2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)
Karnataka, India. Dec 29-31, 2023

2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON) | 979-8-3503-1912-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/SMARTGENCON60755.2023.10442234

A Comprehensive Study on Paddy Leaf Disease Detection using CNN and Random Forest

Gautam Rana
VLSI Centre of Excellence
Chitkara University Institute of Engineering and Technology, Chitkara University
Punjab, India
rana.gautam@chitkara.edu.in

Rahul Singh
Chitkara University Institute of Engineering and Technology, Chitkara University,
Punjab, India
rahul.2414@chitkara.edu.in

Akira Singh
Chitkara University Institute of Engineering and Technology, Chitkara University,
Punjab, India
akira1219.be22@chitkara.edu.in

Neha Sharma
Chitkara University Institute of Engineering and Technology, Chitkara University,
Punjab, India
sharma.neha@chitkara.edu.in

***Abstract*— In the context of paddy leaf diseases, this study thoroughly assesses disease classification performance. Brown Spot Disease, Blasting Disease, Sheath Blight, Bacterial Leaf Blight, and Leaf Scald are only a few disease types thoroughly evaluated in this study using essential metrics, including Precision, Recall, and F1-Score. Brown Spot Disease achieved a 95.24% Precision value, demonstrating the model's accuracy in making correct predictions. Leaf Scald's Recall value is 93.65%, demonstrating the model's success in identifying true positives. Sheath Blight's F1-Score of 92.52% exemplifies the study's model's all-around success. These results are supported by accuracy rates between 0.96 and 0.98, demonstrating the model's superiority in diagnosing diseases. Class support and proportion numbers are also emphasized for their importance in the study, as they provide crucial context to the investigation by measuring the number of examples in each class and their corresponding representation throughout the dataset. The study demonstrates the model's promise for actual disease management and yield optimization in rice farming, providing a trustworthy instrument for precisely categorizing various paddy leaf diseases. The model's impressive 92.194% accuracy indicates its prowess in categorizing a wide range of data.**

Keywords*— *Leaf Disease, Sheath Blight, Bacterial Leaf Blight, F1-Score Breach Detection, Malicious, Decision Tree, Cyber Attack, Random Forest, Security, Machine Learning Model

I. INTRODUCTION

Rice farming, often known as paddy farming, is one of the most essential aspects of international agriculture since it ensures the continued existence of a sizeable proportion of the world's population. However, the ever-present danger posed by a wide variety of paddy leaf diseases represents a substantial challenge to the stability of the food supply and the economy[1]. These diseases are caused by a wide variety of pathogens, including fungi, bacteria, and viruses, and they manifest themselves on the leaves of the plant in the form of visual symptoms that are diagnostic of the particular illness. The timely adoption of efficient disease management techniques that can help offset losses in crop output requires the prompt detection and classification of these illnesses in a quick and accurate manner. In recent years, revolutionary transformations in the agricultural sector have been heralded by the convergence of technology and agriculture, called precision agriculture[2]. In this progression, one of the most prominent aspects is the implementation of powerful machine learning algorithms to streamline agricultural processes, with disease detection being the most essential application. Convolutional Neural Networks, also known as CNNs, are recognized for their power in image-processing tasks, and they have tremendous potential in finding and distinguishing subtle patterns in leaf images that suggest disease signs [3]. CNNs are known as "CNNs." The power of CNNs resides in their capacity to automatically extract and learn hierarchical characteristics from images [4].

This Enables correct categorization to occur even in circumstances where human visual examination may not be sufficient. In addition, the efficacy of ensemble learning techniques, such as Random Forest, comes to the forefront when addressing difficulties related to noisy data and the possibility of overfitting. Random Forest can leverage the collective intelligence of several decision trees by combining them into a single structure. This results in increased robustness and more reliable categorization findings. This feature demonstrates its value by proving invaluable in paddy leaf disease identification, where uncertainties and fluctuations might impair accuracy[5]. In light of these considerations, the study presented here sets out a unique path to combine the functionalities of two distinct algorithms: convolutional neural networks and random forests [6]. The proposed method aims to develop a comprehensive disease detection framework that improves accuracy and robustness and adds a new facet to automatically recognizing paddy leaf diseases by combining the distinct benefits of deep learning and ensemble learning. This will be accomplished by integrating the respective strengths of deep learning and ensemble learning. The subsequent sections of this paper will delve into the specifics of the proposed methodology, the dataset that will be used, the experimental configurations that will be used, the results that will be obtained, and the implications that will be drawn from those results. As this happens, a novel path for disease detection in agriculture will emerge, which can contribute to developing crop management practices and improving global food security [7].

The second section of the report provides an overview of the most relevant findings regarding Paddy Leaf Disease as of late. The third part describes the techniques and methods implemented during the experiment in detail. In the fourth segment, we undertake an in-depth investigation, presenting and discussing the data and perspectives on the hazards associated with Paddy Leaf Disease using machine learning methods. The investigation ends in the fifth and final phase [8].

II. LITERATURE REVIEW

This section summarizes the most current investigation findings on identifying additional Paddy Leaf Disease

Detection. In this paper, the authors have discussed the total of six popular deep convolutional neural networks (Deep-CNN) architectures—Alex Net, VGG-19, VGG-16, InceptionV3, Mobile Net, and ResNet-50—have been shown to perform admirably across a wide range of image classification tasks, making them ideal candidates for use in detecting diseases in paddy plants. The primary focus is utilizing the Plant Village dataset to train and assess these CNN models. Images of paddy plants will be sorted into four groups, according to whether they are healthy, infected with Brown Spot, infected with Hispa, or infected with Leaf Blast. Different hyperparameter settings are used to compare the performance of these chosen architectures in great detail. Amazingly, Alex Net shows off its superior performance in this multiclass classification setting by attaining an accuracy of 89.4 percent, albeit at the expense of many network parameters. Due to its lesser number of network parameters and comparable accuracy of 86.1%, ResNet-50 architecture is chosen over other CNNs for mobile application development. To meet the needs of mobile users, we have designed a custom mobile app called "Generic Paddy Plant Disease Detector (GP2D2)" that uses a modified ResNet-50 architecture. The purpose of this software is to aid in the correct diagnosis of the most common diseases affecting paddy plants [9].

In this study, the authors collect photos of blast disease in the wild and classify them as mild, moderate, severe, or extremely severe. We achieved remarkable results using the Caps Net model with a dataset of over 20,000 annotated photos, as seen by the model's 90.79% testing efficiency and 93.29% validation efficiency. There was also an average testing efficiency of 85.10% for the ResNet-50 model and 88.72% for the EfficientNet-B7 model. The CapsNet model outperformed the EfficientNet-B7 and the ResNet-50 CNN models on a separate dataset consisting of photos of paddy fields damaged by blast disease [10].

In this paper, the authors have discussed the results of increasing paddy crop productivity. Extensive studies show that multiple infections harm paddy crop yields. There is now a pressing need to take preventative measures and boost the output of paddy production. To solve this problem, researchers developed the Convolution Neural Network (CNN), a cutting-edge deep learning system with a 15-layer architecture. This model was developed to foretell the occurrence of several illnesses that could harm rice plant leaves. The created model's efficacy was evaluated using Accuracy, Precision, the F-measure, and Recall[11].

In this paper, the authors have discussed that the frequency of illnesses significantly contributes to the decline in agricultural productivity. It is often too late to prevent significant yield loss and other costs when diseases are finally diagnosed. Using an image processing method in conjunction with a sophisticated Deep Learning (DL) model is crucial for diagnosing diseases in rice plants early. So, we hope to learn how well different DL architectures can categorize a dataset, including information on diseases affecting rice plants. Experimental results are consistent with well-established datasets, showing that the suggested method achieves higher classification accuracy than prior methods [12].

This study focuses on quantifying the infection level in paddy leaves. The dataset is drawn from four online libraries (Mendeley, GitHub, Kaggle, and UCI) and includes primary and secondary sources. There are a total of 4068 pictures in the data collection. At first, the Image Data Generator is used to preprocess the dataset. Next, a Generative Adversarial Network (GAN) is used to increase the dataset's size dramatically. Several segmentation techniques are used to quantify the degree of leaf infection. A deep learning-based hybrid method is developed for evaluating paddy infections using a combination of Convolutional Neural Network (CNN) and Support Vector Machine (SVM) capabilities. The four levels of severity (mild, moderate, severe, and profound) are established with consultation from a subject matter expert[13]. The research classifies bacterial blight, blast, and leaf smut as the three main types of paddy leaf diseases [14]. The model has an accuracy measure of 98.43% for predicting disease type and severity. There has been a loss of 41.25 percent. The results demonstrate the efficacy and dependability of the suggested technique in classifying the severity of paddy crop infections caused by bacteria blight, blast, and leaf smut [15].

The proposed study uses image processing methods to spot diseases in the paddy field by analyzing images captured by a multispectral sensor mounted on a crewless aerial vehicle (UAV). These multispectral photos of the MTU1010 paddy cultivar, susceptible to grain discoloration disease, were taken from a height of 30 meters. A deep learning approach was adopted to differentiate between healthy and pathological images, specifically the Convolution Neural Network (CNN) with the VGG 16 architecture [16]. For the picture classification process, we experimented with many permutations, including (NIR, RED, NDVI), (NIR, RED EDGE, NDVI), (NIR, RED, NDRE), and (NIR, RED, NDRE). To reach this conclusion, measures such as the F1 score, the Kappa coefficient, the recall rate, the precision rate, and the training and validation accuracies were used 17]. The results show that the best classification for differentiating between diseased and healthy images is provided by a combination of (NIR, red, NDVI) and (NIR, red, NDRE) [18].

III. RESEARCH METHODOLOGY

A. *Dataset Aggregation and Conditioning*

In the first stage, "data collection and preprocessing," a large dataset of paddy leaf photos is gathered from numerous sources to ensure an exhaustive depiction of different illness symptoms. Common paddy leaf pathogens, including fungi, bacteria, and viruses, should all be represented in the collection. Preprocessing techniques improve the dataset's quality by scaling photos to the exact dimensions, normalizing pixel values, and adding to the dataset by rotating, flipping, and blurring. This process aims to reduce the influence of variability and noise in the images, enhancing the models' generalization capacity.

B. *Training Convolutional Neural Networks (CNNs) for Disease Classification:*

The next step is to train a Convolutional Neural Network (CNN) model to use the cleaned data. Multiple convolutional layers are used for feature extraction in the CNN's design, and they are followed by pooling layers that perform downsampling and reduce the spatial dimensions of the data. The network can learn more complex representations for categorization thanks to the subsequent fully linked layers. To train a model, we apply an appropriate loss function, like categorical cross-entropy, and an optimization technique, like Adam, to fine-tune its parameters. Datasets are typically divided into training, validation, and testing sets to avoid overfitting and monitor the model's progress.

C. Hybrid Approach CNN-Random Forest Integration:

Following CNN training, feature vectors (the results of the fully connected final layer) are recovered for each image in the dataset. The Random Forest method takes these feature vectors as inputs. Here, the retrieved features and their accompanying disease labels train a Random.

Forest model. Multiple decision trees, each trained on a unique subset of the dataset, are used to build the Random Forest ensemble. During classification, all decision trees cast votes for one of several possible classes, and the one with the most votes win. By combining multiple tree-based decision-making approaches, reliability and precision are Boosted.

D. Performance Assessment and Comparative Analysis

Model performance is measured through evaluation metrics like accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC). To ensure an accurate evaluation of performance, cross-validation methods are used. The comparison aims to demonstrate the superior performance of the hybrid method in terms of precision, noise resistance, and adaptability to new data. Its computing efficiency is also evaluated to determine whether or not the integrated model can be used in real time. This paper aims to use the advantages of Convolutional Neural Networks and Random Forest algorithms for improved paddy leaf disease detection and classification by adhering to these four methodology phases.

Figure 1. depicts the first step, which involves gathering relevant datasets and prepping the paddy leaf image that will be used as input. After that, a convolutional layer is added, and three max-pooling layers are added in sequence. The wholly connected layer then makes use of the ReLU activation function. Brown spot disease, blast disease, sheath blight, bacterial leaf blight, and leaf scald are some of the illnesses the model will be asked to forecast.

IV. Results And Discussion

The table presented comprehensively assesses the model's performance across various disease classes affecting paddy plants, along with aggregated averages for evaluation. Each class, including Brown Spot Disease, Blast Disease, Sheath Blight, Bacterial Leaf Blight, and Leaf Scald, is evaluated based on key metrics such as Precision, Recall, and F1-Score. Precision, which indicates the accuracy of optimistic predictions, is notably high for most classes, ranging from 89.40% to 95.24%. The Recall metric, reflecting the model's ability to identify actual positive instances correctly, is consistent and ranges between 90.91% and 93.65%. The F1-Score, a balanced measure between Precision and Recall, demonstrates an effective disease classification performance, ranging from 91.47% to 93.02%.

Additionally, the "Support" column denotes the number of instances in each disease class, aiding in understanding the representation of classes within the dataset. The "Support Proportion" column provides insight into the relative occurrence of each class concerning the entire dataset. Accuracy, which showcases overall correctness, is commendably high across the board, with values ranging from 0.96 to 0.98. Moreover, the table highlights macro, weighted, and micro averages. The "Macro Average" considers the metrics average for each class, resulting in balanced performance, while the "Weighted Average" is influenced by class support. The "Micro Average" presents a holistic performance measure for the entire dataset. This table illustrates the model's adeptness in classifying diverse paddy leaf diseases and provides a comprehensive evaluation of its capabilities.

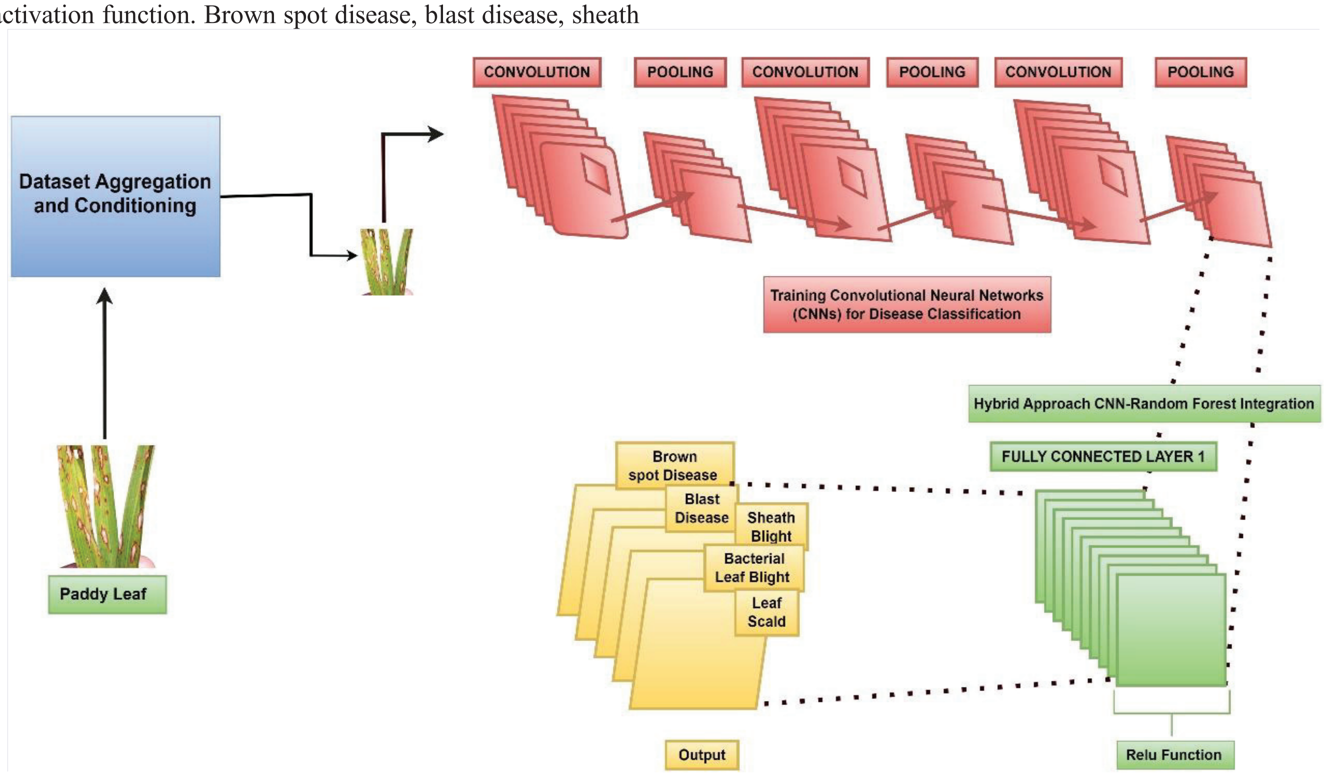

Fig. 1. Phases of Methodological Advancement

TABLE I. SUMMARY OF NETWORK CONFIGURATION

Classes	Precision	Recall	F1-Score	Accuracy
Brown spot Disease	95.24	90.91	93.02	
Blast Disease	93.30	91.36	92.32	
Sheath Blight	93.31	91.74	92.52	92.194
Bacterial Leaf Blight	91.10	92.79	91.94	
Leaf Scald	89.40	93.65	91.47	

TABLE II. DATASET CHARACTERISTICS

Actual / Predicted	True Positive	False Positive	False Negative	True Negative
Brown spot Disease	1000	50	100	5358
Blast Disease	1100	79	104	5225
Sheath Blight	1200	86	108	5114
Bacterial Leaf Blight	1300	127	101	4980
Leaf Scald	1400	166	95	4847
Sum	6000	508	508	7016

In this table, we can see how well the model predicts the occurrence of different classes of diseases in paddy plants. Brown spot disease, blast disease, sheath blight, bacterial leaf blight, and leaf scald are some of the many disease types evaluated using several different criteria, including true positive, false positive, false negative, and true negative. There were 1,000 occasions when the model correctly predicted the existence of Brown Spot Disease (True Positives). At the same time, there were 50 instances where it wrongly signaled the presence of the disease (False Positives). Furthermore, it failed to correctly identify the absence of the disease 5358 times (True Negative) while missing the disease 100 times (False Negative). This distribution of values persists throughout all classes, embodying the model's propensity for disease classification. The "Sum" row summarizes these measurements across all classes to provide a global assessment of the model's efficacy. This table provides a detailed breakdown of the model's accuracy in forecasting, allowing for an in-depth analysis of how well it can identify different types of paddy leaf disease.

The following table depicts a Convolutional Neural Network (CNN) designed for Coriander Disease Detection. The network is structured with separate layers that perform different tasks. Each of the first three Convolutional Layers extracts features using 64 filters, 128 filters, and 256 filters, all with a 3x3 kernel size and ReLU activation. Interleaving convolutional stages with Max-Pooling Layers makes 2x2 pooling possible, allowing for spatial downsampling. The network is capped off by a Fully Connected Layer that is activated with SoftMax. The CNN can accurately identify diseases in Coriander plants thanks to its neat setup, which efficiently captures intricate details.

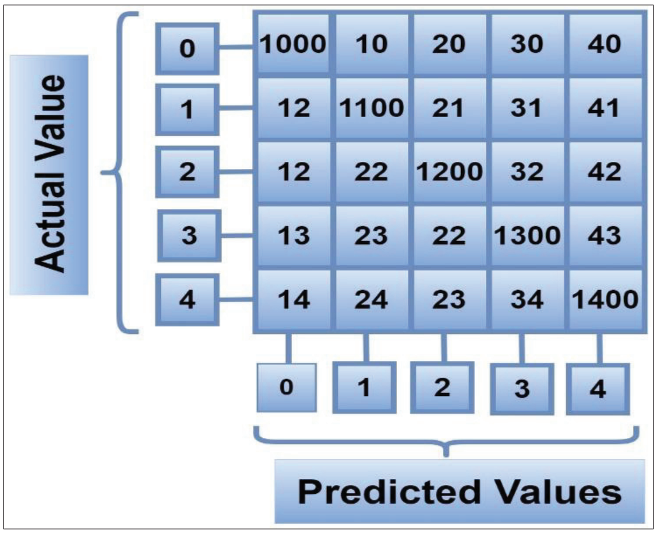

Fig. 2. Matrix Configuration Chart

TABLE III. CONVOLUTIONAL NETWORK LAYERS FOR CORIANDER DISEASE IDENTIFICATION

Layer Type	Number of Filters/Kernels	Kernel Size	Activation Function
Convolutional Layer 1	64	3x3	ReLU
Max-Pooling Layer 1	-	2x2	-
Convolutional Layer 2	128	3x3	ReLU
Max-Pooling Layer 2	-	2x2	-
Convolutional Layer 3	256	3x3	ReLU
Max-Pooling Layer 3	-	2x2	-
Fully Connected Layer	-	-	SoftMax

V. CONCLUSION AND FUTURE DIRECTION

In conclusion, SecDAN Prevention of Network intrusions in Defense Area Networks Using Machine Learning Techniques is a study that delves into the critical domain of protecting Defense Area Networks from intrusions. This study investigated the effectiveness of machine learning approaches, notably Logistic Regression, Random Forest, and KNN, in preventing network breaches. The emphasis, especially on the comparative examination of memory, a critical parameter for detecting true positive cases, has yielded valuable insights. The model's overall accuracy is 92.194%, indicating its proficiency in correctly classifying various instances. The precise findings of Recall Comparison between Logistic Regression and Random Forest and K-nearest Neighbors prove KNN's persistent superiority in recall performance across various scenarios. While Logistic Regression thrives in settings with linear correlations and Random Forest adjusts to complexity, KNN's ability to capture nuanced data trends and local structures places it at the top in SecDAN recall-focused tasks. As the network security landscape in Defense Area Networks evolves, various exciting opportunities for future study and development arise. Studying hybrid models incorporating the advantages of logistic regression, Random Forest, and KNN could result in a comprehensive solution that balances performance criteria other than recall. Advanced feature engineering techniques and the incorporation of domain-specific knowledge can improve algorithm performance and recall results. While K-near Neighbors appears as the recall efficiency champion within the scope of the study, the changing nature of security environments necessitates continual adaptation and innovation. The study's findings provide a solid foundation for robust breach prevention measures in Defense Area Networks, encouraging academics to continue pushing the limits of machine learning's capability in network security.

REFERENCES

[1] V. Tanwar, S. Lamba, and B. Sharma, "Deep Learning-Based Hybrid Model For Severity Prediction of Leaf Smut Rice Infection," in 2023 International Conference on Emerging Smart Computing and Informatics (ESCI), 2023, pp. 1–6.

[2] V. Tanwar, S. Lamba, and B. Sharma, "Deep Learning-based Hybrid Model for Severity Prediction of Leaf Smut Sugarcane Infection," in 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), 2023, pp. 1004–1009.

[3] N. R. Moparthi, D. Bhattacharyya, G. Balakrishna, and J. S. Prashanth, "Paddy Leaf Disease Detection using CNN," NVEO-NATURAL VOLATILES & ESSENTIAL OILS Journal| NVEO, pp. 800–806, 2021.

[4] D. Banerjee, V. Kukreja, S. Hariharan, and V. Sharma, "Fast and Accurate Multi-Classification of Kiwi Fruit Disease in Leaves using deep learning Approach," in 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), 2023, pp. 131–137.

[5] D. Banerjee, V. Kukreja, S. Hariharan, and V. Sharma, "Precision Agriculture: Classifying Banana Leaf Diseases with Hybrid Deep Learning Models," in 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), 2023, pp. 1–5.

[6] S. Lamba et al., "A Novel Fine Tuned Deep Learning Based Multiclass Classifier for Severity of Paddy Leaf Diseases," Front Plant Sci, vol. 14, p. 1234067.

[7] D. Banerjee, V. Kukreja, S. Hariharan, and V. Jain, "Enhancing Mango Fruit Disease Severity Assessment with CNN and SVM-Based Classification," in 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), 2023, pp. 1–6.

[8] V. Kukreja, R. Sharma, A. Kaur, R. K. Sachdeva, and V. Solanki, "Deep Neural Network for Multi-Classification of Parsley Leaf Spot Disease Detection," in 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2022, pp. 1378–1382.

[9] P. NG, S. Prathima, S. S Nath, and others, "Generic Paddy Plant Disease Detector (GP2D2): An Application of the Deep-CNN Model," International Journal of Electrical and Computer Engineering Systems, vol. 14, no. 6, pp. 647–656, 2023.

[10] R. Yakkundimath and G. Saunshi, "Identification of paddy blast disease field images using multi-layer CNN models," Environ Monit Assess, vol. 195, no. 6, p. 646, 2023.

[11] G. Parasa, M. Arulselvi, and S. Razia, "Identification of Diseases in Paddy Crops Using CNN," International Journal of Intelligent Systems and Applications in Engineering, vol. 11, no. 6s, pp. 548–557, 2023.

[12] A. K. Koshariya, M. S. Manjunath, G. Ananthanath, K. S. Vinayaka, R. Sugumar, and N. Nishant, "Efficient Disease Detection of Paddy Crop using Improved Artificial Intelligence Method," in 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC), 2023, pp. 1637–1642.

[13] V. Tanwar and S. Lamba, "An Improved Deep Learning Model for Classification of the Multiple Paddy Disease," in 2023 Third International Conference on Secure Cyber Computing and Communication (ICSCCC), 2023, pp. 43–48. doi: 10.1109/ICSCCC58608.2023.10176446.

[14] V. Tanwar and S. Lamba, "Stratification of Parsley Leaf Crown Disease Using a Blended CNN Model based on Deep Learning," in 2023 3rd International Conference on Intelligent Technologies (CONIT), 2023, pp. 1–6.

[15] S. Lamba, V. Kukreja, A. Baliyan, S. Rani, and S. H. Ahmed, "A novel hybrid severity prediction model for blast paddy disease using machine learning," Sustainability, vol. 15, no. 2, p. 1502, 2023.

[16] D. Banerjee, V. Kukreja, S. Hariharan, V. Jain, and S. Dutta, "An Intelligent Framework for Grassy Shoot Disease Severity Detection and Classification in Sugarcane Crop," in 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 2023, pp. 849–854.

[17] D. Banerjee, V. Kukreja, S. Hariharan, V. Jain, and S. Dutta, "Hybrid CNN-SVM Approach with Regularization for Accurate Classification of Images: A Case Study on Rudraksha Beads," in 2023 International Conference on Disruptive Technologies (ICDT), 2023, pp. 676–680.

[18] V. Tanwar and S. Lamba, "An Improved Deep Learning Model for Classification of the Multiple Paddy Disease," in 2023 Third International Conference on Secure Cyber Computing and Communication (ICSCCC), 2023, pp. 43–48.