CARPATHIAN J. MATH.
Volume **41** (2025), No. 4,
Pages 1091 - 1100

Online version at `https://www.carpathian.cunbm.utcluj.ro/`
ISSN 1584-2851 (Print edition); ISSN 1843-4401 (Electronic)
DOI: https://doi.org/10.37193/CJM.2025.04.16

Mathematical Simulation of Rice Leaf Lesions for Machine Learning–Based Disease Diagnosis

MARA HAJDU-MACELARU[1]

ABSTRACT. This article presents a research based on mathematical modeling, visual simulation of leaf spots caused by plant diseases. Using analytical functions (simple Gaussian distributions, elliptical and jagged edges), we simulate the spots caused by two diseases quite common in rice crops. Geometric formulas (area, perimeter, circularity) are used for interpretation, while the shape and color features of the spots are analyzed to generate a synthetic dataset. To demonstrate the applicability of the method, we show how, for two rice diseases, adding mathematically generated images to the training images leads to increased accuracy of machine learning models for automated plant disease diagnosis. Thus, the results confirm that synthetic image generation can effectively complement real datasets, making disease detection more robust.

1. INTRODUCTION

Early detection and prevention of plant diseases is a major achievement for agriculture, and if this process could be automated, farmers would not have to use so many pesticides to prevent certain diseases. In the case of plant disease detection from images, a major limitation is the lack of datasets containing diverse examples of diseased plants and leaves, which makes machine learning algorithms less efficient for plant disease detection [15]. This article demonstrates how synthetic images (for example, spots generated by mathematical models) can be created to enrich the training phase, thereby balancing classes [2, 7, 9, 8]. As a result, the model is exposed to a wider variety of shapes, sizes, and colors, making it more robust. Mathematical modeling of biological phenomena offers a rigorous framework for simulating behaviors observed in nature. In the context of smart agriculture, visually simulating foliar diseases provides a valuable resource for training and testing neural networks when large volumes of real data are not available [5, 14, 22].

This paper proposes a methodology for generating synthetic images of rice leaves with spots simulating the symptoms of two common diseases (Brown Spot and Bacterial Leaf Blight), where the spots are mathematically generated, and for testing their effectiveness in a machine learning algorithm for disease detection. Images of the two diseases, Brown Spot and Bacterial Leaf Blight, from the rice leaf dataset were used to observe the disease and to reproduce the two diseases as accurately as possible using mathematical modeling. The work brings some new elements compared to what already exists in the field. Instead of using complex neural networks, as many recent studies do, we propose a method based on simple and controllable mathematical models. These models can be modified by clear parameters (for example: the shape can be round or elongated, the size can vary within a range, the color can be more intense or paler). At the same time, this mathematical approach does not need a large volume of real data to start, which makes it very useful in contexts where image collection is difficult or even expensive [16, 1, 6]. Compared to other studies, the images generated by this method are not just visual examples for theoretical

Received: 17.04.2025. In revised form: 30.07.2025. Accepted: 20.09.2025
2020 *Mathematics Subject Classification.* 65D18, 92C60.
Key words and phrases. *mathematical modeling, rice diseases, Synthetic image generation, Image simulation, machine learning.*

demonstrations, but their use for training and testing machine learning algorithms is even exemplified in this article. Most of the time, researchers and practitioners are faced with the lack of sufficiently large and varied data sets, which greatly limits the performance of automatic plant disease recognition models. Thus, using these synthetic images, we can complement real image collections and create more balanced datasets with a greater diversity of lesion shapes, colors, and distributions [4, 10, 17]. This has a direct impact on the efficiency of machine learning algorithms, which become better trained and more capable of recognizing diseases in real situations. The proposed method offers a practical solution for increasing the accuracy and robustness of detection algorithms.

2. Mathematical Simulation of Foliar Lesions

Using mathematical modeling we can describe and simulate some biological phenomena through well-defined functions, having control over the shape, size, color and variability of lesions [3, 12, 13]. In this study, we proposed a visual simulation framework for two of the most common rice diseases: Brown Spot and Bacterial Leaf Blight. Leaf spots are visible lesions that occur mainly due to infections. If we were to make a phytopathological description of a leaf spot, this would be a delimited area, usually with a different color from the rest of the leaf, which indicates the presence of a disease. These spots can appear in various shapes: circular, elliptical, irregular, with clear or diffuse edges and can vary in color from brown and yellow to black or reddish.

Through mathematical modeling of leaf spots, we aim to simulate the real characteristics of these spots or lesions. This approach is particularly valuable in the context of artificial intelligence (AI), as it allows the generation of labeled synthetic data, which will be useful for training learning models, especially in cases where real data are insufficient. The modeling is performed through parameterizable analytical functions, which define both the intensity distribution within the spot and its geometry. Among the most commonly used functions are Gaussian distributions (circular or elliptical) and oscillatory functions, which can be combined to simulate irregular edges or diffusion effects.

The mathematical modeling was carried out by developing an application written in Python, run in Google Colab Pro, with which the leaf spots that simulate the two diseases was applied on images of healthy rice leaves. For each disease studied, we first analyzed real photos with specific symptoms from the same dataset (for example, Brown Spot and Bacterial Leaf Blight), to better understand the shape, color and spread of the lesions so that we could simulate these spots as best as possible.

For each type of disease, a separate application was implemented, in which the mathematical models were developed in order to simulate the appearance of the spots as best as possible. The generated spots were then superimposed on the healthy leaves images through a linear interpolation technique (alpha blending), which allowed their natural integration into the texture. In this way, we obtained synthetic images that simulate real diseases and that can be used to extend and balance the initial dataset.

2.1. **Mathematical simulation of Bacterial Leaf Blight lesions.**

After studying several real images with Bacterial Leaf Blight disease - see Figure 1, the Bacterial Leaf Blight lesions were modeled as elongated stripes that appear on the main axis of the leaf.

Leaf Axis Estimation was estimated using Principal Component Analysis (PCA) by applying the leaf mask, thus ensuring the alignment of the lesions with the biological structure of the plant:

$$\vec{v}_1 = \arg \max_{\|\vec{v}\|=1} \mathrm{Var}(\vec{x} \cdot \vec{v})$$

where:

- $\vec{x}$ represents the coordinates of the leaf pixels,
- $\vec{v}_1$ is the principal direction of maximum variance.

FIGURE 1. Real images with leaves affected by Bacterial Leaf Blight Disease

The basic shape of the lesion was described using a two-dimensional anisotropic Gaussian distribution:

$$G(x,y) = \exp\left(-\frac{(x-c_x)^2}{2\sigma_x^2} - \frac{(y-c_y)^2}{2\sigma_y^2}\right), \tag{2.1}$$

where:

- (c_x, c_y) are the lesion center coordinates,
- $\sigma_x \gg \sigma_y$, resulting in an elongated structure.

To simulate the fact that the lesion narrows at the extremities, a taper function was applied along the longitudinal axis:

$$T(x) = \left(1 - \frac{|x-c_x|}{L/2}\right)^p, \tag{2.2}$$

where:

- L is the total lesion length,
- p controls the degree of tapering.

A transversal Gaussian profile was introduced to simulate the characteristic necrotic ridge observed in real blight lesions:

$$R(y) = \exp\left(-\frac{(y-c_y)^2}{2\sigma_r^2}\right). \tag{2.3}$$

The lesion shape was perturbed by smoothed random noise in order to increase variability and realism:

$$M(x,y) = G(x,y) \cdot T(x) \cdot (1 + \beta \cdot N(x,y)), \tag{2.4}$$

where:

- $N(x,y)$ is a Gaussian-filtered noise field,
- β controls the irregularity level of lesion contours.

As it can be noticed in Figure 2, in order to simulate better the biological progression of tissue damage, a transversal color gradient was applied, with a pale center and yellow-brown edges, combined with a darker rim.

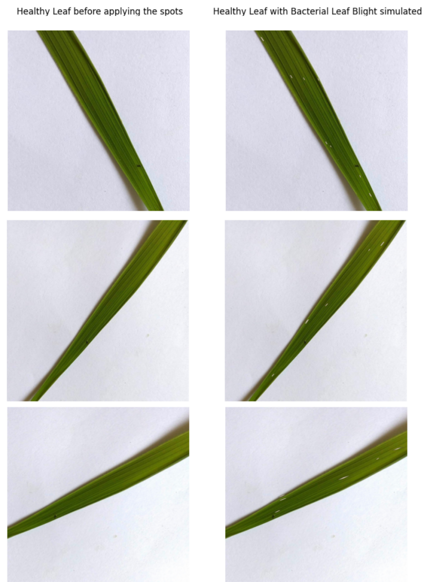


FIGURE 2. Before-and-after samples: healthy leaves versus leaves with simulated disease spots

2.2. **Mathematical simulation of Brown Spot lesions.** After studying several real images with Brown Spot disease - see Figure 3,the *brown spot* lesions were modeled as circular brown spots with diffuse edges and irregular contours.

FIGURE 3. Real images with leaves affected by Brown Spot Disease

A two-dimensional isotropic Gaussian distribution was used to describe the basic shape of each spot:

$$G(x,y) = \exp\left(-\frac{(x-c_x)^2 + (y-c_y)^2}{2\sigma^2}\right), \tag{2.5}$$

where (c_x, c_y)represents the center of the spot and σ controls the effective radius. In order to simulate the irregular contours, so that the spot simulates the spots observed in reality as closely as possible, the Gaussian distribution was perturbed with smooth noise, obtained by Gaussian filtering of a random field:

$$M(x,y) = G(x,y) \cdot (1 + \alpha \cdot N(x,y)), \tag{2.6}$$

where $N(x,y)$ is the filtered noise and α is an irregularity coefficient.

The chromatic effect was accomplished using a radial gradient:

$$C(r) = (1-r) \cdot C_{\text{center}} + r \cdot C_{\text{edge}}, \tag{2.7}$$

where

$$r = \frac{(x-c_x)^2 + (y-c_y)^2}{R}$$

is the normalized distance from the center, C_{center} is the dark brown color in the center, and C_{edge} is the diffuse yellowish-brown color in the periphery.

The comparison between the healthy leaf and the leaf with simulated brown spot lesions is shown in Figure 4,and it can be noticed that by applying a radial color gradient with a dark brown center and diffuse yellow-brown edges, we achieved to simulate the characteristic appearance of the disease.

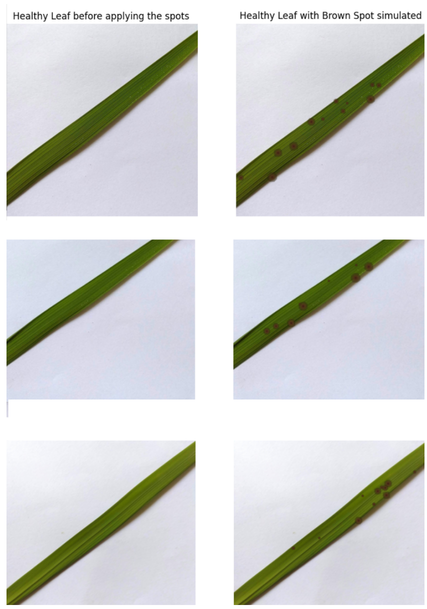


FIGURE 4. Before-and-after samples: healthy leaves versus leaves with simulated disease spots

2.3. **Common Aspects and Integration on Healthy Leaves.** In both cases of simulations, we generated the irregular lesion contours using smoothed Gaussian noise:

$$N\_s(x,y) = (NG\_\sigma)(x,y), \tag{2.8}$$

where:

- N is a uniformly random noise field,
- G_σ is a Gaussian smoothing filter,

- $*$ denotes the convolution operation.

The integration of lesions into healthy leaves was made through linear interpolation (alpha blending):

$$I\_out(x,y) = \big(1 - \alpha(x,y)\big) \cdot I\_leaf(x,y) + \alpha(x,y) \cdot I\_lesion(x,y), \tag{2.9}$$

where:

- $I_{leaf}(x,y)$ is the original healthy leaf image,
- $I_{lesion}(x,y)$ is the simulated lesion texture,
- $\alpha(x,y)$ is the transparency mask that controls blending.

This approach to mathematical modeling of leaf lesions combines deterministic functions (Gaussian, tapering, chromatic gradient) with stochastic processes (noise, parameter randomization) and geometric estimates (PCA for the leaf axis). The result is realistic synthetic images, useful for augmenting datasets for machine learning algorithms in precision agriculture.

3. Experimental Setup

If in the first part we generated images simulating the two diseases specific to rice leaves (Bacterial Leaf Blight and Brown Spot), the next step was to evaluate whether this synthetic data can contribute to increasing the accuracy in automatic disease diagnosis based on images. We want to verify whether in situations where there are not enough real images in the training phase or when the accuracy of a model needs to be improved, whether the generated images simulating certain diseases can be used to augment the dataset and improve the performance of classification algorithms.

For this experiment, we implemented, in Google Colab Pro in the Python language, a classification algorithm based on MobileNetV2 as a feature extractor and on an SVM (Support Vector Machine) for decision. We designed several experimental scenarios to evaluate the impact of using the generated images:

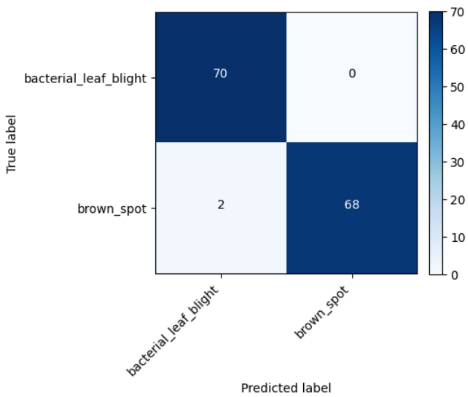
(1) Scenario 1 – Model with original images only. Only real photos of rice leaves affected by the two diseases were used, 350 images for each class. The dataset was divided into 80% for training and 20% for testing.
(2) Scenario 2 – Combined model - original and generated image. Starting from scenario 1, we added 100 images generated for each class to the training set, keeping the same testing set (only real images).
(3) Scenario 3 – Adaptive model. In situations where the model performance from Scenario 1 was reduced for a certain class, we completed the training with 100 additional images generated only for that class. In this case we wanted to see if we can use generated images to increase accuracy where needed.

The performance evaluation was performed using standard metrics (accuracy, precision, recall, F1-score), as well as by analyzing confusion matrices. This experimental framework will give us a clear picture of whether simulated images can be used in the training phase to increase accuracy

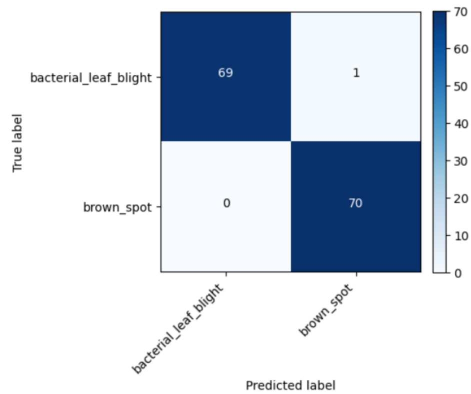
4. Results

In all the experimental scenarios, the same classification algorithm was used, based on MobileNetV2 as a feature extractor and SVM as a decision method. The test set was kept identical, consisting exclusively of original images, and the standard augmentation procedures applied to the images remained the same. The only difference between the scenarios was exclusively in the way the training set was constituted, by introducing or not the generated images.

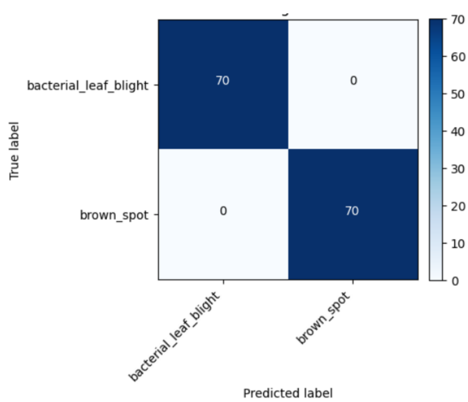
In evaluating the results of a classification algorithm based on machine learning, the confusion matrix shows us how many examples were classified correctly and how many were classified incorrectly. The metrics used to evaluate the results are accuracy (the total proportion of correct predictions), precision (how good the predictions are for a class), recall (how well the real cases were identified) and F1-score (the balance between precision and recall). Together, these metrics provide a complete picture of the model's performance, both globally and on each class.



(A) Confusion matrix for Scenario 1



(B) Confusion matrix for Scenario 2



(C) Confusion matrix for Scenario 3

FIGURE 5. Confusion matrices for the three experimental scenarios.

In Scenario 1, as it can be seen from Table 1 with the metrics results, the model achieved an overall accuracy of 98.6%. The results show that the bacterial leaf blight class was classified perfectly, but for the brown spot class there were two classification errors, which led to a slightly lower recall (97.1%). The confusion matrix presented in Figure 5 highlights these errors.

TABLE 1. Classification results for Scenario 1

Class	Precision	Recall	F1-score	Accuracy
Bacterial_leaf_blight	97.22	100,00	98.59	98.57
Brown_spot	100.00	97.14	98.55	98.57
Overall average Values	98.61	98.57	98.57	98.57

In Scenario 2, by adding 100 images generated for each class to the training set, the overall accuracy increased to 99.3%. As can be seen in Table 2, the Brown spot class was classified perfectly, and for Bacterial leaf blight there was only one classification error (recall = 98.6%). The confusion matrix in Figure 5 shows the almost perfect distribution of predictions. This result confirms that using generated images in the training phase brings an additional performance improvement. [20, 19, 18]

TABLE 2. Classification results for Scenario 2

Class	Precision	Recall	F1-score	Accuracy
Bacterial_leaf_blight	100.00	98.57	99.28	99.29
Brown_spot	98.59	100.00	99.29	99.29
Overall average Values	99.30	99.29	99.29	99.29

In Scenario 3, the training set was completed only with images generated for the brown spot class, where the performance was weaker in Scenario 1. As can be seen in Table 3, the result was a perfect accuracy of 100%, with all test images being correctly classified. The confusion matrix in Figure 5 shows a complete diagonal, without errors. This demonstrates that the generated images used only for the classes to be improved can completely eliminate classification errors.

TABLE 3. Classification results for Scenario 3

Class	Precision	Recall	F1-score	Accuracy
Bacterial_leaf_blight	100.00	100.00	100.00	100.00
Brown_spot	100.00	100.00	100.00	100.00
Overall average Values	100.00	100.00	100.00	100.00

As can be seen in Figure 6, which shows a comparison of the performance in the three scenarios tested through a graph, it is observed that in Scenario 2 all metrics increase slightly compared to Scenario 1, but in Scenario 3 all metrics reach their maximum value, confirming that an adaptive strategy of completing the training set with generated images is the most efficient.

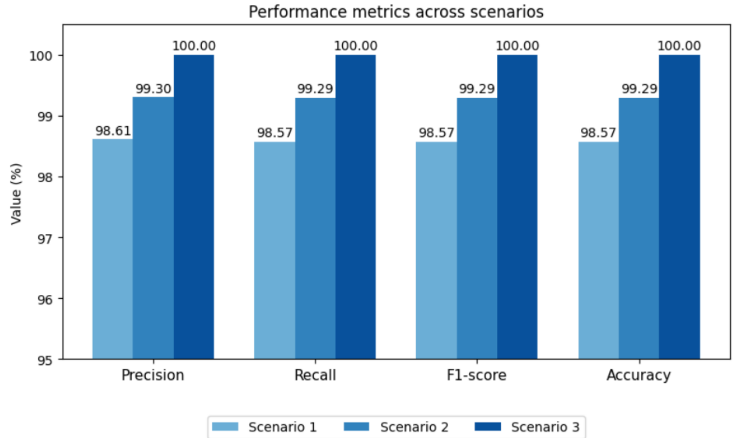


FIGURE 6. Performance comparison across experimental scenarios

5. CONCLUSION

The presented method shows that we can use mathematical models to generate images that realistically imitate the plant diseases. Using mathematical models we can reproduce important features such as the shape, size, color and distribution of lesions, even if they do not faithfully copy every detail from reality. We can thus obtain a framework that provides a flexible tool that can be adapted for multiple types of diseases and that can be used to obtain better results in artificial intelligence models for detecting plant diseases from images. The generated images can be added to real datasets and used to train disease recognition algorithms, thus increasing their accuracy and efficiency. This is important in situations where real images are often difficult to obtain [11, 21, 23].

As a final conclusion, the proposed method not only provides a realistic way to simulate plant diseases, but also represents a practical aid for the development of intelligent applications, with direct impact in agriculture and research.

ACKNOWLEDGMENTS

This work was supported by the project ”New Tool Based on Intelligent Algorithms and Machine Learning for Detecting Plant Diseases from Images”, funded through the National Research Grant Competition – GNaC ARUT 2023.

REFERENCES

[1] Alhammad, S. M.; et al. Deep learning and explainable AI for classification of plant leaf diseases: a review. *Front. Artif. Intell.* **8** (2025), 1449329.

[2] Antwi, K.; et al. On the application of image augmentation for plant disease detection: a systematic review. *Plant Communications* **4** (2024), 1001953.

[3] Aslam, A.; et al. Multi-convolutional neural networks for cotton disease classification and yield prediction. *Sci. Rep.* **15** (2025), 12112145.

[4] Klein, J.; Waller, R.; Pirk, S.; Pałubicki, W.; Tester, M.; Dominik L. Michels. Synthetic data at scale: a development model to efficiently leverage machine learning in agriculture. *Front. Plant Sci.* **15** (2024), 1360113.

[5] Benfenati, A.; Bolzi, D.; Causin, P.; Oberti, R. A deep learning generative model approach for image synthesis of plant leaves. *PLOS ONE* **17** (2022), e0276972.

[6] Isinkaye, F. O.; et al. Deep learning and content-based filtering techniques for improving plant disease identification and treatment recommendations: A comprehensive review. *Heliyon* **10** (2024), e29583.

[7] Ngugi, H. N.; Ezugwu, A. E.; Akinyelu, A. A.; Abualigah, L.; et al. Revolutionizing crop disease detection with computational deep learning: a comprehensive review. *Environ. Monit. Assess.* **196** (2024), 302.

[8] Xu, M.; Park, J.-E.; Lee, J.; Yang, J.; Yoon, S. Plant disease recognition datasets in the age of deep learning: challenges and opportunities. *Front. Plant Sci.* **15** (2024), 1452551.

[9] George, R.; et al. Past, present and future of deep plant leaf disease diagnosis: datasets and methods. *Comput. Electron. Agric.* **218** (2025), 108058.

[10] Klein, J.; et al. Synthetic data at scale: a development model to efficiently leverage ML in agriculture. *Front. Plant Sci.* **15** (2024), 1360113.

[11] Sajitha, P.; Andrushia, A. Diana; Anand, N.; Naser, M. Z. A Review on Machine Learning and Deep Learning Image-based Plant Disease Classification for Industrial Farming Systems. *Journal of Industrial Information Integration* **38** (2024), 100572. doi:10.1016/j.jii.2024.100572

[12] Shwetha, V.; et al. LeafSpotNet: A deep learning framework for detecting leaf blight spot disease stages. *Plant Communications* **9** (2024), 1000035.

[13] Pal, C.; et al. A lightweight and explainable CNN model for empowering plant disease diagnosis. *Sci. Rep.* **15** (2025), 94083.

[14] Lopes, F. A.; Sagan, V.; Esposito, M. PlantPlotGAN: A physics-informed generative adversarial network for plant disease prediction. *Proc. IEEE/CVF WACV* (2024), 1234–1244.

[15] Mara, M.; Pop, P.; Barata, J. A comparative study of machine learning models. *Proc. 19th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2024)*, Springer, 2024, 123–134.

[16] Mohanty, S. P.; Hughes, D. P.; Salathé, M. Using deep learning for image-based plant disease detection. *Front. Plant Sci.* **7** (2016), 1419.

[17] Muhammad, A.; et al. Harnessing the power of diffusion models for plant disease detection. *PLoS ONE* **18** (2023), e0289345.

[18] Sagar, S.; Javed, M.; Doermann, D. S. Leaf-Based Plant Disease Detection and Explainable AI. *arXiv preprint arXiv:2404.16833*, (2023). doi:10.48550/arXiv.2404.16833

[19] Rahat, I. S.; et al. Towards precision agriculture: tea leaf disease detection with deep learning. *Sci. Rep.* **15** (2025), 12345.

[20] Rashid, M. R. A.; et al. An ensemble learning framework with explainable AI for cucumber leaf disease diagnosis. *Intell. Agri.* **7** (2025), 100013.

[21] Nagaraju, M.; Chawla, P. Systematic review of deep learning techniques in plant disease detection. *International Journal of System Assurance Engineering and Management* **11**(3) (2020), 547-560. doi:10.1007/s13198-020-00972-1

[22] Zhang, K.; et al. Detecting soybean leaf disease from synthetic images using deep learning. *Comput. Electron. Agric.* **187** (2021), 106305.

[23] Wang, S.; Xu, D.; Liang, H.; Bai, Y.; Li, X.; Zhou, J.; Su, C.; Wei, W. Advances in Deep Learning Applications for Plant Disease and Pest Detection: A Review. *Remote Sensing* **17**(4) (2025), 698. doi:10.3390/rs17040698

[1] Technical University of Cluj-Napoca, Faculty of Science, Computer Science Department
Email address: `mara.hajdu@mi.utcluj.ro`