SICASH 2023
SLIIT INTERNATIONAL CONFERENCE ON ADVANCEMENTS IN SCIENCES & HUMANITIES

Disease Identification and Mapping using CNN in Paddy Fields

* [1]W. D Nilakshi Sandeepanie, [2]Samadhi Rathnayake, [3]Amali Gunasinghe

[1,2,3]Department of Data Science, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka

Corresponding author - *sandeepanien@gmail.com

ARTICLE INFO

Article History:
Received: 10 September 2023
Accepted: 01 November 2023

Keywords:
Machine learning; Object detection; Paddy cultivation; Web development; YOLO v8

Citation:

W. D Nilakshi Sandeepanie, Samadhi Rathnayake, Amali Gunasinghe. (2023). Disease Identification and Mapping using CNN in Paddy Fields. Proceedings of SLIITInternational Conference on Advancements in Sciences and Humanities, 1-2 December, Colombo, pages 285-289.

ABSTRACT

Rice, a globally vital staple crop, sustains over half of the world's caloric needs while supporting the livelihoods of small-scale farmers and landless laborers. The escalating global population has led to an increased demand for rice production. Sri Lanka, renowned for its premium rice quality, has a rich history of paddy cultivation. However, a substantial portion of the country's 708,000 hectares of paddy land remains underutilized due to water scarcity and unstable terrain. The objective of this project is to enhance paddy crop quality during the critical vegetative phase by employing machine learning and web development for early disease identification. The vegetative phase significantly influences overall yield, resistance to pests and diseases, nutrient assimilation, and environmental sustainability in agriculture. This project primarily focuses on early disease identification during this phase and presents the findings through a user-friendly map interface. Early identification of paddy diseases is vital for effective crop management and high yields. These diseases, caused by various pathogens, can severely impede plant growth and productivity if not promptly detected and treated. Identifying them early enables farmers and experts to take timely, targeted actions such as applying suitable fungicides or implementing

sandeepanien@gmail.com

https://doi.org/10.54389/NKKJ6476

cultural practices to control their spread and minimize crop damage. A logical map, displaying disease spread percentages, will gauge the impact of infections on paddy plants. The reliability of this mapping process hinges on model accuracy, which was rigorously validated using multiple metrics to ensure its effectiveness.

1. INTRODUCTION

The paddy crop undergoes a comprehensive lifecycle encompassing seven distinct stages. These stages include the Pre-planting stage, Planting stage, Vegetative stage, Reproductive stage, ripening stage, Harvesting stage, and post-harvest stage. The initial phase, known as the Pre-planting stage, involves meticulous land preparation and the careful selection of suitable seed varieties. It encompasses tasks such as land ploughing, levelling, and irrigation. Subsequently, the second stage entails either direct seeding or transplanting of the chosen seeds. The ensuing Vegetative stage marks the commencement of the paddy plant's growth. During this phase, leaves emerge from the shoot apex, and the root system undergoes development. Notably, the Vegetative stage is crucial for the successful growth and development of paddy plants as it facilitates photosynthesis and stem elongation. It lays the groundwork for the subsequent stages of the plant's lifecycle. In this study, the researchers have specifically chosen to focus on the pivotal Vegetative phase and have selected the 'Broadcasting method' for planting. Object detection plays a crucial role in identifying diseases in paddy crops. By employing advanced computer vision techniques and machine learning algorithms, object detection systems can analyse images or video footage of paddy fields and accurately detect signs of diseases or infections. The system can identify specific symptoms such as discoloration, lesions, or unusual growth patterns on the leaves or stems of paddy plants. With the help of object detection, farmers and agricultural experts can quickly and efficiently assess the health status of paddy crops over large areas, enabling them to take timely actions to prevent the spread of diseases.

This project proposes a way to recognize diseased crops using an object detection technique. The pre-identified diseased crops or the clusters of crops with symptoms will be displayed using a map to the end user. Additionally, a high-level overview of the spread of the diseases inside a chunk of land will be provided to the end user.

During the vegetative phase, rice plants are vulnerable to a range of diseases, including Blast, Tungro, Sheath Blight, Bacterial Leaf Blight, and Brown Spot. These diseases can cause significant damage to the plants, reducing their ability to photosynthesize and produce healthy grains. In severe cases, they can even lead to plant death. Therefore, pre-identification of diseases in a paddy field during the vegetative phase is important to prevent or control disease, improve crop yields and quality, and make informed decisions about inputs and management practices.

2. MATERIALS AND METHODS

Crop diseases pose a significant threat to agricultural productivity and food security worldwide. Timely identification and management of these diseases are crucial to mitigate losses. In this research, the focus is on the disease identification in paddy crops, utilizing the Osmo V3 device for image collection and the YOLO v8 algorithm for automated disease detection. The objective is to develop an accurate, efficient, and scalable solution to aid farmers in early disease detection and effective crop management.

2.1. DATASET COLLECTION

To build a robust disease identification system, a diverse and representative dataset is essential. An image dataset comprising around 5000 high-resolution images of paddy crops was collected,and captured using the DJI Osmo V3 device and a smart mobile phone. The gimbal stabilization system of the device helps reduce camera shake, allowing for smoother and more professional-looking shots. The device is particularly useful when capturing footage from moving vehicles or when walking through uneven terrain. Although a drone is a matching solution for the given scope, the wind generated by the drone's propellers potentially affects the quality of photographs taken while flying over a paddy field. This movement results in blurry or distorted images, especially if the exposure time of the camera is relatively long. To get the expected output from the system, the images are recommended to be captured row-wise. In brief, the capturing process should be done according to a pattern. The images were acquired from various geographical locations and encompassed different stages of disease progression.

2.2. LABELLING THE DATASET

The labelling process involved annotating or marking objects of interest within the images with bounding boxes and corresponding class labels. In this case, the exact places infected by diseases were bound using a box. After bounding the affected areas, the disease type related to the bounded box should be chosen using a drop-down list. For the labelling process, an inbuilt labelling software specialized for the YOLO algorithm was chosen.

2.3. PREPROCESSING AND AUGMENTATION

To enhance the quality of the dataset and to improve the generalization ability of the model, preprocessing and augmentation techniques were applied. Noise reduction techniques, such as image denoising and contrast enhancement, were employed to improve image clarity. Data augmentation techniques, including random rotations, flips, and translations, were applied to increase the dataset's diversity and robustness.

2.4. SELECTING SUITABLE MODEL

The pre-processed data is then divided into three major categories known as 'Train', 'Test' and 'Valid' to be deployed in the YOLO v8 (You Only Look Once Version 8) model. The training dataset is used to train the YOLO v8 model. It consisted of many labelled images, where each image is annotated with bounding box coordinates and class labels for the diseases present. The test dataset is used to evaluate the performance of the trained YOLO v8 model. It contained a separate set of images that were not seen during the training process. The validation dataset is used to fine-tune the hyperparameters and monitor the training progress. The reason for choosing YOLO V8 is due to its state-of -art performance and real-time processing capabilities. YOLO v8 utilizes a single deep neural network to simultaneously predict bounding boxes and class probabilities in a single pass. This architecture enables fast and accurate detection, making it suitable for large-scale disease identification in agricultural settings.

2.5. TRAINING AND MODEL DEVELOPMENT

The first install the necessary libraries and dependencies such as 'OpenCV', 'Numpy' and 'Matplotlib'. Defined the YOLOv8 configuration and downloaded the pre-trained weights. Then, loaded the YOLOv8 model and labels while providing the number of epochs. The Yolo V8 model is compatible with arbitrary sized images as long as both sides of the images are multiple of 32. Therefore, in this case, image resizing techniques were not applied.

3. RESULTS AND DISCUSSION

3.1. PERFORMANCE EVALUATION

To assess the performance of the developed model, a separate code was written based on the testing image set. A set of testing image data set affected by diseases, bounded by a frame with an accuracy score of identifying the disease was get as the output of the code. Additionally, a visual inspection of the detected bounding boxes and class predictions was conducted to analyse the model's performance qualitatively. A performance test with few parameters is conducted to test the accuracy of the model.

As the first parameter, 'precision' was considered as a measure of the accuracy of positive predictions. In the context of object detection, it represents the proportion of predicted bounding boxes that contain objects of interest (true positives) out of all predicted bounding boxes (Table 1).

The second parameter was 'recall' used to measure the proportion of actual positive objects that are correctly identified by the model (Table 1).

The third parameter is metrics/mAP50(B): Mean Average Precision and calculates the average precision at a detection threshold of 0.5 (Table 1). '50' represents the IoU (Intersection over Union) and is a measure of the overlap between predicted bounding box and ground truth bounding box. metrics/mAP50-95(B): mAP50-95 calculates the average precision over different confidence thresholds ranging from 0.5 to 0.95.(Table. 1).

Val/box_loss: Box loss measures the discrepancy between the predicted bounding box coordinates and the ground truth box coordinates. Further, it quantifies the localization accuracy of the model. (Table. 1)

Val/cls_loss: Class loss represents the error in predicting the object class labels. It captures the accuracy of object classification (Table 1).

Val/dfl_loss: DFL (Dynamic Feature Learning) loss is specific to YOLO models and is used to optimize the feature learning process. It helps in adapting the network to better represent the features of objects of different scales and aspect ratios (Table 1). To ensure robust detection, it is imperative that the metrics 'val/box_loss,' 'val/cls_loss,' and 'val/dfl_loss' are minimized.

Automatically generated graphs depict (Graph of correlogram and graph of Dispersion) the distribution of various diseases among the given images of the dataset. Upon successfully developing the disease detection model, a computation is incorporated to determine the prevalence rate of each disease within a designated land area. For each selected disease in a specific plot of land, a corresponding percentage value of the infection is generated. The outcome is then visually represented on a map that is integrated into a web application. (Fig. 1)

To increase the precision and recall values generated by the model, it is expected to adjust the model architecture and parameters further. Additionally, the model training process will be optimized by applying techniques like gradient clipping and weight decay to prevent overfitting. While the YOLO v8 algorithm demonstrated promising results in disease detection and classification, it is essential to acknowledge its limitations. The algorithm's performance might be affected by variations in lighting conditions, image quality, and the presence of occlusions. Additionally, the dataset used in this study focused on a limited number of common diseases, and further research is needed to expand its applicability to a broader range of diseases in paddy cultivation. Future work should also explore the integration of remote sensing techniques and other advanced machine learning algorithms to enhance disease detection accuracy and scalability.

Table. 1: First ten records of the values obtained for the validation metrics

metrics/ precision(B)	metrics/ recall(B)	metrics/ mAP50(B)	metrics/ mAP50-95(B)	Val /box_loss	Val /cls_loss	Val /dfl_loss
0.07697	0.11361	0.04084	0.01462	2.4456	3.6537	2.0365
0.08058	0.07372	0.0443	0.0133	2.608	3.5708	2.1174
0.06924	0.121	0.03306	0.01135	2.5094	3.8878	2.0698
0.08478	0.10746	0.02731	0.00884	2.606	3.8265	2.2848
0.15414	0.13464	0.06067	0.01815	2.4844	3.5594	2.1708
0.1368	0.1264	0.06717	0.02198	2.3708	3.0507	2.0595
0.20784	0.11725	0.08405	0.03786	2.3703	3.2134	2.0631
0.20251	0.15282	0.08612	0.03362	2.2689	2.9988	1.9588
0.21186	0.09919	0.0967	0.02722	2.2958	2.9934	1.9492

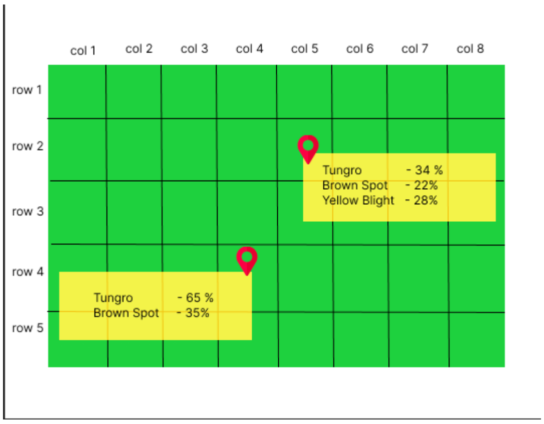

Figure 1: Percentage-wise spread of diseases in a plot of land

4. CONCLUSIONS

In conclusion, the study depicts the effectiveness of the YOLO v8 algorithm in detecting and classifying diseases in paddy cultivation. The algorithm's precision, recall, and F1-scores, coupled with its real-time processing capabilities, make it a valuable tool for farmers in managing and mitigating disease-related challenges. The findings of this research contribute to the advancement of precision agriculture and hold significant potential for improving crop productivity and food security in paddy cultivation.

ACKNOWLEDGMENT

I am grateful to 'Yakkala Govijana Sewa Center', 'Bathlegoda Rice Research Institute', and 'Pelwehera CIC Agro Business Farm' for their support. Special thanks to Mr. Samadhi Rathnayake, my supervisor, and Mrs. Amali Gunasinghe, my co-supervisor, along with everyone who contributed to this project.

REFERENCES

Wang, Y., Wang, Y., & Zhao, J. (2022). MGA-YOLO: A lightweight one-stage network for apple leaf disease detection. *Frontiers in Plant Science, 13*. https://doi.org/10.3389/fpls.2022.927424

Li, Y., Jian-Chun, W., Wu, H., Yu, Y., Sun, H., & Zhang, H. (2022). Detection of powdery mildew on strawberry leaves based on DAC-YOLOv4 model. *Computers and Electronics in Agriculture, 202,* 107418. https://doi.org/10.1016/j.compag.2022.107418

Tang, Z., He, X., Zhou, G., Chen, A., Wang, Y., Li, L., & Hu, Y. (2023). A precise Image-Based tomato leaf disease detection approach using PLPNET. *Plant Phenomics, 5*. https://doi.org/10.34133/plantphenomics.0042

Peng, Y., & Wang, Y. (2022). Leaf disease image retrieval with object detection and deep metric learning. *Frontiers in Plant Science, 13*. https://doi.org/10.3389/fpls.2022.963302

Gong, X., & Zhang, S. (2023). An analysis of plant diseases identification based on deep learning methods. *Plant Pathology Journal (Suwon), 39*(4), 319–334. https://doi.org/10.5423/ppj.oa.02.2023.0034

Kundu, R. (2023, April 20). *YOLO: Algorithm for Object Detection Explained [+Examples].7.* https://www.v7labs.com/blog/yolo-object-detection