

Smart Paddy identification and soil suitability prediction using Deep Learning

¹Chidaraboina Haritha

¹M.Tech, Department of Computer Science and Engineering
DNR College of Engineering & Technology (Autonomous)
Bhimavaram – 534202, Andhra Pradesh, India
harithaabhi28@gmail.com

²Dr. A. Ramamurthy

²Professor, Department of Computer Science and Engineering
DNR College of Engineering & Technology (Autonomous)
Bhimavaram – 534202, Andhra Pradesh, India
ram111.sai@gmail.com

Abstract— Agriculture plays a vital role in the economy and food security of many countries, with paddy remaining one of the most important staple crops cultivated worldwide. Accurate crop detection and appropriate soil recommendation are essential for improving yield and reducing losses. Traditional farming practices rely heavily on farmer experience and manual observation, often leading to inefficient resource usage and reduced productivity. Artificial Intelligence provides advanced solutions for smart agriculture by enabling automated analysis of crop and soil conditions. This project focuses on paddy crop detection and soil suitability prediction using deep learning techniques. Image processing and machine-learning models are employed to detect paddy crop conditions, while soil parameters are analyzed to assess fertility and compatibility. Data collected from sensors and crop images is processed by AI models to generate accurate predictions, supporting early detection of crop issues and helping farmers make informed decisions. The system recommends suitable soil treatments and fertilizers, reduces manual effort through automation, and improves both crop yield and soil health. The proposed solution is scalable, cost-effective, and promotes sustainable farming practices, demonstrating the effectiveness of AI-based approaches in agriculture.

Keywords--- Paddy Crop Detection, Soil Suitability Prediction, Deep Learning, Machine Learning, Image Processing, Precision Agriculture, Fertilizer Recommendation, Smart Farming, Crop Yield Improvement, Sustainable Agriculture.

I. INTRODUCTION

Agriculture is an important component of global food production, contributing to food security, employment, economic development, and the growing demand for staple crops. Among these, paddy (rice) remains one of the most widely cultivated and consumed crops, and its productivity strongly depends on accurate crop identification and suitable soil conditions. Parameters such as soil pH, nitrogen, phosphorus, potassium (NPK), moisture, and temperature directly affect the growth, health, and yield of paddy crops. These parameters can change due to irrigation practices, fertilization habits, seasonal variation, and regional soil composition. Recent advances in artificial intelligence and deep learning have created opportunities for intelligent paddy identification and predictive soil-suitability assessment [1]–[3]. Conventional paddy monitoring generally relies on manual field observation, farmer experience, and periodic soil testing. Although these methods provide useful information, they have limited capability for continuous, large-scale, and objective analysis. Machine-learning techniques can learn relationships between crop image features and disease/health status from historical observations [9], [12]. Deep-learning and hybrid

CNN-based models can further identify complex visual patterns in paddy leaf images with high accuracy [1], [2], [4]–[6]. Soil-suitability prediction is equally important because crop yield depends heavily on matching the crop to appropriate soil and nutrient conditions. Machine-learning and deep-learning approaches have therefore been explored to recommend suitable crops and fertilizers based on soil image and nutrient data [10]–[13]. Sensor-assisted and image-based monitoring have also improved data collection for both crop and soil analysis [3], [10]. However, limited datasets, environmental variability, computational requirements, and differences between soil types across regions remain important challenges [1], [6], [10].

Prediction alone cannot provide complete intelligent agricultural management. A practical AI-based system should integrate data acquisition, preprocessing, paddy identification, soil-suitability prediction, and treatment/fertilizer recommendation. This project therefore aims to:

1. Review machine-learning and deep-learning approaches for paddy crop identification.
2. Examine soil-suitability prediction techniques using NPK, pH, moisture, and related parameters.
3. Analyze image-based and sensor-assisted data acquisition approaches.
4. Compare the strengths, limitations, and applicability of existing AI methods.
5. Examine early detection, decision support, and recommendation generation.
6. Formulate an integrated AI-based paddy identification and soil-suitability framework.
7. Identify evaluation requirements, research challenges, and future directions for reliable and sustainable paddy farming.

The remainder of this paper discusses existing detection and prediction approaches, the proposed system, data representation, evaluation requirements, challenges, and concluding observations.

II. EXISTING METHODS FOR AI-BASED PADDY IDENTIFICATION AND SOIL SUITABILITY PREDICTION

Existing approaches can be broadly classified into manual observation, sensor-based monitoring, statistical methods, machine-learning approaches, deep-learning models, and hybrid/remote-sensing frameworks.

A. Manual Field Observation

Manual monitoring depends on visual inspection of paddy leaves and periodic soil testing conducted by farmers or

agricultural experts. It can provide reliable human interpretation but becomes time-consuming and inconsistent across large fields, limiting early identification of crop issues [1], [9].

B. Sensor-Based and IoT Monitoring

Sensor-based systems continuously collect soil parameters such as moisture, temperature, and nutrient levels, providing more frequent observations than manual sampling. These systems support real-time assessment and generate datasets for predictive analysis. However, sensor noise, calibration requirements, and deployment cost can affect data reliability [3].

C. Conventional Machine Learning

Machine-learning approaches learn relationships between soil/crop parameters and outcomes from historical data. Algorithms such as Random Forest, SVM, and KNN have been used for crop recommendation and disease classification [11]–[13]. Their advantage is efficient computation, but performance depends heavily on feature selection and dataset quality.

D. Deep Learning-Based Detection

Deep-learning models, particularly CNNs, can capture complex visual patterns in paddy leaf images for disease and crop-condition detection [1], [2], [4]–[6]. These approaches provide strong classification accuracy but generally require large training datasets and computational resources.

E. Hybrid and Ensemble Models

Hybrid models combine multiple learning mechanisms — e.g., CNN feature extraction with optimization or boosting classifiers — to improve accuracy. Recent studies have investigated hybrid deep-learning frameworks for paddy disease classification and multimodal soil-crop recommendation [1], [10]. Such models improve reliability but increase model complexity.

F. Remote Sensing and Image-Based Soil Classification

Image-based soil classification using CNN architectures (ResNet, EfficientNet, Vision Transformers) provides an additional source of soil-type information beyond tabular sensor data [10], [16]. Although this improves prediction accuracy, challenges include image data diversity, computational cost, and regional soil variation.

TABLE I: COMPARISON OF EXISTING PADDY IDENTIFICATION AND SOIL SUITABILITY METHODS

Method	Main Technique	Strength	Major Limitation
Manual observation	Human inspection	Direct assessment	Time-consuming, inconsistent
Sensor-based	Continuous sensing	Frequent measurements	Noise, calibration cost
Machine learning	RF, SVM, KNN	Automated prediction	Data dependent
Deep learning	CNN	High visual accuracy	Requires large datasets
Hybrid models	CNN + optimization/boosting	Improved accuracy	Model complexity
Image-based soil classification	ResNet, EfficientNet, ViT	Rich soil-type information	Data diversity, compute cost
Proposed framework	Integrated AI	Detection + suitability + recommendation	Requires comprehensive validation

The literature indicates a transition from manual, isolated methods toward continuous, predictive, and integrated AI-based paddy and soil management. However, most existing approaches address paddy detection and soil recommendation as separate problems. A comprehensive architecture should integrate crop-image data, soil data, prediction, and recommendation into a unified decision-support pipeline.

III. PROPOSED AI-BASED SMART PADDY

IDENTIFICATION AND SOIL SUITABILITY SYSTEM

The proposed system considers paddy crop and soil management as an integrated sequence of interconnected processing stages:

Paddy Field Sensors + Crop Images

Data Acquisition

Data Preprocessing

Crop and Soil Feature Analysis

AI-Based Prediction

Paddy Crop Identification / Abnormality Detection

Soil Suitability Assessment

Fertilizer and Treatment Recommendation

Farmer Monitoring and Intervention

The architecture consists of five major functional layers: data acquisition, preprocessing, intelligent analysis, prediction, and decision support. The first layer collects paddy crop images and soil parameters such as pH, nitrogen, phosphorus, potassium, moisture, and temperature. The preprocessing layer removes noise, handles missing values, and converts heterogeneous image and sensor data into suitable input features.

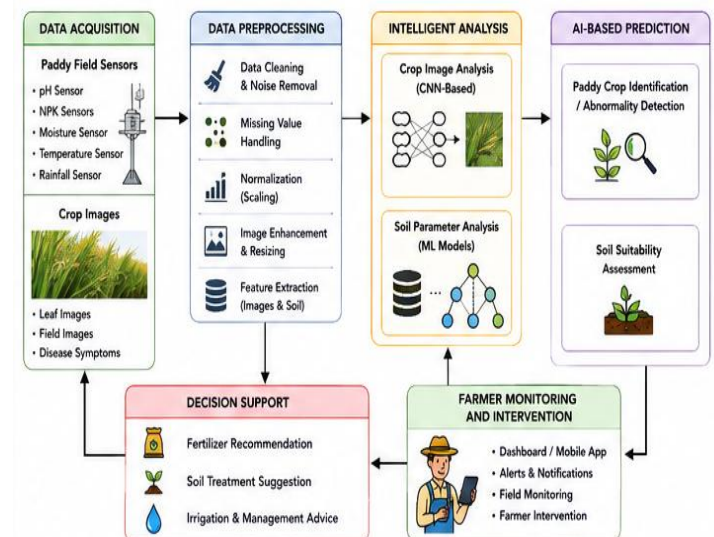


Fig. 1. Overall Architecture of the Proposed AI-Based Smart Paddy Identification and Soil Suitability System

The intelligent analysis layer applies CNN-based deep-learning techniques to identify paddy crop conditions from images, while a parallel module analyzes soil nutrient data. The prediction layer estimates crop health/disease status and soil-crop suitability based on the combined data. The decision-support layer presents predictions and recommended fertilizer/treatment actions to farmers through an accessible interface, assisting rather than replacing farmer judgement. This integrated approach enables early detection, informed decision-making, and more efficient paddy farming.

IV. PADDY AND SOIL DATA REPRESENTATION

A. Data Acquisition and Processing

The proposed system collects paddy crop images along with soil parameters such as pH, nitrogen (N), phosphorus (P), potassium (K), moisture, and temperature. Preprocessing handles missing, duplicate, inconsistent, and abnormal data, and normalizes numerical values.

The structured representation is defined as:

$$P_i = \{C_i, S_i, E_i, H_i\} \quad (1)$$

where C_i represents crop-image features, S_i represents soil parameters, E_i represents environmental conditions, and H_i represents crop-health indicators.

Crop, soil, and health indicators are analyzed jointly to assess paddy field conditions. Normalization converts parameters with different scales and units into comparable ranges, improving joint model learning and prediction.

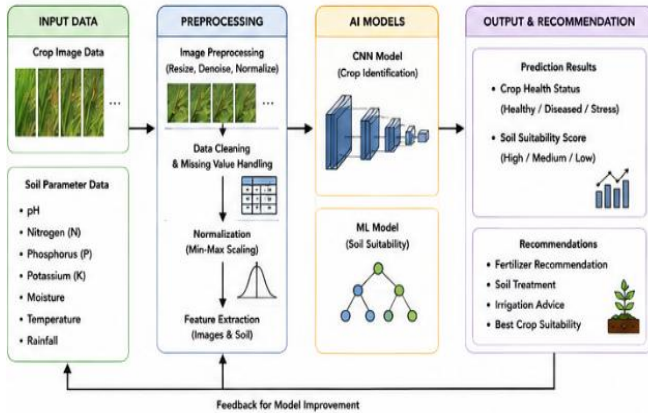


Fig. 2. Paddy Crop and Soil Data Processing and Recommendation Generation

B. Paddy Field Condition Representation

The condition of a paddy field is represented using the relevant crop and soil parameters. The system extracts:

- Crop leaf image features
- Soil pH level
- Nitrogen (N) level
- Phosphorus (P) level
- Potassium (K) level
- Soil moisture
- Temperature
- Rainfall
- Crop health indicators

The paddy condition representation is defined as:

$$Q = \{C, pH, N, P, K, M, T, Rf, H\} \quad (2)$$

where C represents crop-image features, pH represents soil acidity/alkalinity, N , P , K represent nutrient levels, M represents moisture, T represents temperature, Rf represents rainfall, and H represents crop-health indicators.

Representing the collected parameters using a unified structure enables the proposed system to analyze the relationship between crop appearance, soil properties, and environmental conditions. This representation is subsequently provided to the deep-learning model for paddy crop identification, disease detection, or soil-suitability prediction, depending on the objective of the system.

V. SEMANTIC PADDY CONDITION-RECOMMENDATION MATCHING AND FIELD SUITABILITY SCORING

A central component of the proposed system is semantic condition-recommendation matching. Unlike simple threshold-based analysis, the proposed approach identifies relationships between the observed paddy field conditions and the recommended soil-management requirements. The system considers parameters such as soil pH, nitrogen, phosphorus, potassium, moisture, crop-image features, and other available field parameters to determine the suitability of the current farming environment for paddy cultivation.

The observed field condition and the recommended management condition are represented as feature embeddings, and their semantic similarity is computed to assess how closely the current field state matches an ideal or recommended paddy-growing condition. A higher similarity value indicates stronger compatibility between the observed field conditions and the recommended soil-management practices.

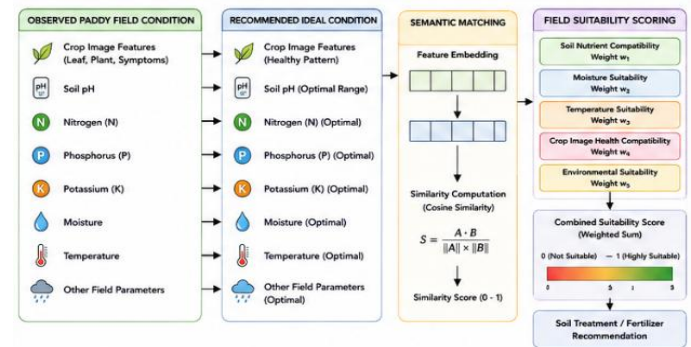


Fig. 3. Semantic Paddy Condition-Recommendation Matching and Field Suitability Scoring

Semantic similarity alone, however, should not determine the final recommendation. A field condition with high overall similarity may still contain a critical deficiency, such as insufficient nitrogen or unsuitable pH, that can negatively affect paddy growth. Therefore, the proposed system combines semantic similarity with explicit paddy-specific factors, including soil-nutrient compatibility, moisture suitability, temperature suitability, crop-image-based health compatibility, and overall environmental suitability. Each of these factors is assigned a parameter-specific importance, allowing the system to give greater weight to whichever conditions matter most for a given objective — for example, a crop-health-monitoring

application may emphasize crop-image features and disease indicators, whereas a soil-recommendation application may emphasize nutrient and pH compatibility.

The field conditions are subsequently evaluated according to their combined suitability score to generate a field-suitability status and corresponding soil-treatment or fertilizer recommendation, enabling timely intervention and improved paddy field management.

VI. EXPLAINABLE AI AND DECISION-AWARE PADDY MANAGEMENT

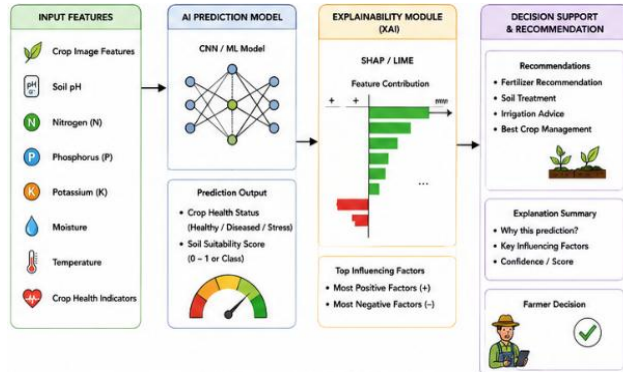


Fig. 4. Explainable AI Framework for Paddy Identification and Soil Suitability Prediction

AI-based paddy management systems should provide more than a prediction or numerical output. Farmers need to understand why the system generated a particular prediction or recommendation. Explainable AI (XAI) techniques such as SHAP and LIME can provide feature-level explanations of model predictions.

For each prediction, the proposed system can generate an explanation containing:

- Crop-image feature contribution
- Soil pH influence
- Nitrogen, phosphorus, and potassium contribution
- Moisture and temperature influence
- Crop health/disease indicators
- Major factors contributing to the final prediction
- Recommended soil treatment or fertilizer action

For example:

Paddy Field Condition — 91% Suitable

- Crop condition: healthy.
- Soil pH: suitable.
- Nitrogen level: adequate.
- Phosphorus level: adequate.
- Potassium level: adequate.
- Moisture: within optimal range.
- Major influencing factor: balanced soil-nutrient composition.
- Recommendation: Maintain current fertilization schedule and continue regular monitoring.

Such explanations help farmers understand whether the AI prediction is supported by meaningful crop and soil parameters rather than relying only on the final prediction value.

A. Explainability and Decision Monitoring

Paddy prediction models can be influenced by multiple crop-image and soil-nutrient parameters. Therefore, the system should continuously monitor the contribution of each input feature to the prediction. Explainability helps identify which parameters have the greatest influence on crop health, soil suitability, or other target outcomes considered by the proposed system.

The proposed framework therefore integrates explainability as a decision-support component. The system should report both: *Performance* — How accurately does the system predict paddy crop condition and soil suitability?

Explainability — Can the farmer understand which crop and soil parameters caused the prediction?

This dual objective is important because a highly accurate model that provides no understandable reasoning may be difficult to trust or use in practical farming environments.

VII. ADAPTIVE PADDY INTELLIGENCE AND HUMAN-IN-THE-LOOP DECISION SUPPORT

Paddy field conditions are not static. Soil nutrient levels, moisture, temperature, and crop growth stages continuously change during the cultivation cycle. A prediction model trained using fixed field conditions may therefore become less effective when field conditions change across seasons or regions.

Model Update — Adaptive mechanisms can update soil-parameter ranges, nutrient thresholds, feature importance, prediction knowledge, and fertilizer recommendations while maintaining model version control. The system can periodically incorporate newly collected field data to improve its understanding of changing paddy conditions. However, adaptive AI should not automatically modify critical fertilizer or treatment recommendations without validation — model updates should first be evaluated using historical and newly collected datasets before being deployed.

Human-in-the-Loop Workflow — The final decision-support workflow is:

Data Collection → AI Prediction → Condition Assessment → Explanation → Farmer Verification → Recommendation → Final Farmer Decision

The farmer or agricultural expert remains responsible for the final decision, particularly when field conditions are unusual or when additional local knowledge is required that may not be available in the collected data. This human-in-the-loop design allows farmers to combine AI-based predictions with practical farming experience and real-time field observations. Farmer feedback can also be captured and used for future system improvement, while preventing the AI system from making completely autonomous decisions without human supervision.

VIII. EVALUATION FRAMEWORK

A comprehensive evaluation framework is adopted to assess the proposed paddy identification and soil-suitability system across detection accuracy, prediction reliability, explainability, efficiency, and robustness. Rather than relying on a single performance measure, the evaluation considers multiple complementary dimensions to determine the practical effectiveness of the proposed framework.

A. Paddy Detection Performance

The paddy identification module is evaluated using accuracy, precision, recall, and F1-score. Accuracy measures overall

classification correctness, precision indicates the proportion of correctly identified healthy/diseased crops, recall measures the ability to correctly detect all affected crops, and F1-score provides a balanced assessment of precision and recall.

B. Soil Suitability Prediction Performance

Since soil-suitability prediction is a core objective, the system is evaluated using regression and classification metrics such as RMSE, MAE, and suitability-classification accuracy, determining how closely predicted soil-suitability scores match actual field outcomes.

C. Recommendation Relevance

The recommendation component is compared with conventional threshold-based fertilizer guidance. The evaluation examines whether the AI-generated recommendations align with agronomic best practices across varying soil and crop conditions.

D. Explainability

The generated explanations are evaluated based on faithfulness, consistency, completeness, and human understandability, ensuring farmers can verify which crop and soil factors influenced a given prediction.

E. Efficiency and Robustness

Computational evaluation considers image-processing time, prediction latency, memory consumption, and scalability. Robustness is evaluated through testing across different paddy varieties, soil types, regional conditions, and image-quality variations, ensuring the system maintains reliable performance under practical farming conditions.

TABLE II: EVALUATION DIMENSIONS OF THE PROPOSED PADDY IDENTIFICATION AND SOIL SUITABILITY FRAMEWORK

Dimension	Evaluation Metrics / Criteria	Purpose
Paddy Detection	Accuracy, Precision, Recall, F1-Score	Evaluate crop identification performance
Soil Suitability	RMSE, MAE, Classification Accuracy	Evaluate suitability-prediction quality
Recommendation Relevance	Comparison with agronomic baseline	Assess recommendation quality
Explainability	Faithfulness, Consistency, Completeness, Human Understandability	Evaluate decision transparency
Efficiency	Processing Time, Prediction Latency, Memory	Assess computational efficiency
Robustness	Cross-variety Testing, Soil-type Variation, Image-quality Variation	Evaluate generalization capability

IX. RESEARCH CHALLENGES AND FUTURE ROADMAP

Although AI-powered paddy identification and soil-suitability prediction can improve farming decisions, several challenges remain in developing a reliable and practical system.

A. Dataset Availability — Paddy and soil datasets may contain limited regional diversity, incomplete nutrient profiles, and variations in image quality. Developing sufficiently diverse and representative datasets is important for reliable prediction.

B. Crop–Soil Variation — Paddy varieties and soil types differ across regions and growing conditions. The proposed system must therefore handle variations in crop appearance, soil composition, and environmental factors to achieve reliable predictions.

C. Prediction Reliability — Soil-suitability scoring should consider multiple field-relevant factors rather than depending on a single parameter. Future improvements should focus on maintaining consistent predictions across different regions and seasons.

D. Explainability — AI-generated recommendations should provide understandable reasons. Future systems should improve the transparency of nutrient analysis, crop-health assessment, and fertilizer-recommendation reasoning.

E. Data Quality and Noise — Field images and sensor data may contain noise, occlusion, or inconsistent lighting/environmental conditions. Robust preprocessing and noise-handling techniques are required for reliable predictions.

F. Scalability — New paddy varieties, soil conditions, and farming practices continuously emerge. Future systems should support updated models and evolving field conditions without reducing previously learned capabilities.

G. Human–AI Interaction — AI recommendations should assist farmers rather than completely replace their judgement. Future systems should provide effective farmer-feedback mechanisms and accessible decision-support interfaces.

H. Cost and Accessibility — Practical deployment requires low-cost, scalable solutions suitable for small and marginal farmers. Affordable sensor and imaging infrastructure is therefore important for real-world adoption.

The research roadmap can be organized into three stages:

Short Term: Improve paddy image preprocessing, soil-parameter analysis, multi-factor suitability scoring, and explainability.

Medium Term: Develop adaptive models for diverse paddy varieties, improved region-specific recommendations, diverse agricultural datasets, and robust cross-region evaluation.

Long Term: Develop reliable, human-centered smart farming intelligence that combines paddy identification, explainable soil-suitability prediction, adaptive learning, and human-in-the-loop decision support at scale.

X. CONCLUSION

This paper presented an AI-powered framework for intelligent paddy crop identification and soil-suitability prediction. The framework addresses the limitations of conventional manual farming practices by incorporating image-based crop detection, soil-parameter analysis, semantic condition–recommendation matching, multi-factor suitability scoring, explainability, and human-in-the-loop decision support.

The proposed approach evaluates paddy field conditions using crop-image features, soil pH, nitrogen, phosphorus, potassium,

moisture, and other relevant agronomic parameters. Semantic matching enables the system to identify contextual relationships between observed field conditions and recommended management practices beyond simple threshold-based rules. The suitability-scoring component further prioritizes recommendations according to their overall relevance to the specific field condition.

Explainability is incorporated to improve the transparency and reliability of the system's recommendations. Farmers can examine the major factors contributing to a given prediction, improving trust and adoption. The human-in-the-loop design ensures that AI-generated recommendations support farmer decision-making rather than completely replacing human judgement.

Future research should focus on diverse regional datasets, adaptive crop-and-soil representations, robust suitability evaluation, faithful explainability, and reliable, cost-effective AI-based decision support for smallholder paddy farmers.

REFERENCES

- [1] C. Subbarayudu and M. Kubendiran, "An automated hybrid deep learning framework for paddy leaf disease identification and classification," *Scientific Reports*, vol. 15, Art. no. 26873, 2025, doi: 10.1038/s41598-025-08071-6. (Open access — PDF: nature.com/articles/s41598-025-08071-6.pdf)
- [2] G. Latif, S. E. Abdelhamid, R. E. Mallouhy, J. Alghazo, and Z. A. Kazimi, "Deep learning utilization in agriculture: Detection of rice plant diseases using an improved CNN model," *Plants*, vol. 11, no. 17, Art. no. 2230, 2022, doi: 10.3390/plants11172230. (Open access on PMC)
- [3] O. Debnath and H. N. Saha, "An IoT-based intelligent farming using CNN for early disease detection in rice paddy," *Microprocessors and Microsystems*, vol. 94, Art. no. 104631, 2022.
- [4] A. Haridasan, J. Thomas, and E. D. Raj, "Deep learning system for paddy plant disease detection and classification," *Environmental Monitoring and Assessment*, vol. 195, no. 1, Art. no. 120, 2023, doi: 10.1007/s10661-022-10656-x.
- [5] A. Singh, J. Kaur, K. Singh, and M. L. Singh, "Deep transfer learning-based automated detection of blast disease in paddy crop," *Signal, Image and Video Processing*, vol. 18, no. 1, pp. 569–577, 2024, doi: 10.1007/s11760-023-02735-4.
- [6] R. K. Dubey and D. K. Choubey, "Adaptive feature selection with deep learning MBi-LSTM model-based paddy plant leaf disease classification," *Multimedia Tools and Applications*, vol. 83, no. 9, pp. 25543–25571, 2024, doi: 10.1007/s11042-023-16475-7.
- [7] S. Lamba, A. Baliyan, and V. Kukreja, "A novel GCL hybrid classification model for paddy diseases," *International Journal of Information Technology*, vol. 15, no. 2, pp. 1127–1136, 2023.
- [8] R. Dogra et al., "Deep learning model for detection of brown spot rice leaf disease with smart agriculture," *Computers and Electrical Engineering*, vol. 109, Art. no. 108659, 2023.
- [9] P. K. Sethy, N. K. Barpanda, A. K. Rath, and S. K. Behera, "Deep feature-based rice leaf disease identification using support vector machine," *Computers and Electronics in Agriculture*, vol. 175, Art. no. 105527, 2020.
- [10] V. Pandey, R. Das, and D. Biswas, "AgroSense: An integrated deep learning system for crop recommendation via soil image analysis and nutrient profiling," *arXiv preprint arXiv:2509.01344*, 2025. (Freely downloadable: arxiv.org/pdf/2509.01344)
- [11] A. Motwani, P. Patil, V. Nagaria, S. Verma, and S. Ghane, "Soil analysis and crop recommendation using machine learning," in *Proc. IEEE Conf.*, 2022, doi via IEEE Xplore document 9725901.
- [12] E. Manjula and S. Djodiltachoumy, "Efficient prediction of recommended crop variety through soil nutrients using deep learning algorithm," *Journal of Postharvest Technology*, vol. 10, no. 2, pp. 66–80, 2022.
- [13] J. Madhuri and M. Indiramma, "Artificial neural networks-based integrated crop recommendation system using soil and climatic parameters," *Indian Journal of Science and Technology*, vol. 14, no. 19, pp. 1587–1597, 2021.
- [14] Z. Doshi, S. Nadkarni, R. Agrawal, and N. Shah, "AgroConsultant: Intelligent crop recommendation system using machine learning algorithms," in *Proc. 4th Int. Conf. Computing Communication Control and Automation (ICCUBEA)*, IEEE, 2018.
- [15] N. H. Kulkarni et al., "Improving crop productivity through a crop recommendation system using ensembling technique," in *Proc. 3rd Int. Conf. Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, IEEE, Bangalore, India, 2018.
- [16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.