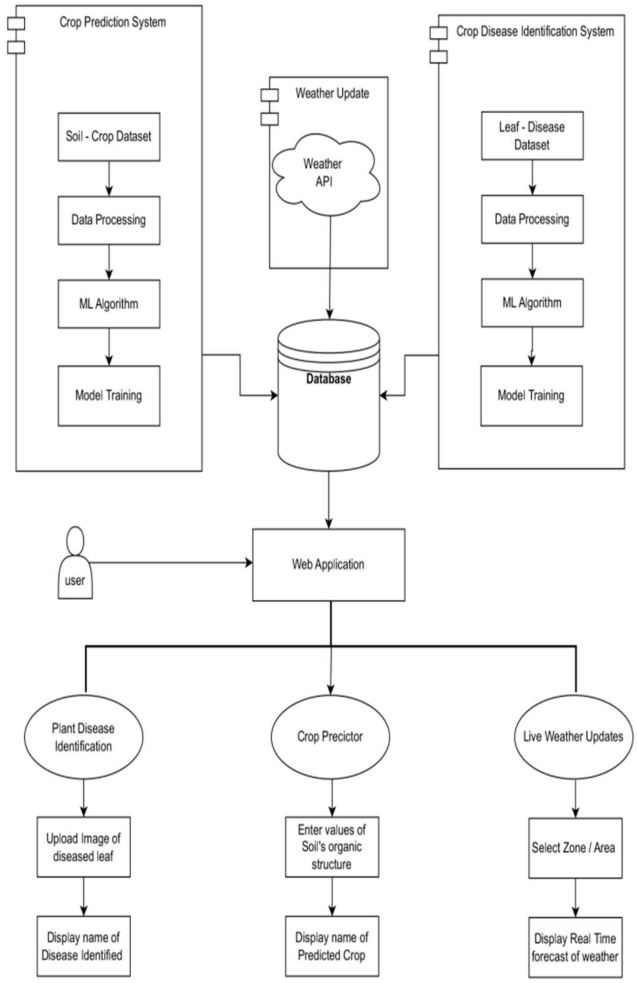

Fig. 1. System architecture

these platforms is paramount for their effective utilization by farmers and agricultural stakeholders. This optimization involves several key strategies. Firstly, ensuring responsive design is crucial to enable seamless access to the application across various devices, including desktop computers, tablets, and smartphones "Fig.1". Additionally, prioritizing fast loading times enhances user engagement and satisfaction, particularly for farmers in remote areas with limited internet connectivity. Lastly, integrating external data sources, such as weather forecasts, enriches the system with comprehensive and up-to-date information to support informed decision-making.

III. PROPOSED STUDY

The implementation serves as a pivotal factor inside the framework of the Smart Crop Recommendations with Plant Disease Identification. This phase encapsulates the interpretation of theoretical underpinnings and methodologies elucidated in preceding sections into tangible packages [10]. It encompasses the systematic improvement of the system, rigorous execution of trying out protocols, and meticulous analysis of outcomes to check adherence to stipulated necessities. The subsequent subsections provide an exhaustive elucidation of the methodologies embraced, the checking out and verification methodologies deployed, the complete analysis of effects derived, and the first-class guarantee mechanisms enacted to uphold the reliability and efficacy of the gadget [8].

A. Methodology

1. Datasets

This dataset consists of 2200 rows in total. Each row has 8 columns representing Nitrogen, Phosphorous, Potassium, Temperature, Humidity, PH, Rainfall and Label NPK(Nitrogen, Phosphorous and Potassium) values represent the NPK values in the soil. Temperature, humidity and rainfall are the average values of the sorroundings environment respectively. PH is the PH value present in the soil. The Label column tells us the type of crop that's best suited to grow based on these conditions. Label is the value we will be predicting.

Dataset for crop disease identification consists of the secondary dataset encompasses a collection of 70,295 plant snap shots exhibiting a spectrum of illnesses. This dataset amounts to five gigabytes in length, all standardized to a decision of 128x128 pixels. It encompasses a complete of 38 distinct instructions, which includes 14 distinct plant types and 26 identifiable illnesses. Following training, the set of rules demonstrating the highest accuracy became selected for similarly evaluation and application.

2. Data Preprocessing

Data preprocessing, the procedure of reworking uncooked information right into a format amenable to evaluation by way of analysts and records scientists for the purpose of making use of machine learning algorithms to derive insights or expect effects, is imperative to statistics education. In Plant disease identification model the raw pictures acquired from the dataset may also incorporate noise, therefore necessitating preprocessing earlier than integration into the mastering module.

3. Train and Test Split

The technique includes partitioning the dataset into awesome education and trying out subsets using the `train_test_split()` method to be had inside the scikit-examine module. Specifically, inside this framework, the dataset comprising 2200 information factors has been apportioned such that 80% of the dataset, equating to 1760 records factors, is allotted to the training subset, while the closing 20%, amounting to 440 information points, paperwork the trying out subset.

4. Predicting the model

In the context of forecasting the chance of a particular final results, the time period "prediction" denotes the result acquired from an algorithm subsequent to its schooling on a previous dataset and its next software to novel statistics. The prediction of the version is accomplished via the usage of the predict technique, which operates on the test feature dataset. This process yields an output within the form of an array containing the anticipated values [9].

Model evaluation of plant disease to examine the version's performance, the following steps are carried out:

- Initially, 80% of the pics from an excellent dataset are allocated for training purposes, at the same time as the ultimate 20% are reserved for validation.
- Validation statistics is utilized to assess accuracy by using employing the expect function and exactly extracting features.
- Additional photos are captured to affirm detection accuracy once best outcomes are obtained from the validation process.

5. Confusion Matrix and Classification Report

The strategies of Confusion Matrix and Classification Report are imported from the metrics module within the scikit-learn library. They are computed via using the actual labels present in the check datasets along the expected values [1]. The Confusion Matrix affords a comprehensive breakdown of the frequency of real negatives, false negatives, real positives, and false positives within the dataset [1].

Precision denotes the capability of a classifier to correctly pick out the wide variety of fine predictions among the ones which might be deemed fantastic which is shown in "(1)". It is computed because the ratio of real positives to the sum of genuine positives and fake positives for every class [1]. Where the accuracy of Precision-Positive Predictions; True Positive (TP) and False Positive (FP)

$$\text{Precision} = \frac{TP}{\text{TP} + FP} \tag{1}$$

Recall, in the context of classification, signifies the classifier's ability to correctly identify all high quality times from the confusion matrix which is shown in equation. Mathematically, it's miles computed because the quotient of real positives divided through the sum of proper positives and fake negatives for every magnificence which is shown in "(2)".

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

Where recall is the proportion of correctly detected positives; False Negative, or FN

The F1-Score rating represents a weighted harmonic suggest of precision and don't forget, in which a price of zero. 0 denotes the poorest overall performance, at the same time as 1.0 signifies most efficient performance. As precision and do not forget are imperative additives in its calculation, F1 ratings commonly yield lower values as compared to accuracy measurements which is shown in "(3)".

$$\text{F1} - \text{Score} = \frac{2*PR}{(P+R)} \tag{3}$$

Where P-Precision; R-Recall

6. Accuracy

Model accuracy, a metric indicating the proportion of correct predictions relative to the full number of predictions made, is a fundamental aspect of model assessment. It is computed the use of the `accuracy_score()` technique available in the scikit-examine metrics module which is shown in "(4)".

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{4}$$

Where TP-True Positive; FP-False Positive; TN-True Negative; FN-False Negative

IV. Result Analysis

Seven classification algorithms have been selected, and their respective accuracies have been assessed. For the second task, the preeminent photograph classification convolutional neural network (CNN) architectures were skilled. The ensuing tables gift the corresponding outcomes.

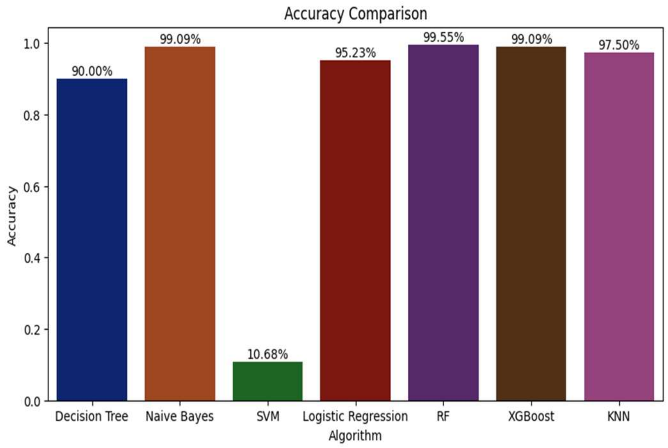

Fig. 2. Accuracy comparison

Table.1 offers the accuracy of crop advice responsibilities across diverse algorithms [4]. This desk unequivocally demonstrates that the random wooded area set of rules surpasses all others, accomplishing a extraordinary accuracy of 99.54%.

TABLE I. ACCURACY VS ALGORITHMS

Algorithm	Accuracy
Decision Tree	90.0
Gaussian Naive Bayes	99.09
Support Vector Machine (SVM)	10.68
Logistic Regression	95.23
Random Forest	99.55
XGBoost	99.09
KNN	97.50

"Fig.3" show that Random Forest models consistently perform well across different crops, often outperforming other models. While some models excel with specific crops, they show varying accuracy overall. Random Forest stands out for its stability and versatility, making it a reliable choice for a wide range of crops. This highlights the model's effectiveness in providing consistent accuracy compared to the more specialized performance of other models.

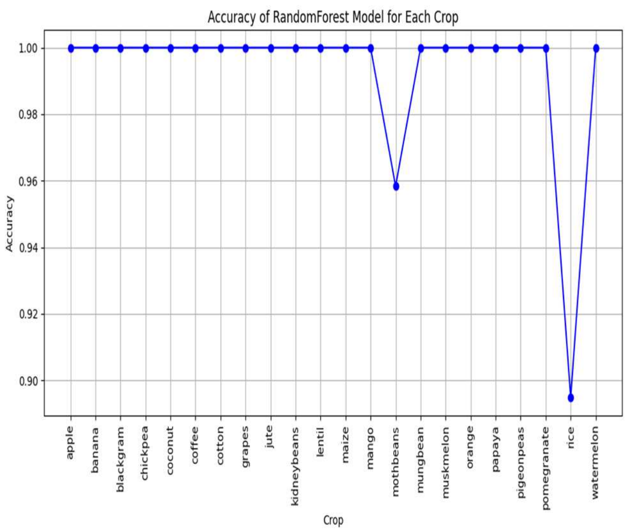

Fig. 3. Accuracy vs Crop graph for random forest model

In this study, we employed a convolutional neural network to detect plant diseases from a dataset containing images of various plant species. The dataset, consisting of 17,572 images across 38 distinct classes, was organized into a validation set using the tf.keras.utils.image_dataset_from_directory function, ensuring that each image was resized to 128x128 pixels to match the input requirements of our model. The class names, representing various plant diseases and healthy conditions, were automatically inferred from the directory structure. We utilized a pre-trained CNN model, which was loaded from a saved.keras file, to perform the predictions [5]. Here we use dropout layer to prevent overfitting by randomly setting a fraction of input unit to zero at each updated during training. For visualization and testing purposes, a sample image of a potato plant affected by early blight was processed. This image was read, converted from BGR to RGB using OpenCV, and resized to the required dimensions. The preprocessed image was then fed into the model to obtain predictions. The model's output was a probability distribution across all classes, from which the class with the highest probability was identified using the np.argmax function. The predicted class label, in this case, was displayed alongside the original image, confirming the model's ability to accurately identify the plant disease. This approach demonstrates the efficacy of CNNs in the automated detection of plant diseases, which can significantly aid in early diagnosis and management, ultimately contributing to better agricultural practices. "Fig.4" illustrates an example of a diseased plant image used for plant disease identification.

The detailed features in the image, such as spots, discoloration, and other symptoms, enable the models to learn and identify various plant diseases, contributing to smarter crop management and healthier yields.

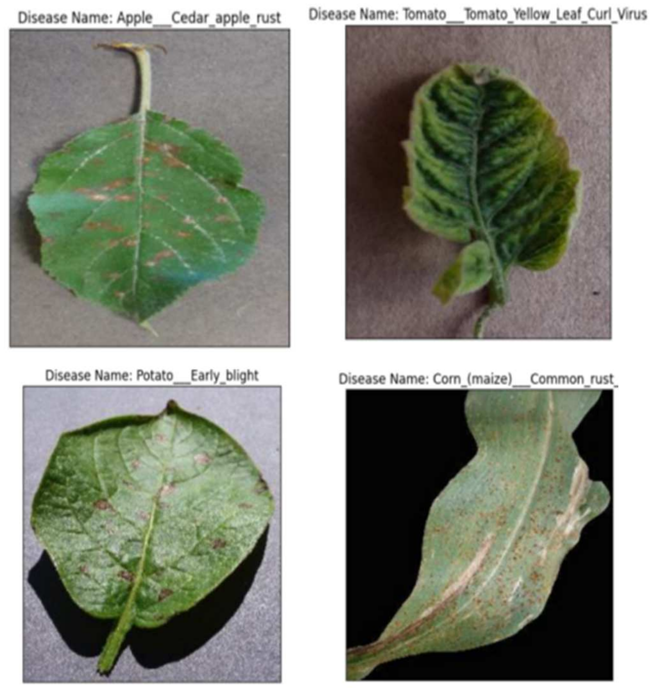

Fig. 4. Example of diseased image

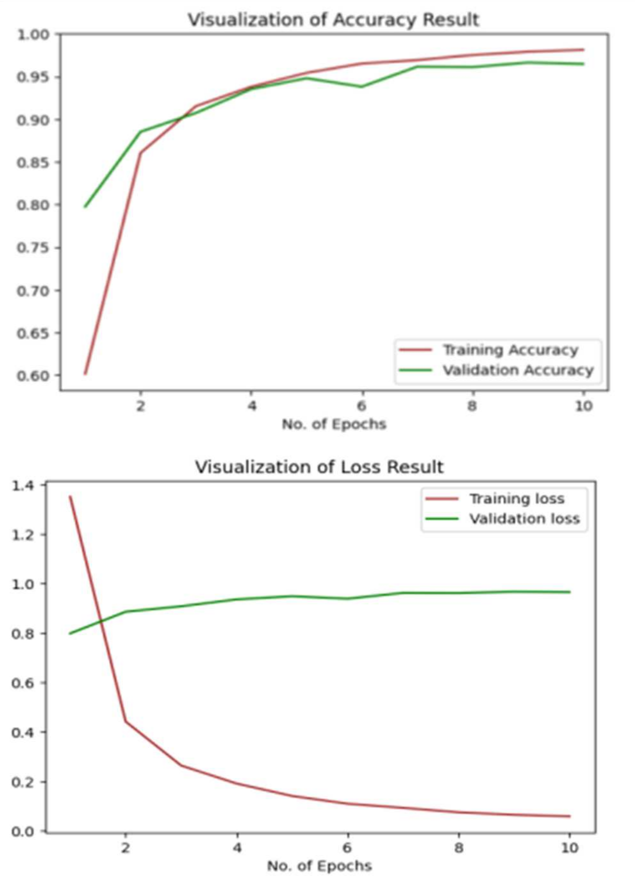

Fig. 5. Training and Validation (accuracy and loss) simple CNN

The graph depicts the performance of a plant disease detection model using a convolutional neural network (CNN). The model's training and validation accuracy are plotted alongside the training and validation loss. It is important for the training and validation accuracy to be high, and the training and validation loss to be low [5]. High training and validation accuracy indicate that the model can accurately distinguish between healthy and diseased plants [12]. Low training and validation loss signifies that the model is learning from the training data and generalizes well to unseen data. Learning curves for a deep learning algorithm show how well it is learning the dataset throughout training in an incremental manner.

The accuracy and loss of CNN during training and validation are shown in "Fig.5". Although there was little volatility during the validation test, the loss curve shows that the training and validation losses dropped over time and that the interval between them was short over the experiments. The training and validation accuracy and loss for the proposal are displayed in "Fig.5". In contrast, the loss graph shows strong fitting for both training and validating loss curves [2].

V. Conclusion and Future Scope

In conclusion, the project has successfully developed a sophisticated system for smart crop recommendation and disease detection using advanced technologies and techniques. By combining machine learning algorithms, and real-time data processing, the project provides insights that can be applied to improving crop management and reducing the impact of plant diseases. Modular and scalable design patterns adopted while adhering to industry standards ensured the reliability, scalability and maintainability of the solution Using best practices of database design and user-centric interface design principles, provided project's farmers regardless of their technical skills. facilitated seamless communication and usability. Overall, precision agriculture has improved dramatically through integration of strategies, innovation and adherence to industry standards Integration of cutting-edge technologies for smart crops recommendations and disease detection have provided farmers with the necessary tools for informed decision making and sustainable agricultural practices.

In future, the dataset can be frequently updated with new samples and cases. The Performance of Crop Recommendation and Plant Disease Identification, implemented by Machine learning classification and prediction models can be improved by integrating it with practices, such as Multispectral and Hyperspectral imaging data for insights on plant's health and Reinforced Learning to optimize the performance of existing strategies. Including the Temperature and rainfall data of the region in which the farm is can also help in suggesting best suited crop for them, potentially reducing crop losses and improving yields. More features like Profit oriented Crop prediction for the farmers based on the selling price of the crops and investment needed for that crop can be added. We can also include Crop Growth Schedule managers in getting the new farmer acquainted with timeline of different process in farming and to help in watering, fertilizing and pest cleaning timely and regularly. The concepts of edge computing can also be implemented. By integrating the system with a moveable device with camera and sensors can also help the farmers in real time monitoring of their farms. Furthermore, development of a collaborative platform, inviting various farmers, researchers and agronomists to share data and insights on various cases and info about any new ones will also have a significant impact as such Crowd-sourced data collection could also help in significantly improving model accuracy.

References

[1] M. Shripathi Rao, A. Singh, N. V. Subba Reddy, and D. U. Acharya, "Crop prediction using machine learning," Journal of Physics: Conference Series, vol. 2161, no. 1, p. 012033, Jan. 2022.

[2] Md. Manowarul Islam et al., "DeepCrop: Deep learning-based crop disease prediction with web application," Journal of Agriculture and Food Research, vol. 14, pp. 100764–100764, Dec. 2023.

[3] A. Kar, N. Nath, U. Kemprai, and Aman, "Performance Analysis of Support Vector Machine (SVM) on Challenging Datasets for Forest Fire Detection," International Journal of Communications, Network and System Sciences, vol. 17, no. 2, pp. 11–29, Feb. 2024.

[4] G. Rani, E. T. Venkatesh, K. Balaji, Balasaraswathi Yugandher, Adiki Nithin Kumar, and M Sakthimohan, "An automated prediction of crop and fertilizer disease using Convolutional Neural Networks (CNN)," 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Apr. 2022.

[5] D. Tirkey, K. K. Singh, and S. Tripathi, "Performance analysis of AI-based solutions for crop disease identification, detection, and classification," Smart Agricultural Technology, vol. 5, p. 100238, Oct. 2023.

[6] K. Ramu and K. Priyadarsini, "A Review on Crop Yield prediction Using Machine Learning Methods," 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), Oct. 2021.

[7] A. Bouguettaya, H. Zarzour, A. Kechida, and A. M. Taberkit, "A survey on deep learning-based identification of plant and crop diseases from UAV-based aerial images," Cluster Computing, Aug. 2022.

[8] K. Jhajharia, P. Mathur, S. Jain, and S. Nijhawan, "Crop Yield Prediction using Machine Learning and Deep Learning Techniques," Procedia Computer Science, vol. 218, pp. 406–417, 2023.

[9] M. Shripathi Rao, A. Singh, N. V. Subba Reddy, and D. U. Acharya, “Crop prediction using machine learning,” Journal of Physics: Conference Series, vol. 2161, no. 1, p. 012033, Jan. 2022.

[10] S. M. PANDE, P. K. RAMESH, A. ANMOL, B. R. AISHWARYA, K. ROHILLA, and K. SHAURYA, “Crop Recommender System Using Machine Learning Approach,” 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Apr. 2021.

[11] C. Raju, A. D.v., and A. P. B.v., “CropCast: Harvesting the future with interfused machine learning and advanced stacking ensemble for precise crop prediction,” Kuwait Journal of Science, p. 100160, Dec. 2023.

[12] P. Chauhan, Hardwari Lal Mandoria, A. Negi, and R. S. Rajput, “Plant Diseases Concept in Smart Agriculture Using Deep Learning,” Advances in environmental engineering and green technologies book series, pp. 139–153, Oct. 2020.