An AI-Driven Disease Prediction System for Major Indian Cash Crops

Abhay Baliyan, Anushka Verma, Indervati

Department of Computer Science and Engineering
Galgotias University, Greater Noida –203201, India
Corresponding Author: abhaybaliyan11@gmail.com

ABSTRACT

The early and accurate identification of crop diseases remains a critical challenge in the agricultural sector. Symptoms vary for crops and diseases, making it extremely difficult for farmers to take crucial steps in a timely manner to sustain yields. This work presents a crop disease prediction system for early disease identification using images of plant leaves. Separate deep learning models were trained for several crops, namely - sugarcane, rice, wheat, cotton, and tea, enabling crop-specific learning of disease characteristics. On being provided a crop leaf as input, the system predicts the two most probable disease classes along with their associated probability. To demonstrate its practical usability, the trained models are deployed through a graphical user interface developed using PyQt5. A user can select a crop type and upload an image of the leaf to generate the probability of the leaf to be diseased and its Gradient-weighted Class Activation Mapping (Grad-CAM) for explain ability. This work expects to improve upon existing research by presenting a multi-crop and explainable framework for image-based plant disease prediction. The proposed system demonstrates the potential of integrating deep learning and explainable AI to support reliable decision-making in agriculture.

Keywords: Crop Disease Prediction, Deep Learning, Plant Leaf Image Analysis, Crop-Specific Models, Convolutional Neural Networks (CNN), Explainable AI, Gradient-weighted Class Activation Mapping (Grad-CAM).

I. INTRODUCTION

Agriculture in India can be considered to be the backbone of the economy as it provides livelihoods to a major part of the population, ensures national food security and contributes significantly to the GDP [1]. Crop diseases are considered one of the major concerns for farmers as they may lead to stunted growth and yield loss due to lack of timely detection [2]. Manual identification of diseases can be difficult for farmers due to lack of knowledge and different varieties of crops.

In recent studies, the use of Artificial Intelligence (AI) is shown to be promising in agricultural decision making [2], [3]. Advancements in deep learning have enabled automated prediction of plant leaf images, decreasing the dependency on manual inspection [4], [5]. Though many of the previous work focus on a single crop and limited disease prediction which limits their application in real-world scenarios where multiple crops might be cultivated simultaneously [6], [7]. Additionally, some studies prioritize accuracy of prediction without any explainability, therefore reducing trust of users and making it difficult to validate model decisions [3], [8].

This work focuses on crop-specific deep learning models trained for some of the major cash crops of India, namely – Cotton, Sugarcane, Rice, Tea and Wheat. Unlike most of existing work, separate models for each crop enable the system to focus on crop-specific characteristics instead of relying on general parameters [5], [6]. The model takes leaf images as an input and predicts the probability of the crop to be diseased and by which disease, providing quantitative insight based on the image uploaded and data used to train the model.

To show real-world application, the models are called using a Graphical User Interface (GUI) that allows the user to choose the crop and upload leaf image for prediction. Gradient-weighted Class Activation Mapping (Grad-CAM) has also been integrated into the GUI to show which regions of the image are responsible for the generated prediction [9]. The objective of this work is to present a user-friendly, explainable and practical crop disease prediction framework which targets multiple crops.

II. LITERATURE REVIEW

Earlier, research heavily relied on manual feature extraction techniques [10]. These approaches usually required manual descriptors such as texture, color, shape features, limiting their robustness under varying lighting conditions and field environments [2]. The Convolutional Neural Network (CNN) enhances the process by enabling automatic learning of features directly from leaf images [11].

A significant contribution in these studies is use of CNN in rice disease identification [6]. It shows that deep learning models on plant leaves can learn disease specific patterns without manual feature engineering [6], [11]. Although they show strong performance, usually they are limited to a single crop and a set of diseases, which limits their application across different scenarios [5], [7].

On the other hand, in some studies, models to identify multiple diseases are employed for cotton, containing cross-validation techniques to improve prediction, however, most of them remain crop-specific and dataset dependent [5].

ISBN: 978-81-689089-0-1

Some of these studies have also worked on multi-crop classification by using large datasets like 'PlantVillage' which enables efficient training of deep models avoiding manual feature building and hence saving time, however, these too are trained on lab type data, making them ineffective in real-world scenarios where factors like lighting and background can make a significant impact on results [11].

Recent research has shown importance of Explainable Artificial Intelligence (XAI) in plant disease prediction [8]. Techniques like Grad-CAM highlight specific areas which influence the model's decision making [9]. However, these are rarely shown at the user's end [4]. More advanced studies have also used sensors in their research but their complexity might make them inaccessible by small-scale farmers [12].

Overall, existing works show promising results in image-based crop disease prediction, but their key challenges (namely – crop-specificity, practicality, interpretability, user-friendly design) still remain [5], [8]. Addressing these gaps is essential for the development of suitable crop disease prediction systems.

III. METHODOLOGY

A. System Overview

The proposed crop disease prediction system is designed to identify diseases using deep learning models where each model is trained separately for respective crops. The workflow involves selecting a crop, image acquisition, preprocessing, disease classification, confidence calculation and visual explanation of prediction. Since separate models are present, learning of crop specific characteristics takes place instead of general classification [6].

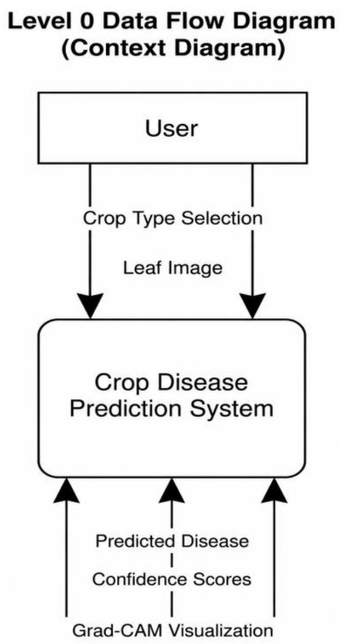

Fig. 1. Level 0 data flow diagram of the proposed system

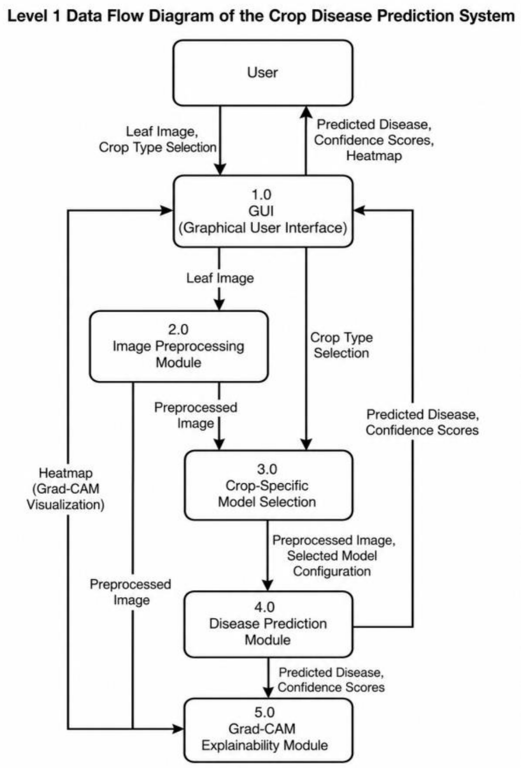

Fig. 2. Level 1 data flow diagram of the proposed system

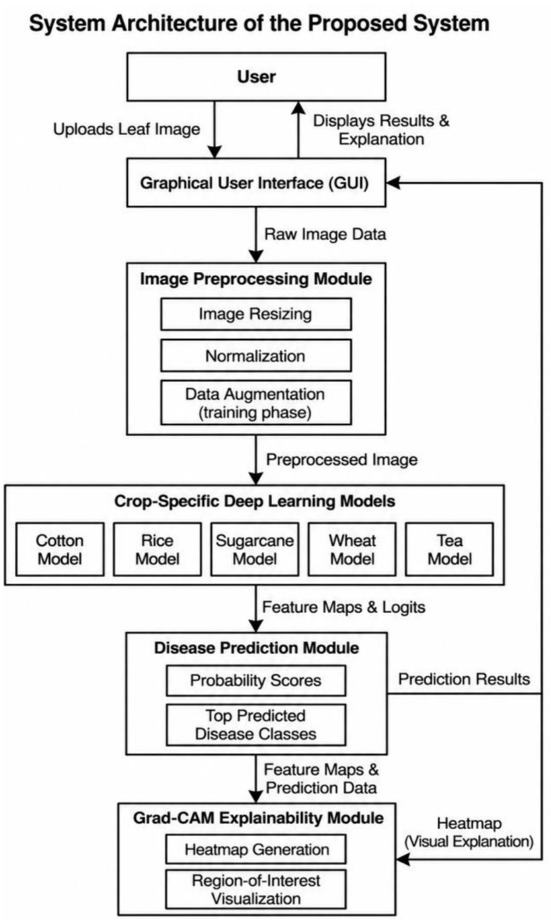

Fig. 3. System architecture of the proposed system

ISBN: 978-81-689089-0-1

B. Dataset Description

The system is trained on publicly available leaf image datasets which are collected from Kaggle for some of the major cash crops, namely – cotton, rice, sugarcane, tea and wheat [13], [14], [15], [16], [17]. Each dataset contains healthy leaves as well as multiple disease categories specific to that crop. For reliable evaluation, the dataset is divided into distinct training and validation subsets, maintaining class balance wherever possible. As some datasets do not provide a separate test subset, the validation subset is used for final performance evaluation.

TABLE I
DATASET DESCRIPTION FOR THE PROPOSED SYSTEM

Crop Name	Number of Classes	Total Images	Training Images	Validation Images	Test
Cotton	2	875	715	109	51
Rice	6	2,628	2,100	528	Not Present
Sugarcane	5	2,522	2,018	504	Not Present
Wheat	15	14,154	13,104	300	750
Tea	8	885	708	177	Not Present

C. Image Preprocessing

Before training, all the images undergo preprocessing to improve the model's robustness. Images are resized to a fixed resolution (224 x 224) and Pixel values are normalized to be compatible with the selected deep learning architecture (VGG16) [18] Augmentation techniques like rotation, horizontal flipping, and change in brightness are applied to during the training phase to reduce overfitting and improve generalization under real-world scenarios where lighting and background may differ.

D. Model Architecture and Training

For classification of diseases, CNN-based architecture is used due to its effectiveness in learning their specific visual features [11]. Transfer learning is achieved by fine-tuning pre-trained models on the respective crop datasets, allowing effective performance with limited data [6].Each model is trained independently using categorical cross-entropy loss and optimized using the Adam optimizer. Model's performance is monitored using validation accuracy and loss during training.

TABLE II
TRAINING CRITERIA FOR THE PROPOSED SYSTEM

Parameter	Value
Input Image Size	224 × 224
Model Architecture	VGG16 (Transfer Learning)
Pre-trained Weights	ImageNet
Optimizer	Adam
Batch Size	32
Loss Function	Categorical Cross-Entropy
Data Augmentation	Rotation, Horizontal Flip, Brightness Adjustment
Normalization	ImageNet Mean and Standard Deviation
Framework	PyTorch

E. Prediction and Confidence

For a pre-processed leaf image, the trained model calculates the probability scores as an output for all disease classes using a SoftMax activation function [11].. The two classes with the highest predicted probabilities are selected as confidence scores and shown as the final prediction results of the model, providing quantitative insight of model's certainty.

F. Explainability

To give a better explainability at user's level, Grad-CAM is integrated into the system. It generates the heatmaps that highlight the particular regions which contributes the most to predicted disease class by analyzing gradients going through the final convolutional layers [9]. The resulting heatmap is overlaid on the original image used as input for the model, allowing the users to visually verify if the model focuses on disease-relevant regions.

G. Graphical User Interface

To demonstrate practical usability, the trained models are deployed through a Graphical User Interface (GUI) which was developed using PyQt5. In the GUI, the user is prompted to select a crop type, upload a leaf image, and view the prediction results, which includes confidence score and Grad-CAM visualizations (if asked for). Input validation and feedback mechanisms are integrated to ensure smooth interaction and clean presentation.

IV. RESULTS

The proposed crop disease prediction system is evaluated using separate validation datasets according to each crop-specific model. The evaluation focuses on the performance, reliability and interpretability of each model.

ISBN: 978-81-689089-0-1

Each trained model is evaluated using standard classification metrics. Namely these metrics include accuracy, precision, recall and F1-score. They provide a deep understanding of model's ability to correctly identify disease classes. Confusion matrices are also generated to analyze misclassification and accurate classification of diseases. This also helps us to identify the diseases with similar patterns.

TABLE III
EVALUATION METRICS OF THE PROPOSED SYSTEM

Crop	Accuracy	Precision	Recall	F1-score
Cotton	98.00	98.00	98.00	98.00
Rice	94.00	94.00	94.00	94.00
Sugarcane	84.00	85.00	84.00	84.00
Tea	92.00	92.00	92.00	92.00
Wheat	89.00	91.00	89.00	88.00

The results show that crop-specific deep learning models are effective in learning disease related visual features from plant leaf images. Training and validation performance shows that the use of transfer learning and data augmentation helps in reducing overfitting and to improve generalization. The use of separate models for each crop contributes in improving learning of crop specific disease characteristics which can be a little challenging in a common multi-crop classifier [4].

Along with quantitative evaluation, the system also provides confidence scores for each prediction which acts as probability values. These scores provide insight about the model's certainty for individual prediction, allowing users to assess the reliability of the predicted disease classes. Showing the top predicted classes allow user to how model interpret in cases where symptoms overlap across diseases.

To enhance explainability, Grad-CAM visualizations are analyzed along with the prediction outputs [9]. The heatmaps highlight disease-relevant regions like discolorations, infected areas, etc. which are present on the leaf's surface. However, it has been observed that in some cases, background noise is also highlighted. This qualitative evaluation shows that the models focus on meaningful visual cues rather than irrelevant background regions, thereby increasing the user's trust in the prediction results [8].

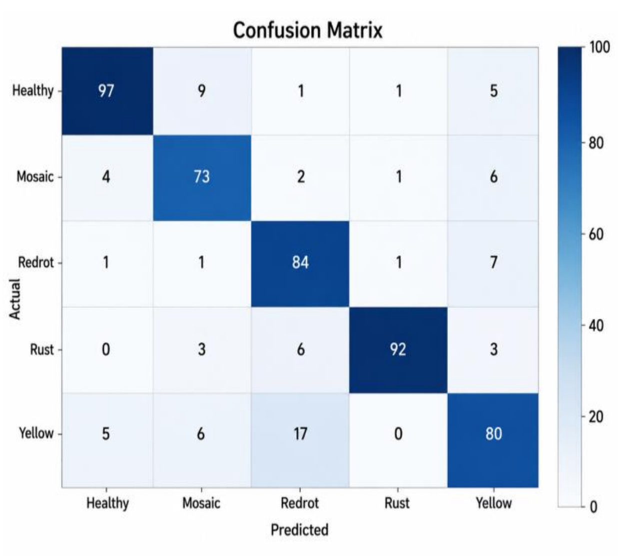

Fig. 4. Confusion matrix for a representative crop (sugarcane)

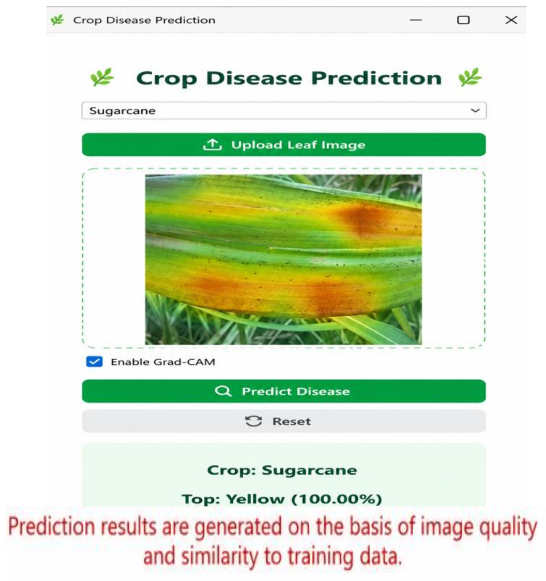

Fig. 5. Graphical user interface of the proposed system

Overall, the evaluation results show that the proposed system achieves reliable disease classification performance against multiple crops while maintaining practical usability. This supports the effectiveness of combining crop-specific models with confidence scores, visual explanations and GUI for early crop disease identification.

V. LIMITATION

Even after showing reliable performance, the proposed crop disease prediction system has some limitations. First, the system is trained on publicly available datasets that are limited in size and diversity. Many of the images in these datasets are captured in laboratory like environment with clean backgrounds. This affects model's performance in real life scenarios where factors like background, shadows, image quality, etc. are different [2].

Second, even though the system has crop-specific models which improve learning of crop-specific disease characteristics, it limits scalability with large number of

ISBN: 978-81-689089-0-1

crops as maintaining and updating multiple models increase computational and storage requirements [4].

Finally, the confidence score calculated by the model is based on data which is learned by the model at the time of training and hence should not be interpreted as absolute certainty. In situations where different diseases have similar symptoms, the model may give comparable scores to multiple disease classes.

VI. FUTURE WORK

The proposed crop disease prediction system can work as a foundation for similar systems. However, many enhancements can be made in future work to continue to improve upon the proposed system.

First, the system can be improved by incorporating more diverse and larger datasets, mostly using images captured in real field conditions with variation in lighting, background, camera quality, etc. to help it improve generalization [2].

Second, future work may focus on adding disease severity estimation along with disease identification to help the user understand the extent of infection and enable better decision-making for treatment.

Third, more information like weather data, soil data, etc. can be added to improve model. Combining visual features with contextual data may help to distinguish between different disease classes.

Overall, the system in the future should focus on improving its prediction, especially in real life conditions, for further refinement of practical relevance of system. A mobile or web-app might be useful as well to improve practicality and take feedbacks for further improvements.

VII. CONCLUSION

This work presents a crop disease prediction system based on deep learning models which are trained separately for multiple crops, which includes cotton, rice, sugarcane, tea and wheat. By using this approach, the system effectively learns disease related visual features from leaf images, avoiding the limitations of general multi crop disease classifiers. The proposed framework shows reliable performance and provides confidence scores that help quantify prediction certainty.

To increase practical usability, the trained models are deployed through a GUI that enables users to select crop type, upload leaf image, and get prediction result in the form of confidence score and Grad-CAM visualization to understand and validate the decision-making process.

This results of this work shows that combining deep learning models, along with visual explanation and confidence estimation, can support early identification of crop diseases in a user-friendly manner. Even though some limitations still remain regarding dataset diversity, the system provides a good foundation for future improvements.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to Ms. Indervati, Department of Computer Science and Engineering, Galgotias University, for her guidance and encouragement, and valuable insights throughout the course of this research work.

REFERENCES

[1] S. Jibran and A. Mufti, “Issues and challenges in Indian agriculture,” INTERNATIONAL JOURNAL OF COMMERCE AND BUSINESS MANAGEMENT, vol. 12, no. 2, pp. 85–88, Oct. 2019, DOI: 10.15740/HAS/IJCBM/12.2/85-88.

[2] V. Singh, N. Sharma, and S. Singh, “A review of imaging techniques for plant disease detection,” Artificial Intelligence in Agriculture, vol. 4, pp. 229–242, 2020, DOI: 10.1016/j.aiia.2020.10.002.

[3] J. M. Sánchez, J. P. Rodríguez, and H. E. Espitia, “Review of Artificial Intelligence Applied in Decision-Making Processes in Agricultural Public Policy,” Processes, vol. 8, no. 11, p. 1374, Oct. 2020, DOI: 10.3390/pr8111374.

[4] Md. M. Islam et al., “DeepCrop: Deep learning-based crop disease prediction with web application,” J Agric Food Res, vol. 14, p. 100764, Dec. 2023, DOI: 10.1016/j.jafr.2023.100764.

[5] U. G and G. BV, “A survey on plant disease prediction using machine learning and deep learning techniques,” Inteligencia Artificial, vol. 23, no. 65, pp. 136–154, Jul. 2020, DOI: 10.4114/intartif.vol23iss65pp136-154.

[6] K. N, L. V. Narasimha Prasad, C. S. Pavan Kumar, B. Subedi, H. B. Abraha, and S. V E, “Rice leaf diseases prediction using deep neural networks with transfer learning,” Environ Res, vol. 198, p. 111275, Jul. 2021, DOI: 10.1016/j.envres.2021.111275.

[7] L. Xu et al., “Wheat leaf disease identification based on deep learning algorithms,” Physiol Mol Plant Pathol, vol. 123, p. 101940, Jan. 2023, DOI: 10.1016/j.pmpp.2022.101940.

[8] M. Ryo, “Explainable artificial intelligence and interpretable machine learning for agricultural data analysis,” Artificial Intelligence in Agriculture, vol. 6, pp. 257–265, 2022, DOI: 10.1016/j.aiia.2022.11.003.

[9] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-cam: Visual explanations from deep networks via gradient-based localization,” in Proceedings of the IEEE international conference on computer vision, 2017, pp. 618–626.

[10] S. Barburiceanu, S. Meza, B. Orza, R. Malutan, and R. Terebes, “Convolutional Neural Networks for Texture Feature Extraction. Applications to Leaf Disease Classification in Precision

ISBN: 978-81-689089-0-1

Agriculture," IEEE Access, vol. 9, pp. 160085–160103, 2021, DOI: 10.1109/ACCESS.2021.3131002.

[11] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," Front Plant Sci, vol. 7, Sep. 2016, DOI: 10.3389/fpls.2016.01419.

[12] G. S. Hundal, C. M. Laux, D. Buckmaster, M. J. Sutton, and M. Langemeier, "Exploring Barriers to the Adoption of Internet of Things-Based Precision Agriculture Practices," Agriculture, vol. 13, no. 1, p. 163, Jan. 2023, DOI: 10.3390/agriculture13010163.

[13] "Rice Leaf Disease Dataset," Kaggle. Accessed: Jan. 02, 2026. [Online]. Available: https://www.kaggle.com/datasets/dedeikhsandwisaputra/rice-leafs-disease-dataset

[14] "Cotton Disease Dataset," Kaggle. Accessed: Jan. 02, 2026. [Online]. Available: https://www.kaggle.com/datasets/janmejaybhoi/cotton-disease-dataset/data

[15] Daphal, Swapnil, Koli, and Sanjay, "Sugarcane Leaf Disease Dataset," Kaggle. Accessed: Jan. 02, 2026. [Online]. Available: https://www.kaggle.com/datasets/nirmalsankalana/sugarcane-leaf-disease-dataset

[16] "Tea Leaves Disease Classification," Kaggle. Accessed: Jan. 02, 2026. [Online]. Available: https://www.kaggle.com/code/rizqyad/tea-leaves-disease-classification/input

[17] Kushagra Agarwal, Vikas Yadav, and Tanish Suthar, "Wheat Plant Diseases," Kaggle. Accessed: Jan. 02, 2026. [Online]. Available: https://www.kaggle.com/datasets/kushagra3204/wheat-plant-diseases

[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

Cite this article as:

Abhay, Indervati and et.al. "An AI-Driven Disease Prediction System for Major Indian Cash Crops", Proceedings of 13th international conference on Microelectronics, Circuits and Systems, Micro2026 (Vol-II)

Displayed as online on 19th August 2026.

Link: http://actsoft.org/science/micro2026-pro/365-micro2026.pdf

ISBN: 978-81-689089-0-1