International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Agrovision: Enhancing Agricultural Precision through Machine Learning

Abhishek K V
Information and Science and Engineering
Malnad College of Engineering
Hassan-573202, India
abhishekkv812003@gmail.com

Akhil Teotia
Information and Science and Engineering
Malnad College of Engineering
Hassan-573202, India
akhilteotia19@gmail.com

Chandan K
Information and Science and Engineering
Malnad College of Engineering
Hassan-573202, India
chandansatish1906@gmail.com

Likhith H V
Information and Science and Engineering
Malnad College of Engineering
Hassan-573202, India
likhithhv8@gmail.com

Sindhu K S
Information and Science and Engineering
Malnad College of Engineering
Hassan-573202, India
sk@mcehassan.ac.in

***Abstract*— Machine learning (ML) and Deep learning (DL) have transformed agriculture by addressing challenges such as soil fertility assessment, crop recommendation, yield prediction, and plant disease detection. These technologies enable farmers to make informed decisions, enhancing productivity and sustainability. Notable advancements include Random Forest achieving 84% accuracy in soil fertility prediction and CNN-based soil classification reaching 95.21% accuracy, both integrated into accessible web platforms. Yield prediction models, such as Random Forest and LSTM, have delivered up to 98.96% accuracy, while innovations in disease detection in plants using CenterNet with DenseNet-77 achieved a mean Average Precision (mAP) of 99%. These tools empower farmers by streamlining processes, optimizing resource utilization, and reducing losses due to uncertainties. Collectively, the integration of ML and DL is shaping a more resilient and efficient agricultural sector.**

Keywords*— Machine learning, deep learning, soil fertility, crop recommendation, yield prediction, plant disease detection, precision agriculture.

I. Introduction

The global agricultural landscape faces significant challenges due to population growth, environmental changes, and shrinking arable land. Traditional farming practices often fall short of addressing these issues, leading to inefficient resource use and economic losses. Recent advancements in Machine learning (ML) and Deep learning (DL) offer transformative solutions, enabling farmers to adopt data-driven approaches for sustainable farming.

ML techniques provide robust frameworks for soil fertility analysis, crop recommendation, and yield prediction, while DL models enhance precision through advanced pattern recognition. For example, Random Forest algorithm has been applied to predict soil fertility with 84% accuracy, offering actionable insights through user-friendly web applications. Similarly, CNN-based soil classification achieved 95.21% accuracy, aiding in crop selection and bypassing intermediaries through direct farmer-to-buyer platforms.

In yield prediction, models like Random Forest and LSTM have delivered exceptional results, with accuracy levels as high as 98.96%. Additionally, breakthroughs in plant disease detection using custom CenterNet frameworks with DenseNet-77 have demonstrated state-of-the-art performance, achieving mAP of 99% and robust classification under challenging conditions. These advancements address critical agricultural challenges such as optimizing resource allocation, reducing losses, and enhancing productivity.This paper explores the integration of ML and DL in agriculture, highlighting their potential to revolutionize farming by providing scalable, precise, and actionable solutions. By adopting these technologies, the agricultural sector can transition toward a more efficient, sustainable, and data-driven future.

II. Related Works

A. Paper Title: Soil Analysis and Crop Recommendation using Machine Learning, Authors: Aditya Motwani, Param Patil, Vatsa Nagaria, Shobhit Verma, Sunil Ghane, Year of publication: 2022

Description: This paper introduces a system that assists Indian farmers by analyzing soil data and recommending suitable crops using machine learning. The study focuses on optimizing crop selection to enhance yield and profitability.

Methodology: A CNN model was used to classify soils into four types (Red, Black, Clay, and Alluvial) with an impressive accuracy of 95.21%.

Random Forest was applied to predict crop yields based on soil type, location, and cultivation conditions.

A web portal was created to connect farmers with buyers, eliminating the need for middlemen and increasing farmers' profits.*Limitations:* The system was limited to major crops and did not include mixed soil types or consider trade-offs between crop quality and revenue.

The dataset used for training was relatively small, which may affect its scalability.

Key Insights: The CNN model was highly effective in soil classification, while the Random Forest model achieved 75% accuracy for crop yield prediction.

The integration of a farmer-friendly web portal provided a practical solution for crop selling without intermediaries.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

B. Paper title: Prediction of Soil Fertility using ML Algorithms and Fertilizer Recommendation System , Authors: Madhumathi R., Dhanishtaa R., Elakkiya E., Gunashri R., Arumuganathan T, Year of publication : 2023

Description: This study addresses the prediction of soil fertility using machine learning methods, focusing specifically on the macro-nutrients—Nitrogen (N), Phosphorus (P), and Potassium (K). It also proposes a fertilizer recommendation framework to enhance agricultural productivity and optimize nutrient usage.
Methodology: Soil data was collected from Kaggle and divided into training and testing sets.
Several algorithms, including Random Forest, Decision Tree, SVM, KNN, and Naïve Bayes, were tested for classification tasks.
Random Forest achieved the highest accuracy of 84%, making it the most reliable algorithm for the study.
A web-based application was developed to offer fertilizer recommendations to farmers based on the model's predictions.
Limitations: The study focuses exclusively on macro-nutrients, neglecting other soil properties such as moisture, pH, and micronutrient levels.
Dynamic factors, such as changes in environmental conditions, are not considered.
Key Insights: The proposed system proved effective in predicting soil fertility, with Random Forest outperforming other methods.

C. Paper title: Predicting Agriculture Yields Based on Machine Learning Using Regression and Deep Learning, Authors: Priyanka Sharma, Pankaj Dadheech, Nagender Aneja, Sandhya Aneja, Year of publication : 2023

Description: This paper employs machine learning and deep learning to forecast crop yields, aiming to improve agricultural efficiency and decision-making for farmers. By using historical data on rainfall, soil, and meteorological factors, it provides actionable insights for better crop management.
Methodology: Machine learning methods, such as Decision Tree, Random Forest, and XGBoost, were used for yield prediction.
Deep learning models, including CNN and LSTM, were implemented to further enhance predictive accuracy.
Metrics like RMSE, MAE, and standard deviation were used to evaluate model performance.
Limitations: The reliance on diverse datasets makes it challenging for farmers with limited access to data.
The study did not incorporate external variables such as market prices or supply chain factors.
Key Insights: Random Forest delivered the best accuracy (98.96%), while CNN achieved the lowest loss (0.00060).
The combined use of machine learning and deep learning resulted in reliable and accurate predictions, highlighting the benefits of integrating multiple approaches.

D. Paper title: A Novel Deep Learning Method for Detection and Classification of Plant Diseases, Authors: Waleed Albattah et al., Year of publication : 2021

Description: This paper introduces an enhanced deep learning framework for detecting and classifying plant diseases with high precision. The system leverages a custom CenterNet model integrated with DenseNet-77 for effective feature extraction, targeting improved performance under real-world conditions.
Methodology: The framework was trained on the PlantVillage dataset, which includes 38 classes of healthy and diseased plants.
CenterNet with DenseNet-77 was used to create accurate heatmap-based localization and classification of plant diseases.
The system's robustness was evaluated under various challenging conditions, such as noise, brightness variations, and image distortions.
The model outperformed traditional object detection methods like YOLO and Faster-RCNN in both accuracy and computational efficiency.
Limitations: High computational requirements make the system less suitable for low-resource environments or mobile devices.
Further evaluation on datasets with greater real-world complexity is necessary to enhance scalability.
Key Insights: Achieved a mean Average Precision (mAP) of 99% and Intersection over Union (IoU) of 99.3%.
Demonstrated superior performance in challenging conditions compared to competing models.

III. METHODOLOGY

Innovative applications of machine learning (ML) and deep learning (DL) address key agricultural challenges like soil fertility prediction, crop recommendation, and plant disease detection. These methodologies involve systematic processes of data acquisition, preprocessing, model development, evaluation, and deployment.

A. Soil Fertility Prediction

- *Objective:* Analyze macro-nutrients (N, P, K) to assess soil fertility and recommend fertilizers.
- *Steps:* Data sourced from Kaggle, preprocessed for missing values, split into training/testing sets. Random Forest, SVM, and Naïve Bayes classified soil fertility levels, supported by a farmer-friendly web app.

B. Crop Recommendation System

- *Objective:* Recommend crops based on soil classification and yield prediction.
- *Steps:* Soil images processed using CNN for type classification. Random Forest analyzed yield data, and a web portal facilitated direct farmer-to-buyer connections.

C. Yield Prediction

- *Objective:* Predict crop yields using environmental and soil factors.
- *Steps:* Preprocessed datasets analyzed with Random Forest, Decision Tree, CNN, and LSTM. Performance evaluated using RMSE, MAE, and accuracy metrics.

D. Ensemble-Based Crop Recommendation

- *Objective:* Use ensemble methods (e.g., CHAID, Random Tree, KNN) for crop recommendations.

- *Steps:* Standardized soil/yield data aggregated predictions through majority voting.

E. Plant Disease Detection

- *Objective:* Localize and classify plant diseases using DL.
- *Steps:* Leveraged PlantVillage dataset, Custom CenterNet with DenseNet-77 for feature extraction and bounding box generation. Fine-tuned with MS-COCO weights, evaluated with mAP, IoU, precision, and recall.

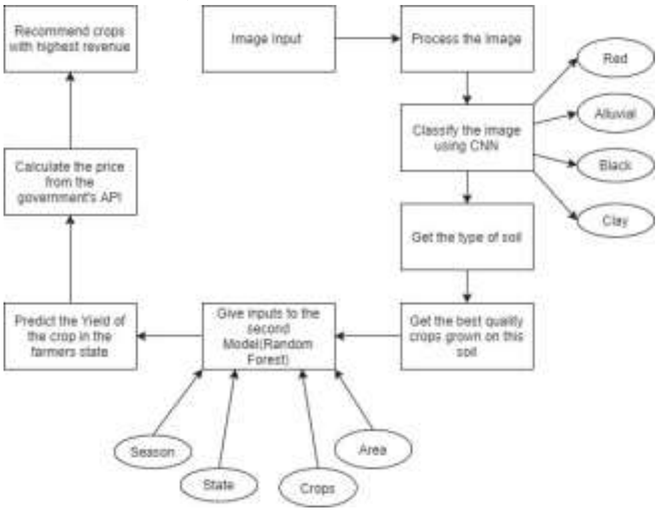

Fig. 1. Methodology used for crop recommendation

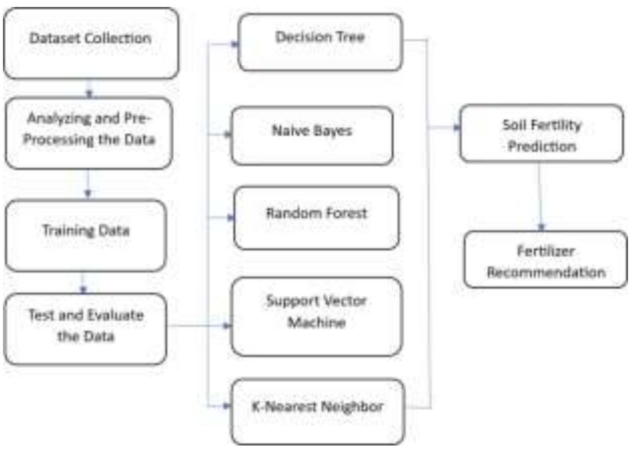

Fig. 2. Methodology used for Fertilizer recommendation

IV. ALGORITHMS USED

A. Random Forest Algorithm

Overview: Random Forest combines multiple decision trees, each trained on random subsets of data and features, to enhance prediction accuracy and reduce overfitting.
Applications: Predicting soil fertility by analyzing macro-nutrient levels.
Estimating crop yields using environmental and soil data.
Procedure: Sample subsets of the dataset to create individual decision trees.
Aggregate the outputs of all trees (e.g., majority voting for classification or averaging for regression).
Evaluate the combined model's accuracy, recall, and precision.

Formula for Prediction:

- *Classification:*
$\hat{y} = \text{mode}(f_1(x), f_2(x), \ldots, f_N(x))$

- *Regression:*

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} f_i(x)$$

Where:
N is the number of decision trees.
$f_i(x)$ is the prediction from the i^{th} tree.
Entropy for Splitting:

$$Entropy(S) = -\sum_{i=1}^{k} p_i \log_2(p_i)$$

pi: Proportion of samples belonging to class i.

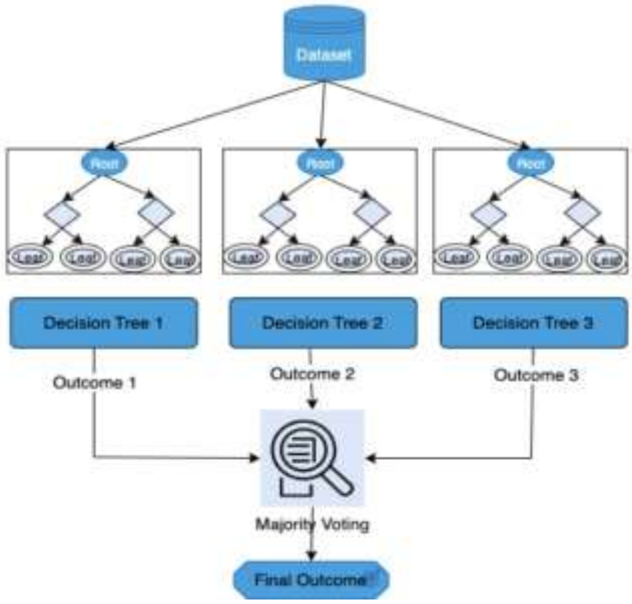

Fig. 3. Random forest

B. Convolutional Neural Network (CNN)

Overview: CNNs specialize in processing image data by extracting spatial features using convolutional layers and reducing dimensions with pooling layers.
Applications: Classifying soil types based on images.
Identifying patterns in soil data for crop recommendations.
Procedure: Preprocess soil images by resizing and normalizing their pixel values.
Pass the images through convolutional layers to extract features like edges and textures.
Use pooling layers to reduce the data's dimensions while retaining essential information.
Feed the output into fully connected layers for final classification.

The mathematical formula for Decision tree is:
Convolution:

$$Z[i,j] = \sum_{m=1}^{M}\sum_{n=1}^{N} X[i+m, j+n] \cdot W[m,n] + b$$

X: Input matrix.
W: Filter matrix (kernel).
b: Bias term.

ReLU Activation:
f(x)=max(0,x)
Pooling (e.g., Max Pooling):

$$P[i,j] = \max_{m,n \in k \times k} Z[i+m, j+n]$$

k: Pooling window size.

C. Decision Tree Algorithm

Overview: Decision Trees split data into branches based on feature thresholds, creating interpretable models for classification or regression.
Applications: Forecasting crop yields based on soil properties and weather data.
Identifying patterns in soil data for crop recommendations.
Procedure: Choose the best feature for splitting data using metrics like Information Gain or Gini Index.
Create branches for each split and repeat until all data is classified or a stopping criterion is met.
Use the tree to make predictions based on input conditions.

The mathematical formula for Decision tree is:
*Information Gain= Entropy(S)-[(Weighted Avg) *Entropy (each feature)*
Entropy: Entropy is a metric to measure the impurity in a given attribute. It specifies randomness in data. Entropy can be calculated as:
Entropy(s)= -P(yes)log2 P(yes)- P(no) log2 P(no)
Where,
S= Total number of samples
P(yes)= probability of yes
P(no)= probability of no

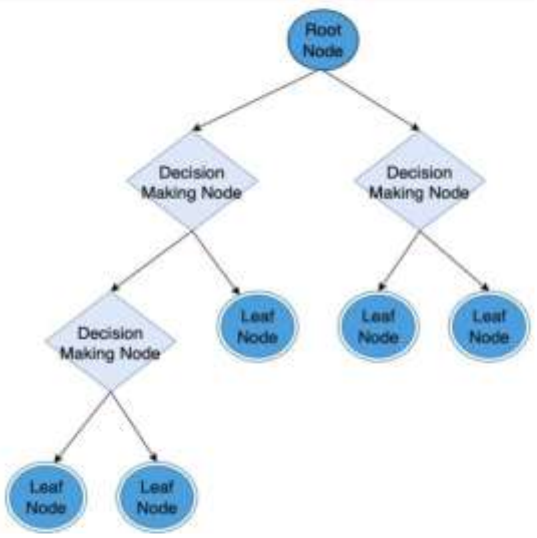

Fig. 4. Decision Tree

D. Long Short-Term Memory (LSTM)

Overview: LSTMs are a type of recurrent neural network (RNN) capable of learning sequential dependencies in time-series data.
Applications: Predicting crop yields by analyzing historical environmental trends.
Procedure: Preprocess data to create sequences suitable for time-series modeling.
Pass data through LSTM cells, where input, forget, and output gates manage information flow.
Generate predictions from the final output layer, optimized for long-term dependencies.
Core LSTM Equations:

1. Forget Gate:
$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$
2. Input Gate:
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$
3. Cell State Update:
$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$
4. Output Gate:
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t \cdot \tanh(C_t)$$

Where,
σ : Sigmoid activation.
tanh : Hyperbolic tangent activation.
W: Weight matrices.
b: Bias vectors.

E. Support Vector Machine (SVM)

Overview: SVMs classify data by finding the hyperplane that maximally separates different classes.
Applications: Classifying soil types or fertility levels based on features like nutrient content.
Procedure: Transform data into a higher-dimensional space using kernel functions.
Identify the hyperplane with the maximum margin between data classes.
Use the trained hyperplane to classify new data points.
The equation for the linear hyperplane can be written as:

$$w^T x + b = 0$$

The vector W represents the normal vector to the hyperplane. i.e the direction perpendicular to the hyperplane. The parameter b in the equation represents the offset or distance of the hyperplane from the origin along the normal vector w.
The distance between a data point xi and the decision boundary can be calculated as:

$$d_i = \frac{w^T x_i + b}{||w||}$$

where ||w|| represents the Euclidean norm of the weight vector w.

F. Ensemble Learning (CHAID, KNN, etc.)

Overview: Ensemble learning combines the predictions of multiple algorithms to improve robustness and accuracy.
Applications: Recommending crops by aggregating outputs from CHAID, Random Tree, Naïve Bayes, and KNN models.
Procedure: Train individual algorithms on the same dataset.
Combine their predictions using majority voting or weighted averaging.
Evaluate the overall model performance to ensure consistency.
General Formula:

$$\hat{y} = \sum_{i=1}^{N} w_i f_i(x)$$

Where:

w_i: Weight assigned to each model $f_i(x)$.
N: Total number of models.
For Bagging (e.g., Random Forest):
Equal weights ($w_i=1/N$).
For Boosting (e.g., AdaBoost):

$$w_i = \log\left(\frac{1-e_i}{e_i}\right)$$

e_i: Error rate of model i.

G. Custom CenterNet with DenseNet-77

Overview: A one-stage object detection model that integrates DenseNet-77 for efficient feature extraction and high-accuracy disease localization and classification.
Applications: Detecting and classifying plant diseases in the PlantVillage dataset.
Monitoring plant health in real-time for precision agriculture.
Procedure: DenseNet-77 extracted high-quality feature maps for accurate disease classification.
Heatmaps generated localized affected plant regions.
Multi-loss optimization was applied to improve model convergence and accuracy.

V. RESULTS

Machine learning and Deep learning techniques have demonstrated significant efficacy across various agricultural applications:

A. Soil Fertility Prediction:
Random Forest achieved 84% accuracy in predicting soil fertility (NPK levels).
A web application enabled real-time fertilizer recommendations for farmers.

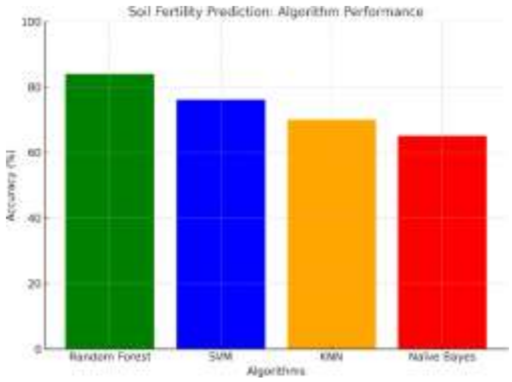

Fig. 5. Soil fertility prediction

B. Crop Recommendation and Soil Classification:
CNN achieved 95.21% accuracy in soil classification, and Random Forest delivered 75% accuracy for yield predictions.
The system improved farmer profitability through a direct farmer-to-buyer web portal.
C. Yield Prediction:
Random Forest reached 98.96% accuracy, while CNN minimized prediction loss to 0.00060.
LSTM effectively captured time-series patterns, enhancing yield prediction accuracy.
D. Ensemble Crop Recommendation:
Ensemble models combined CHAID, KNN, Random Tree, and Naïve Bayes, ensuring robust and accurate crop recommendations.

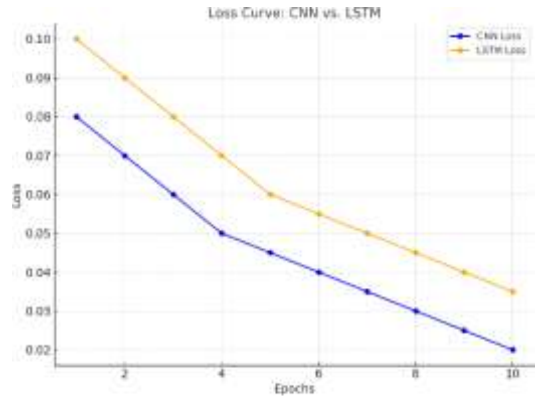

Fig. 6. CNN vs. LSTM Loss Curve

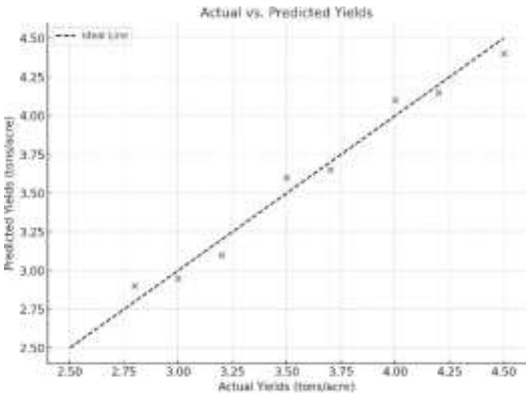

Fig. 7. Actual vs. Predicted Yields

E. Plant Disease Detection (PlantVillage Dataset):
Achieved mAP of 99% and IoU of 99.3%, outperforming Faster-RCNN (mAP 97%) and YOLOv3 (mAP 83%).
Maintained >98% precision and recall across all classes with efficient processing (0.21s/image).
Generated precise disease-localized bounding boxes, excelling in challenging conditions.
These results highlight the transformative impact of AI-driven solutions in modern agriculture.

VI. CONCLUSION

Machine learning and deep learning techniques have transformed agricultural practices by improving precision and efficiency in critical tasks such as plant disease detection.
Key insights from this study include:
Effectiveness of the Custom CenterNet Model:
Achieved a state-of-the-art mAP of 99% and IoU of 99.3%, demonstrating exceptional accuracy.
Delivered faster inference times and reduced computational overhead, making it practical for agricultural applications.
Real-World Applicability:
Robust performance under challenging conditions, including noise, brightness, and distortion.
Suitable for integration into smart farming systems to provide real-time disease monitoring.
Challenges and Future Work:
Current model scalability is limited for low-resource devices; future work aims to create lightweight variants.
Extending the framework to additional datasets and real-world scenarios will enhance versatility.
Exploring integration with IoT devices for seamless real-time plant health monitoring.

By addressing these challenges, the study lays a foundation for sustainable and efficient farming practices powered by advanced AI techniques.

References

[1] A. Motwani et al., "Soil Analysis and Crop Recommendation using Machine Learning," 2022.

[2] P. Pudumalar et al., "Crop Recommendation System for Precision Agriculture," 2016.

[3] M. R. Madhumathi et al., "Prediction of Soil Fertility using ML Algorithms and Fertilizer Recommendation System," 2023.

[4] P. Sharma et al., "Predicting Agriculture Yields Based on Machine Learning Using Regression and Deep Learning," 2023.

[5] W. Albattah et al., "A Novel Deep Learning Method for Detection and Classification of Plant Diseases," 2021.

[6] 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/GHTCSAS.2013.6629944.

[7] S. Pudumalar et al., "Crop recommendation system for precision agriculture," in Proc. 2016 Eighth Int. Conf. Adv. Comput, pp. 32-36, doi:10.1109/ICoAC.2017.7951740

[8] R. K. Rajak et al.,"Crop Recommendation System to Maximize Crop Yield using Machine Learning Technique," Int. Res. J. of Eng. and Technol., issue 12, vol. 4, pp. 950-953, Dec 2017

[9] A.K.M. Reddy, S. Chithra, H.M. Hemashree and Thanu K., "Soil Classification and Crop Suggestion using Machine Learning," Int. J. for Res. in Appl. Sci. and Eng. Technol., issue 7, vol. 8, pp. 1625-1628, Jul 2020

[10] M. S. Rao, A. Singh, N. V. S. Reddy, and D. U. Acharya, ''Crop prediction using machine learning,'' J. Phys., Conf. Ser., vol. 2161, no. 1, 2022, Art. no. 012033

[11] A. Savla et al., "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

[12] 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

[13] Priya P., Muthaiah U. and Balamurugan M., "Predicting Yield of the Crop Using Machine Learning Algorithm," Int. J. of Eng. Sci and Res. Technol., issue 11,vol. 29, pp. 1248-1255, 2020

[14] 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

[15] Abhinav Sharma, Arpit Jain, Prateek gupta, and Vinay Chowdary "Machine Learning Applications for Precision Agriculture: A Comprehensive Review", IEEE Access, vol. 9, 2020.