AI-Driven Detection of Potato Leaf Diseases and Yield Optimization

Jihene Rezgui*, Ilian Djorf*, Lina Nour Slama*$, Noah Favreau*, Anh-Thi Giang*
Laboratoire **R**echerche **I**nformatique **M**aisonneuve (LRIMa) *,
Montreal, Canada
jrezgui@cmaisonneuve.qc.ca

Abstract, **– In recent times, the usage of machine learning models in agribusiness has significantly expanded intelligent farming techniques. This paper outlines our research towards improving the detection of diseases in potato plants in indoor greenhouse environments using advanced object detection models. Using this set of data, we evaluated and compared three object identification models: Faster R-CNN, YOLOv11, and YOLO-NAS for accurately identifying and categorizing potato leaf diseases. Our approach enables us to train a model on a more realistic set of images, making disease diagnosis more automated and timelier for farmers. With a mAP50 score of 99.5%, our results indicate that the YOLOv11 is the most effective, followed by our Faster R-CNN achieving an average mAP50 of 95.93 %, in detecting diseased leaves.**

Keywords: Machine Learning, Smart Farming, Disease Detection, Greenhouse Monitoring.

I. Introduction

Agriculture is a fundamental pillar of national prosperity. However, it is becoming an inconsistent field. One of the main challenges farmers must face includes plant diseases, which propagate to the production cycle and supply chain. For instance, the late blight pathogen, responsible for the Irish Potato Famine (1845–1852), continues to devastate potato crops. While global trade and the movement of agricultural goods contribute to its spread, it thrives in wet and humid conditions—factors exacerbated by climate change. This disease has caused significant economic losses, forcing farmers to adopt more rigorous monitoring and control measures.

Classic methods, like manually removing affected leaves, are not viable solutions. Also, the increased use of chemical pesticides in response to rising pest and disease pressure can have negative impacts on human health and the environment. In this era of smart agriculture, where precision farming techniques enhance productivity and sustainability, there is a growing need for automated, reliable, and efficient solutions to identify and manage plant diseases in real time.

To address this issue, we propose to automate the identification of potato plant diseases based on plant images. Early detection and classification are essential for effective disease management and crop loss prevention. This project uses deep learning techniques, including YOLO (You Only Look Once) and region-based convolutional neural network models.

This paper builds on the 2024 research *Optimizing Disease Detection Models in School Greenhouses: An AI and IoT-Based Approach for Smart Agriculture* [4] which mainly focused on training and comparing object detection models tasked with identifying diseases for tomato plants. This paper enhances the dataset with new images, benchmarks additional object detection models, and introduces disease classification capabilities. We focus on three state-of-the-art object detection models — Faster R-CNN, YOLOv11, and YOLO NAS — each offering advantages for detecting plant diseases. Our main objective is to identify the most efficient model for this application, considering factors such as detection accuracy, training speed, prediction speed, and the model's ability to handle the complex and diverse conditions of a greenhouse.

With the right image detection model, the industry can enhance the resilience of potato production systems, reduce environmental impacts, and ensure the long-term sustainability of potato farming.

Our paper can be summarized as follows: **(1)** We selected over 2,000 images from the PlantVillage dataset [1] and categorized them using Roboflow; **(2)** We augmented the PlantVillage dataset using various transformations; **(3)** We trained and benchmarked three object detection models on the PlantVillage dataset; **(4)** We trained an object detection model on the Leaf Detection dataset [2]; and **(5)** We discussed the results of the three models and the two-step detection process.

Outline: Section II gives a brief overview of similar research done in the field of potato leaf disease detection. Section III presents the transformation and preparation of the data used to train the models. Section IV explains our choices of object detection models. Section V shows our results. Finally, section VI concludes the paper.

II. Related work

Potato disease detection has been an important area of agricultural research. Many studies have used Convolutional Neural Networks (CNNs) to classify diseases in potato plants by analyzing leaf images [15], identifying healthy plants and detecting common diseases like Early Blight and Late Blight.[5] Our work advances this research by integrating several improvements: we implemented transfer learning using pre-trained models like ResNet and adopted the R CNN architecture for better feature extraction. To facilitate real-world use, we developed a web and mobile application and are working on installing a 360-degree rotation camera for continuous, high-quality image capture, which further enhances model accuracy and applicability. In terms of model optimization, we transitioned to YOLO-based architectures (YOLOv11 and YOLO-NAS), which offer faster inference by using single-stage models and reducing the need for additional layers compared to traditional CNNs. Furthermore, there exists another study providing a thorough analysis of the differences between of prominent convolutional neural network architectures [13], namely VGG16 [14], VGG19, ResNet50, ResNet152, and InceptionV3 [6], for classifying distinct disease classes in potato plants as mentioned above, comprising 2152 images which was revaluated in our work using an enhanced dataset of 7127 images. This larger dataset, which more closely simulates real-world conditions, improves the model's accuracy and generalization [15].

To the best of our knowledge, there are no papers using versions of YOLO more recent than YOLOv11 for Potato leaf disease detection. Therefore, we are among the first to experiment with YOLOv11 and YOLO-NAS in this context.

III. Dataset and models

This section highlights the key differences from our previous work [3-4], and the improvements made. It is organized into two main subsections. The dataset preparation section covers the methods and processes used to gather and preprocess the data, which is later used to train the models. The models used section outlines the models used in our approach, along with modifications made compared to our previous work.

A. Dataset

Similar to our work last year, we utilized images from a Kaggle repository called PlantVillage to detect plant diseases. We selected three directories from this dataset, which contain images of individual leaves — both healthy and diseased [12]. This allows our models to classify the leaf's condition based on visual information contained within the image.

However, a major limitation of the PlantVillage dataset is that it contains images of isolated leaves taken in controlled conditions. Leaves in actual greenhouse environments contain clusters of overlapping and folded leaves, making disease detection more complex. As opposed to our work in [4], we didn't have access to any actual greenhouse environment to take pictures in, which would've been useful for detection in greenhouses.

We decided to use a second dataset called Leaf Detection, which is designed to find individual leaves. We trained a separate model to detect leaves, which would help by cropping out each leaf in the greenhouse before using the disease detection model.

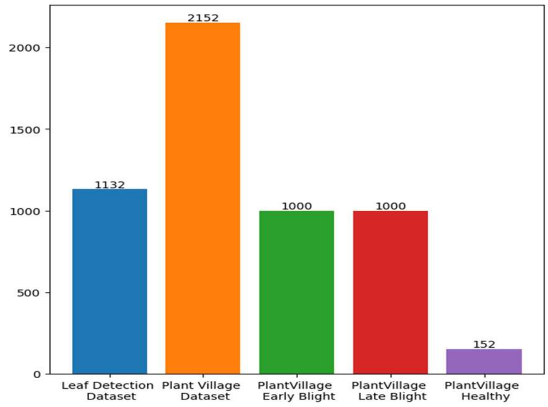

Fig.1. **Distribution of Images Across Different Classes and Datasets**

As we can see in Fig.1, the Leaf Detection dataset has fewer images than the PlantVillage dataset, but we still used it because it is intended for a separate model. For each image in the PlantVillage dataset, we generate 1-3 images to augment the data. A random rotation is applied to these images. The final dataset used in training was 7127 images. This dataset was split, 80% for the model training, 10% for validation and 10% for testing. **(Fig 2.)** The same split was applied on the Leaf Detection dataset. We did not generate new images for the Leaf Detection dataset.

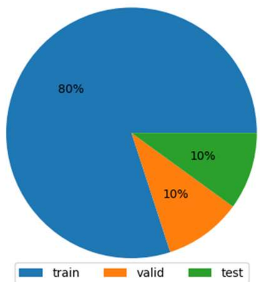

Fig.2. **PlantVillage and Leaf Detection dataset distribution into Train, Valid and Test splits.**

B. The Models

For the object detection models used in this research, we implemented 3 different models to compare them and determine which one would be more accurate in a greenhouse context. They will be evaluated on their ability to accurately detect the disease affecting the potato plant, their ability to differentiate different plants in different states and the detection speed of each model. These criteria will be important in an actual greenhouse context, since the models will have to be accurate, fast and able to separate different leaves from each other.

The first model will be the same as in [4], a Faster R-CNN model with a ResNet-50 backbone combined with a Feature Pyramid Network (FPN). We will later compare this model with others based on the previously mentioned criteria. The second model will be YOLOv11, an upgraded version of YOLOv10 [7]. This newer version shows improvements in detecting smaller objects due to key advancements, such as dual label assignments [10] that boost accuracy and enhanced down sampling techniques that preserve fine details. The third model in our study is YOLO-NAS, a state-of-the-art object detection model that balances speed and accuracy using a neural architecture search (NAS) approach. YOLO-NAS is designed to optimize performance across various hardware configurations while maintaining high precision, making it a strong candidate for detecting the fine details of potato leaves essential for identifying diseases.

IV. Our Greenhouse AI models

A. Faster R-CNN

The first model we trained was the Faster R-CNN (Region-based Convolutional Neural Network) with a ResNet-50 version 2 backbone combined with a Feature Pyramid Network (FPN) for better precision. This was the best model in our previous work, and we will compare it with other models in this work, namely the YOLOv11 and YOLO-NAS models. These two models are more recent, and we'd like to see how the Faster R-CNN model can perform against more modern object detection models.

B. YOLOv11

The second model we're exploring is YOLO (You Only Look Once). In our latest research paper, we utilized YOLOv10, but since then, Ultralytics has released a new iteration in the YOLO series: YOLOv11. Built on the YOLOv8 codebase, this version introduces several architectural modifications that elevate its performance. **Fig.3** shows the architecture for the YOLOv11 model.

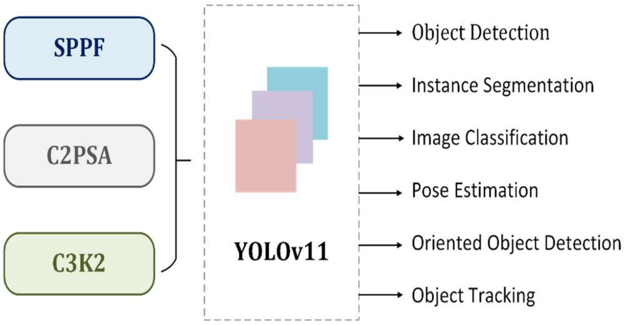

Fig.3. **Key architectural modules in YOLOv11. [6]**

YOLOv11 brings a handful of key improvements that make it stand out compared to its predecessors. For example, it features enhanced attention mechanisms with the Cross Stage Partial with Spatial Attention (C2PSA), which helps it focus better on critical parts of an image. It also incorporates new architectures like the Spatial Pyramid Pooling - Fast (SPPF) and C3K2 block, boosting its ability to process visual data efficiently. Another significant upgrade is its streamlined parameter count. On the COCO dataset, YOLOv11 achieves a 22% reduction in parameters compared to YOLOv8, making it notably leaner. Compared to YOLOv10, which we previously used, these optimizations lead to faster performance and greater efficiency without compromising accuracy.

C. YOLO-NAS

The third model we chose is YOLO-NAS (You Only Look Once -Neural Architecture Search) is an advanced object detection model that improves upon previous YOLO versions by using Neural Architecture Search to automatically optimize its structure for better accuracy and efficiency. It maintains YOLO's core philosophy of single-shot object detection while introducing enhancements in feature extraction, bounding box regression, and classification. The detection process in YOLO-NAS follows a single-stage pipeline. When an image of a potato plant is fed into YOLO-NAS, it is first resized to a fixed input size (1280x1280 pixels) and normalized to ensure consistent scale across all images. The pixel values are transformed and standardized to improve network stability and convergence.

The backbone network of YOLO-NAS consists of multiple convolutional layers designed to extract hierarchical features from the image. The first few layers are responsible for detecting basic patterns such as edges and corners which are crucial for distinguishing between different regions of the image. As the data flows through deeper layers, more complex features are captured, enabling the model to recognize fine-grained structures on potato leaves, such as small lesions, dark spots, or discoloration, all of which are critical indicators of diseases like early blight and late blight. These extracted features are then passed through the detection head, which refines the bounding box coordinates and classifies each detected region as either healthy, early blight, or late blight.

V. Results

The following subsections will present the results of our research for the 3 models.

A. Faster R-CNN

The model was trained over 27 epochs, where we then stopped it due to potential overfitting.

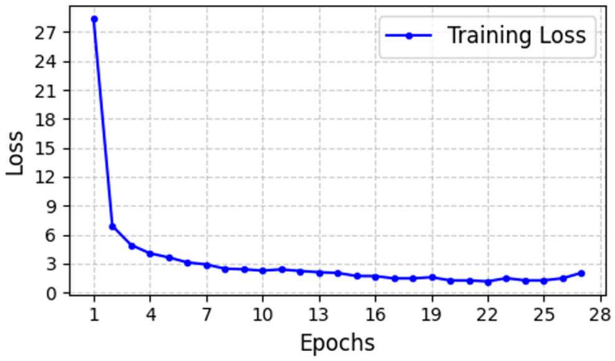

Fig 4. **Training loss of the Faster R-CNN model over training epochs.**

As shown in **Fig.4**, the Faster R-CNN model is displaying an efficient convergence, with the losses quickly stabilizing at a value near 1.25. The training was halted early after observing fluctuations in loss values during epochs 19 and 27. This was a sign of potential overfitting. The final average mAP score for the Faster R-CNN model was 95.86% and the average mAP50 score was 95.93%. The training time for all 27 epochs was about 14 hours. These findings will later help us compare the Faster R-CNN model to the YOLOv11 and YOLO-NAS models.

B. YOLOv11

For the model using YOLOv11, the average mAP50 score across all the epochs came to 99.5%. Fig. 5 illustrates the training and validation performance for box loss, classification loss (cls loss), distribution focal loss (dfl loss), mAP50, mAP50-95, precision, and recall.

For this model, the time of training is about 1 hour and 27 minutes and for prediction of one image it's between 2 and 5 milliseconds on average.

Fig. 6 illustrates the Precision-Confidence curve over our validating dataset. With such a high curve and a very high mAP50, it is likely that the model experienced overfitting after 50 epochs of training.

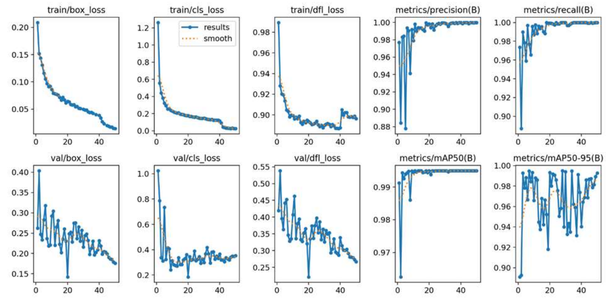

Fig.5 **Training Metrics throughout the epochs for YOLOv11.**

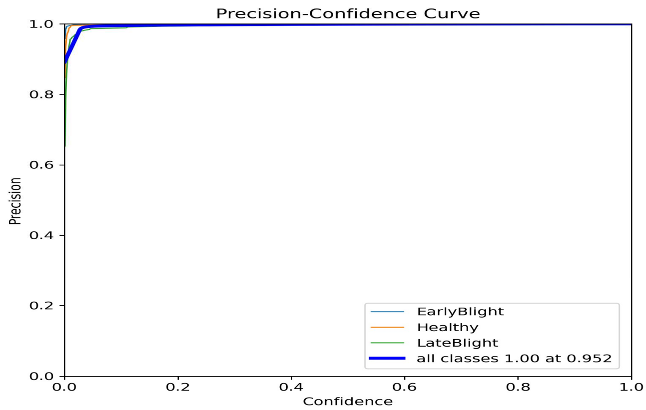

Fig. 6. **Precision-Confidence curve of YOLOv11 over validating dataset.**

C- YOLO-NAS

The training of the YOLO-based object detection model for the classification and localization of potato leaf conditions was carried out over a total of 4 hours and 30 minutes, utilizing a dataset composed of 7127 annotated images in COCO/14 format. The model exhibited rapid convergence across all key loss components, indicating effective learning dynamics and optimization behavior.

The classification loss decreased markedly from an initial value of approximately 3.0 to a stabilized average of ~0.15, indicating the model's effective learning of inter-class discriminative features. Concurrently, the bounding box regression loss declined sharply from 0.30 to an average of 0.01, reflecting high localization accuracy in object bounding box predictions. The object loss, responsible for discerning foreground from background regions, converged from ~0.55 to ~0.34, denoting improved confidence calibration in object detection.

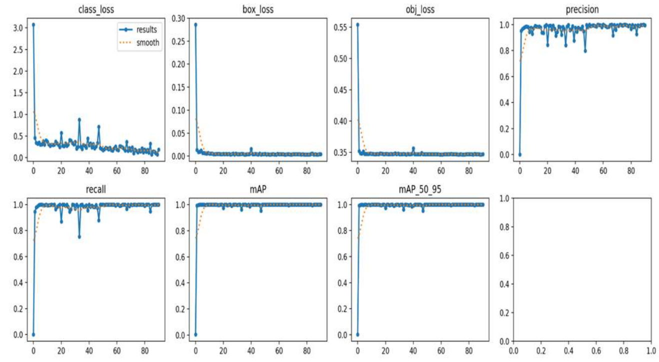

Fig. 7**. Training metrics throughout the epochs of YOLO-NAS.**

In terms of detection performance, as shown in **Fig.7**, the model achieved high levels of precision and recall, averaging 97% and 94%, respectively. These results reflect a low incidence of both false positives and false negatives across the validation set. The mean Average Precision (mAP50) converged to near-optimal values of 98%, while the mAP50-90, a more stringent and comprehensive evaluation metric, remained consistently high with an average of 94.9% . This performance underscores the model's robustness across varying Intersection over Union (IoU) thresholds and its ability to generalize well across diverse annotation overlaps.

Overall, the loss function trajectories and performance metrics collectively indicate that the model converged efficiently, learned robust spatial and semantic representations, and achieved high generalization capacity on the task of multi-class leaf disease detection in potato plants.

D. **Comparison of all the models**

1. Metrics Comparison

Across all three models, YOLOv11 has a map50 score of 99.5%, Faster R-CNN has 95.96%, and YOLO-NAS has 94.9%. As we can see, although Faster R-CNN takes the longest to train, it is overall the highest performing model that we trained with a map50 score of 92.13% and prediction time varying between 250 and 350 ms.

2. Solutions

When testing our models, we encountered an issue similar to the one in our previous work. All three models worked well on individual leaves but showed more difficulties when processing clusters of leaves due to the nature of the PlantVillage dataset. The first option we had to solve this was to take images from an actual greenhouse environment, however this solution would've been a tedious work of capturing data and annotating, which would've taken a lot more time than our second option.

The second option was to create a leaf detection model, this model would be able to accurately detect clear individual leaves among clusters of leaves, which we could then crop out of an image and give to our disease detection models. Seeing as YOLOv11 was our most promising model when making the disease detections models, we created another YOLOv11 model for leaf detection which we trained on the Leaf Detection dataset; a fully annotated dataset available for free use on Kaggle.

By using these two YOLOv11 models in tandem, we were able to get much more accurate results on leaf clusters, as observable in **Fig.8**

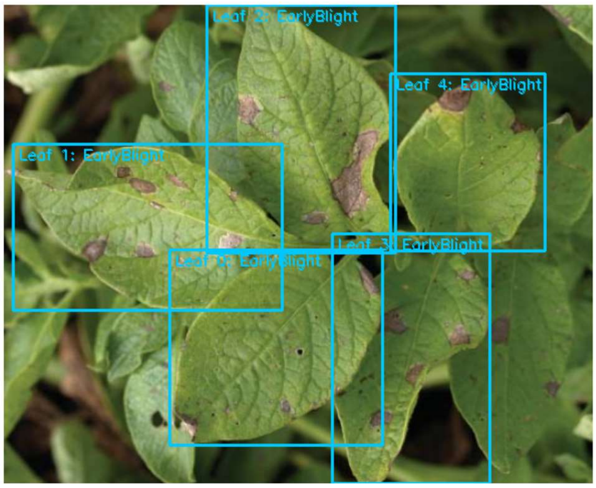

Fig.8. **Result of a prediction done by using both YOLOv11 models on a plant affected by early blight [8]**

3. Discussion

The overall accuracy we managed to achieve during this study surpasses what we had achieved in previous work. When compared to our work last year [4], our approach has better accuracy in classifying the disease and detecting individual leaf clusters. Before, our best model was Faster R-CNN since it was able to somewhat detect individual leaves among clusters, however with our new 2 step approach the YOLOv11 model is far more suited to the task, itself achieving an accuracy of 99.5% instead of 92.13%.

While our new two-step approach shows promise, there are still several ways to enhance our solution. Despite its accuracy, the two-step detection process significantly slows down performance, as it requires running two separate models. To address this, we propose the following improvements:

(1) Switching to Image Classification Models – In the current two-step approach, we use two object detection models. However, the model responsible for disease detection does not perform actual object detection, as it was trained on the PlantVillage dataset. Since object detection models are generally slower than classification models, we could replace the disease detection model with a classification-based architecture such as EfficientNet [11], GhostNet, or a ResNet variant. This would improve the efficiency of the second detection step.

(2) Enhancing the Leaf Detection Model – To improve both speed and accuracy, future work could explore more recent object detection architectures. Notably, Ultralytics released YOLOv12 in February 2025, which shows promise as an even faster and more efficient detection model.

VI. CONCLUSION

This study highlights the potential of using models for disease detection in greenhouses. The approach we used in this paper for detecting diseases in potato plants can be easily adapted to other types of plants, easily, as the PlantVillage dataset is now a viable option, even for leaf clusters. The leaf detection model can be used for any plant, making it compatible with other disease detection models.

In the future, we plan to implement our approach for a real-time monitoring system within the greenhouse. We will also focus on developing classification models and evaluating their performance in conjunction with the leaf detection model.

REFERENCES

[1] PlantVillage Kaggle repository https://www.kaggle.com/datasets/abdallahalidev/plantvillagedataset [last visited 10 January 2025]

[2] Leaf Detection Kaggle repository https://www.kaggle.com/datasets/alexo98/leaf-detection?resource=download [last visited 8 January 2025]

[3] Y. Lakhdari, E. Soldevila and J. Rezgui, "Detection of plant diseases in an industrial greenhouse: Development, Validation & Exploitation", accepted IEEE ISNCC 2023, Qatar.

[4] J. Rezgui, E. Soldevila, L. Lafleur-Hébert, Y. Asmour, M. Tardy, "Optimizing Disease Detection Models in School Greenhouses: An AI and IoT-Based Approach for Smart Agriculture", accepted IEEE ICAA 2024

[5] S. Ren, K. He, R. Girshick, J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," Advances in Neural Information Processing Systems 28 (NeurIPS 2015).

[6] Journal, I. RJ ET. "IRJET- POTATO LEAF DISEASE DETECTION USING INCEPTION V3." IRJET, 2020.Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024).

[7] Wharton, P., & Kirk, W. (2007). *Early Blight*. Extension Bulletin E-2991. Michigan State University. https://www.canr.msu.edu/psbp/resources/early-blight-bulletin.pdf?utm_source=chatgpt.com

[8] Ghosh, A. (2024, October 8). *YOLO11: Redefining real-time object detection*. LearnOpenCV. https://learnopencv.com/yolo11/

[9] Mukherjee, S. (2024, October 18). *What's new in YOLOv11: Transforming object detection once again, part 1*. DigitalOcean. https://www.digitalocean.com/community/tutorials/what-is-new-with-yolo

[10] Afzaal, H., Farooque, A. A., Schumann, A. W., Hussain, N., McKenzie-Gopsill, A., Esau, T., Abbas, F., & Acharya, B. (2021). Detection of a potato disease (early blight) using artificial intelligence. Qatar University: .https://www.mdpi.com/973322

[11] Chowdhury, M. E. H., Rahman, T., Khandakar, A., Ayari, M. A., Khan, A. U., Khan, M. S., Al-Emadi, N., Reaz, M. B. I., Islam, M. T., & Ali, S. H. M. (2021). Automatic and reliable leaf disease detection using deep learning techniques. AgriEngineering, 3(2), 20. Qatar University: https://www.mdpi.com/2624-7402/3/2/20.

[12] Gavhale, K. R., & Gawande, U. (2014). An overview of the research on plant leaves disease detection using image processing techniques. IOSR Journal of Computer Engineering, 16(1), Ver. V, 10–16. Yeshwantrao Chavan College of Engineering: www.iosrjournals.org.

[13] Alhammad, S. M., Khafaga, D. S., El-hady, W. M., Samy, F. M., & Hosny, K. M. (2025). Deep learning and explainable AI for classification of potato leaf diseases.University, Riyadh, Saudi Arabia :https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2024.1449329/full

[14] Lu, J., Tan, L., & Jiang, H. (2021). Review on convolutional neural network (CNN) applied to plant leaf disease classification. Xihua University. https://www.mdpi.com/2077-0472/11/8/707.