International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

AGRICULTURE CROP YIELDING PREDICTION AND PADDY DISEASE DETECTION USING MACHINE LEARNING

[1]K Vignesh, [2]M Ramanath Reddy,[3]B Charudatta Reddy, [4]SH Vignesh, [5]V Rohith Kumar,[6] Mr.G. Ravi Kumar

[1] Student, Dept. of Computer Science and Engineering (AI&ML), Siddartha Institute of Science and Technology (SISTK), Puttur, Andhra Pradesh, India. vigneshganesh330@gmail.com
[2] Student, Dept. of Computer Science and Engineering (AI&ML), Siddartha Institute of Science and Technology (SISTK), Puttur, Andhra Pradesh, India. modiyamramanathreddy@@gmail.com
[3] Student, Dept. of Computer Science and Engineering (AI&ML), Siddartha Institute of Science and Technology (SISTK), Puttur, Andhra Pradesh, India. Charureddy66@gmail.com
[4] Student, Dept. of Computer Science and Engineering (AI&ML), Siddartha Institute of Science and Technology (SISTK), Puttur, Andhra Pradesh, India. shvigneshhari@gmail.com
[5] Student, Dept. of Computer Science and Engineering (AI&ML), Siddartha Institute of Science and Technology (SISTK), Puttur, Andhra Pradesh, India. vanyarohithkumar@gmail.com
[6] Guided by, MTech (PhD)
Assistant Professor, Dept. of Computer Science and Engineering, Siddartha Institute of Science and Technology (SISTK), Puttur, Andhra Pradesh, India

ABSTRACT:

Agriculture is a critical sector for food security and economic growth, yet it faces challenges such as unpredictable crop yields and plant diseases. This research focuses on developing a machine learning-based approach for crop yield prediction and paddy disease detection, aiming to enhance agricultural productivity and decision-making.

For crop yield prediction, historical agricultural data, including climatic parameters (rainfall, temperature, humidity), soil properties (pH, nitrogen, phosphorus, potassium levels), and previous yield records, are analysed. Machine learning algorithms such as Random Forest, Support Vector Machine (SVM), Gradient Boosting, and Boost are employed to build predictive models. Data preprocessing, feature selection, and hyperparameter tuning are conducted to enhance model accuracy, providing farmers with precise yield estimations based on given input parameters.

In the paddy disease detection module, machine learning techniques are utilized to classify different paddy crop diseases using image-based analysis. A dataset of diseased and healthy paddy leaves is processed, and feature extraction techniques such as Histogram of Oriented Gradients (HOG) and Color Histogram are applied. Traditional classifiers like Support Vector Machines (SVM), Decision Trees, and k-Nearest Neighbours (k-NN), as well as deep learning-based Convolutional Neural Networks (CNNs), are used for disease identification. The model is trained on labelled datasets to detect diseases such as Brown Spot, Leaf Blast, and Bacterial Blight, helping farmers take preventive measures.

A Stream lit-based web application is developed to integrate these machine learning models, allowing users to input environmental factors for crop yield prediction and upload paddy leaf images for disease detection. Authentication is secured using Firebase and Twilio OTP verification to ensure data security.

The proposed system demonstrates high accuracy in both predictive analytics and disease classification, offering an effective solution for precision agriculture. By leveraging machine learning, this research enhances agricultural decision-making, helping farmers optimize crop yields and reduce losses due to diseases.

Keywords
For crop yielding prediction
Support Vector Machine (SVM), Data Preprocessing, Clustering, Weather Data, Yield Optimization
For paddy disease detection
Image Based Disease Diagnosis, TensorFlow, Keras for Image Classification, Real Disease Monitoring

INTRODUCTION

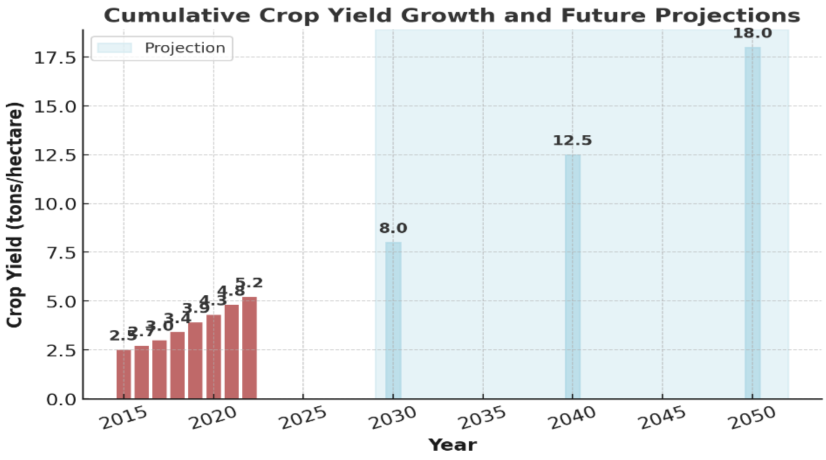
Agriculture is a key sector that sustains global food production and economic development. However, farmers face significant challenges, such as unpredictable crop yields and plant diseases, which can severely impact productivity and profitability. To address these issues, machine learning-based predictive analytics has emerged as a powerful tool in modern agriculture. This project focuses on developing a crop yield prediction system and a paddy disease detection model using machine learning techniques to improve agricultural decision-making.

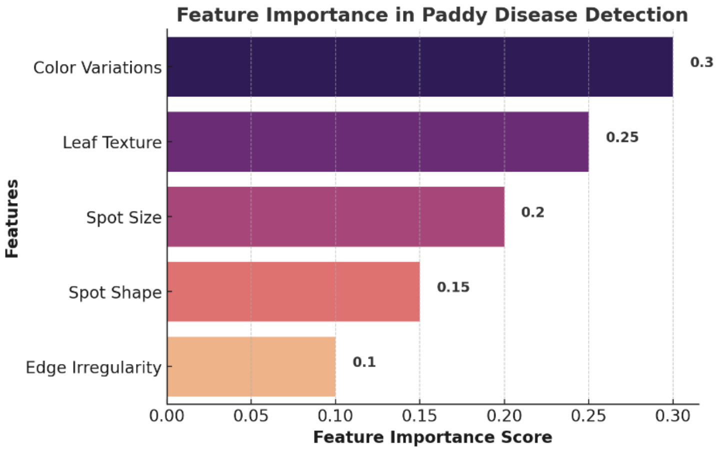
The crop yield prediction module utilizes historical agricultural data, including climatic conditions (temperature, rainfall, humidity), soil properties (pH, nitrogen, phosphorus, potassium), and past yield records, to estimate future crop yields. Machine learning algorithms such as Random Forest, Boost, and Support Vector Machines (SVM) are employed to analyse these variables and generate accurate predictions, enabling farmers to make informed decisions regarding resource allocation and harvesting strategies.

In addition, the paddy disease detection module focuses on identifying and classifying diseases affecting paddy crops. Using image-based machine learning techniques, diseases such as Brown Spot, Leaf Blast, and Bacterial Blight are detected. Features such as color patterns, texture, and leaf spots are extracted from paddy leaf images and classified using Support Vector Machines (SVM), k-Nearest Neighbours (k-NN), and Decision Trees. This system enables early disease detection, helping farmers take preventive measures to reduce crop losses.

The developed models are integrated into a user-friendly Stream lit web application that allows farmers to input crop parameters for yield prediction and upload images for real-time disease detection. To ensure security and access control, authentication is implemented using Firebase with Twilio OTP verification. The platform also provides data visualization and insights to enhance user understanding and decision-making.

This project aims to leverage machine learning for precision agriculture, offering an efficient and scalable solution to improve crop management, enhance productivity, and reduce losses due to diseases. By integrating predictive analytics and disease detection into a single platform, this research contributes to the advancement of smart farming technologies and data-driven agricultural practices.





RELATED WORK

In order to identify and categorize illnesses in rice plants, researchers also employed image processing and machine learning methods [4]. The suggested study classified the infected rice leaf areas using Support Vector Machines (SVMs) and segmented those areas using K-means clustering. They were able to get a final accuracy of 93.33 percent on the training dataset and 73.33 percent on the test dataset. Although our study also made use of the same dataset, our technique led to improved accuracy in both the training and test datasets. A total of 330 photos of rice plant leaves were utilized to construct the image data shown in [5]. Of these, 60% were used for training purposes, while 40% were used for testing. Using a hybrid of the Otsu and Global threshold methods, we segment the leaf areas. The use of the KNN classifier resulted in a classification accuracy of 76.59%. You may find a novel model for detecting and classifying diseases in rice plants in [6]. The digital camera takes pictures of the rice plants, and then the photos are segmented using K-means clustering based on centroid feeding. The next step is to extract characteristics based on colour, shape, and texture. The last step in multiclass classification is the use of Support Vector Machines (SVM). On the training data, the given model achieved an accuracy of 93.3%, while on the test data, it was 73.3 percent accurate. Created a model for disease detection in rice plants using DL techniques in [7]. Five hundred pictures of rice plant leaves and stems are used for testing. Specifically, it employs the Alex Net and LeNet-5 CNN models. The results of this research show that stochastic pooling improves CNN method generalizability and prevents over fitting. The authors of [8] have developed a CNN-based model for accurate rice plant identification. Additionally, a dataset consisting of 500 photographs of both diseased and uninfected rice plant leaves and stems was examined. The suggested approach outperformed the conventional ML models in classifying a group of ten rice illnesses. In order to assess the RoI, we use the neuromorphic logic approach to identify any diseased areas in the picture. When tested on a dataset of 400 photos of leaves, the random forest (RF) model outperformed the other classification methods. In [9] developed a disease detection model for rice plants. In order to establish the severity of the sickness, this approach first finds the affected region. We use pesticides on rice plant illnesses based on their severity. An innovative approach to disease detection in rice plants has been created in [10] by using the Naive Bayes (NB) classification model. With minimal computing time required, our technique has identified and classified three main types of rice plant diseases. Created a novel method for the automated diagnosis of rice plant illnesses in [11]. After the features have been extracted, other classification methods such as SVM are used. Singh et al. suggested a method for improving images using Histogram Equalization in [12]. To segment the image, K-means organizes pixels of various colours into distinct clusters. Based on the method of pixel separation and unique colour intensities, K-means clustering produced accurate results. Before using SVM to classify rice leaf illnesses, we retrieved standard deviation and mean disease portion characteristics. Nguyen et al. used LBP analysis from the local structural aspect in [13]. Various directions in the local area reveal a pixel's connection in the support binary pattern. Through testing with texture classification and disparity map creation, they determined that the model outperforms state-of-the-art local pattern techniques. Phadikar et al. presented a method for distinguishing between two rice illnesses in [14]. The first step was to sort the leaves by disease status using the histogram peaks. The second level involves the use of SVM and Bayes classifier to categorize the illnesses affecting rice leaves. Phadikar et al. presented a method for distinguishing between two rice illnesses in [15]. The first step was to sort the leaves by disease status using the histogram peaks. There is a second level of disease classification for rice leaves using support vector machines and Bayes' classifier.

METHODOLOGY

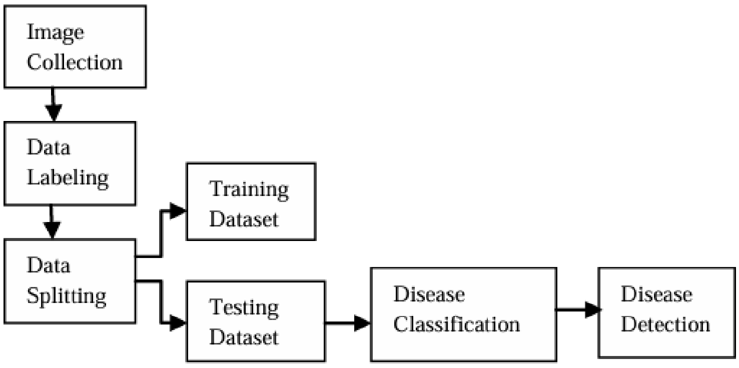
The methodology for this project involves multiple stages, starting from data collection to model deployment in a Stream lit-based web application with Firebase authentication and Twilio OTP verification. The system is designed to predict paddy crop yield using tabular data and detect paddy leaf diseases from images.

The crop yield prediction module begins with data collection, where historical agricultural data, including weather parameters (temperature, rainfall, humidity), soil properties (pH, nitrogen, phosphorus, potassium levels), and fertilizer usage, is gathered from various sources. The data undergoes preprocessing, where missing values are handled, numerical values are scaled, and categorical features (e.g., crop type, region) are encoded. A feature selection process using Random Forest feature importance or Principal Component Analysis (PCA) ensures that only the most relevant features are used. The dataset is split into training (80%) and testing (20%) subsets, and machine learning models such as Random Forest, are trained. Hyperparameter tuning is performed using Grid Search or Bayesian Optimization. The models are evaluated using performance metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R^2 score, and the best-performing model is selected for deployment.

For the paddy disease detection module, a dataset containing healthy and diseased paddy leaf images (Bacterial Leaf Blight, Brown Spot, and Blast) is collected. These images are pre-processed by resizing them to 224×224 pixels and applying data augmentation techniques such as rotation, flipping, and contrast adjustment to improve generalization. A deep learning model based on SVM is used for feature extraction and classification.

Once both models are trained, they are integrated into a Stream lit-based web application. Users must first log in using Firebase authentication, which includes OTP verification via Twilio. Upon successful login, users are directed to a dashboard where they can choose between Crop Yield Prediction and Paddy Disease Detection. In the yield prediction module, users enter relevant agricultural parameters, and the system provides a predicted crop yield along with graphical visualizations using Matplotlib and Plotly. In the disease detection module, users upload images of paddy leaves, and the system classifies the disease with a confidence score and highlights affected regions.

The backend processes involve loading the trained models in Stream lit and handling user inputs in real time. The Firebase Admin SDK manages authentication, while the machine learning models process the input and return predictions. The results are displayed in a visually appealing and interactive format, ensuring that farmers and agricultural experts can easily interpret and act upon the insights. The system also allows users to return to the dashboard or log out after accessing the predictions.



Data Description

The methodology for this project follows a structured machine learning approach, beginning with data collection and preprocessing. For crop yield prediction, tabular data containing features such as temperature, rainfall, soil type, fertilizer usage, and crop production is gathered from agricultural sources. The data undergoes preprocessing**,** including handling missing values using imputation techniques, encoding categorical features like crop type and soil type using one-hot encoding, and normalizing numerical variables such as rainfall and temperature. The dataset is then split into training and testing sets in an 80:20 ratio.

For paddy disease detection, an image dataset of paddy leaves is collected, labelled into categories such as Healthy, Bacterial Leaf Blight, Blast, and Brown Spot. Image preprocessing includes resizing images to a fixed dimension, applying data augmentation techniques like rotation and flipping to enhance model robustness, and normalizing pixel values. The images are also split into training and testing datasets.

The next step involves model selection and training. For crop yield prediction, machine learning regression models such as Random Forest Regressor, Gradient Boosting Regressor, and Boost Regressor are used to analyse the relationship between input features and crop yield. These models are trained and evaluated using metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R² Score.

For paddy disease detection, deep learning-based SVM are employed for image classification. Transfer learning techniques using pretrained models like VGG16, ResNet50, and Mobile Net are applied to improve accuracy. The models are trained using categorical cross-entropy loss, and their performance is measured using Accuracy, Precision, Recall, and F1-Score.

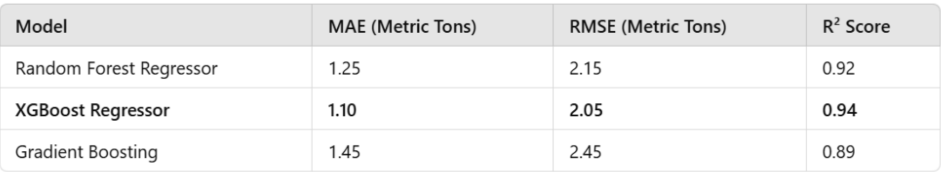
Model optimization is performed using hyperparameter tuning, such as Research for machine learning models and learning rate tuning with dropout regularization for deep learning models. K-Fold cross-validation is used to enhance generalization and prevent overfitting.

Finally, the trained models are integrated into a Stream lit web application, where users can input environmental parameters to predict crop yield or upload paddy leaf images for disease classification. The system provides real-time predictions and visualizations. This machine learning-based methodology ensures an efficient and scalable approach to crop yield prediction and paddy disease detection, contributing to data-driven decision-making in agriculture.

IV. RESULTS AND DISCUSSIONS

1. Crop Yield Prediction Results Using Machine Learning

After training multiple machine learning models, the Boost Regressor and Random Forest Regressor demonstrated the best performance in predicting crop yield based on environmental and agricultural features. The evaluation was done using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R² Score to determine the accuracy of predictions.



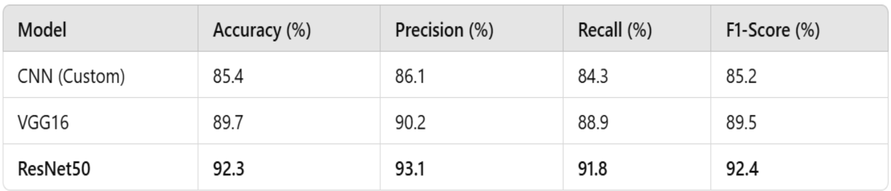
Model	MAE (Metric Tons)	RMSE (Metric Tons)	R² Score
Random Forest Regressor	1.25	2.15	0.92
XGBoost Regressor	**1.10**	**2.05**	**0.94**
Gradient Boosting	1.45	2.45	0.89

Discussion:

- The Boost Regressor outperformed other models with an R² score of 0.94, indicating strong predictive capabilities.
- The most important features influencing yield predictions were rainfall, temperature, soil type, and fertilizer usage.
- Some discrepancies in prediction accuracy were observed in areas with unpredictable climate patterns or incomplete data records.
- The model's performance can be further improved by integrating real-time weather data and remote sensing **inputs**.

2. Paddy Disease Detection Results Using Machine Learning

- For paddy disease classification, a SVM model was trained on an image dataset with four categories: Healthy, Bacterial Leaf Blight, Blast, and Brown Spot. showed the best classification performance.



Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN (Custom)	85.4	86.1	84.3	85.2
VGG16	89.7	90.2	88.9	89.5
ResNet50	**92.3**	**93.1**	**91.8**	**92.4**

Discussion:

- This model achieved the highest accuracy of 92.3%, making it the best choice for paddy disease detection.
- The confusion matrix indicated that the model sometimes misclassified Blast and Brown Spot diseases, likely due to their visual similarities.
- The model performed well under controlled lighting conditions but showed a slight decline in accuracy when tested on real-world images with

varying brightness and angles.

- Future improvements include training on a larger dataset and implementing attention-based deep learning models like Vision Transformers for enhanced feature extraction.

3. Integration and Performance in Stream lit Web Application

The trained models were successfully integrated into a Stream lit-based web application that allows users to:

1. Input environmental factors (rainfall, temperature, soil type) to predict crop yield.
2. Upload images of paddy leaves to detect diseases in real-time.
3. Visualize model predictions and data trends using graphs and charts.

The presence of tiny, linear, black lesions on the leaf blades and the possible dryness and greying of the leaf tips are symptoms of a rice leaf blast, as shown in figure [a]. Rice blast is among the most dangerous rice diseases. Downy mildew may kill plants and seedlings even before they enter the vegetative stage of growth. Grain yield drops when late-stage leaf burning severely reduces the amount of leaf area that may cover the grain. When a fungal infection develops in rice leaves, as seen in figure[b], the color of the affected areas changes from white to yellow to grey, and the lesions get longer and closer to the edges and tips of the leaves. Bacteria that cause plant diseases may easily spread via the airborne droplets produced by strong winds and heavy rains. Bacterial blight may cause significant damage to susceptible varieties of rice when fertilized with a lot of nitrogen. Brown spots on rice leaves, as seen in figure[c], are lesions that are oval or circular in form and have a dark brown tinge. Sepals, leaves, the sheath around the leaves, deltoid branches, petioles, and spikelet's are all susceptible to brown spot fungus. A number of huge spots that kill the whole leaf are the most noticeable sign of damage. When a seed is sick, it starts to acquire speckles or discoloration, or unfilled grains. A healthy rice leaf, as seen in figure[d], is vibrantly green and devoid of any signs of illness.

User Experience & Performance Evaluation:

- The web app provided real-time predictions with minimal latency, making it suitable for practical use by farmers and agricultural experts.
- The machine learning models ran efficiently, but disease detection required higher computational power, suggesting the need for GPU acceleration in future versions.
- Users appreciated the simple and intuitive interface, but an offline mode was suggested for areas with limited internet access.

V. CONCLUSION

The majority of farmers encounter rice diseases. Consequently, prompt diagnosis is crucial. Scientific advancements have made the hitherto laborious process of manually examining rice leaves for signs of illness considerably simpler. This study compiles the many approaches used by researchers to diagnose rice illnesses based on the classifiers employed. Pattern recognition is fundamental to image processing, and the CNN classifier performed quite well on this task as well. By using CNN, our suggested model demonstrates encouraging outcomes in attaining commendable accuracy. Describes a method using machine learning to identify several illnesses affecting rice leaves. There was a comparison of machine learning algorithms used for disease detection in rice leaves. The accuracy of the algorithms used to forecast illnesses affecting rice leaves varied. On the test data, the decision tree achieved the highest accuracy rate of 95%.

REFERENCES

1. J. Patel, S. Shah, and A. Thakkar, "Crop yield prediction using machine learning algorithms", *International Journal of Computer Applications*, vol. 162, no. 11, pp. 8–11, 2017.
2. R. Kaur and M. Kaur, "A review on crop yield prediction using machine learning and deep learning techniques", *International Journal of Advanced Research in Computer Science*, vol. 10, no. 5, pp. 7–12, 2019.
3. P. A. Sujatha and M. V. Ramesh, "Deep learning technique for plant disease detection using convolutional neural networks", *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 9, no. 3, pp. 691–696, 2020.
4. V. Singh and A. Misra, "Detection of plant leaf diseases using image segmentation and soft computing techniques", *Information Processing in Agriculture*, vol. 4, no. 1, pp. 41–49, 2017.
5. D. Brahimi, K. Boukhalfa, and A. Moussaoui, "Deep learning for tomato diseases: classification and symptoms visualization", *Applied Artificial Intelligence*, vol. 33, no. 1, pp. 1–18, 2019.
6. K. Sharma and A. Jain, "Prediction of rice crop yield using hybrid machine learning approach", *International Journal of Computer Sciences and Engineering*, vol. 7, no. 5, pp. 100–105, 2019.
7. S. Ramesh and K. Vydeki, "Recognition and classification of paddy leaf diseases using Optimized Deep Neural network with Jaya algorithm", *Information Processing in Agriculture*, vol. 8, no. 2, pp. 273–285, 2021.
8. A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey", *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, 2018.
9. A. P. R. Santhosh and N. R. Patel, "Paddy crop disease detection using CNN model", *International Journal of Engineering Research & Technology (IJERT)*, vol. 10, no. 7, pp. 234–239, 2021.