Early detection of tomato diseases based on a novel hybrid method combining wavelet decomposition, asymmetric generalized Gaussian distribution, and KNN classification*

Alphonse AZOMA[1,*,†], Boris ZANNOU[1,†], K.B. Audace DIDAVI[1,†], Constant A. ALIHONOU[1,†], Jean Ouèrè TOUPOUVOGUI[1,†], Marc Kokou ASSOGBA[2,†]

1 Mamou Higher Institute of Technology, Computer Science Department

2 Polytechnic School of Abomey-Calavi (EPAC) 01 BP 2009, Benin

Abstract

Early detection of tomato leaf diseases is crucial for food security and reducing pesticide use. Traditional methods based on visual inspection are subjective and often result in delayed detection. Laboratory diagnostics are also unsuitable for real-time applications and have limitations in terms of speed, accessibility, and responsiveness to current needs. This work therefore focuses on early diagnosis and uses a method called AGGD-MM-KNN, which is based on characterizing textures in macroscopic images of tomato leaves. To this end, a reference database of certified healthy and infected leaves extracted using Kaggle was created. Then, using Matlab R2024b, textures were extracted from the digital images using the Haar wavelet transform. The histograms of the resulting detailed images were approximated by a non-zero mean Asymmetric Generalized Gaussian Distribution (AGGD). The parameters of this Gaussian distribution are determined using the method of moments (MM) and constitute characteristic markers of the textures. Finally, based on these markers, the KNN classification achieves a detection rate of 99.29%, compared to 98.02% for the existing IDLFOA-DCTLFD model.

Keywords

Early detection, foliar diseases, wavelet transform, texture characterization, macroscopic images

1. Introduction

The tomato is a rich and widely cultivated food crop [1]. It is the most nutritious crop in the world; its production and cultivation have a major influence on the agricultural economy [2], and its demand continues to grow [3]. According to statistics, smallholder farmers harvest only 80% of agricultural production due to pests and diseases. In agriculture, plant diseases are a major threat [4]. Various tomato diseases can affect the fruit, leaves, roots, and stems of the plant [5]. Common plant diseases are caused by fungi, nematodes, bacteria, and viruses, which cause spots on the stems or leaves, black or brown lesions, yellowing of the lower leaves, black spots, and death of the lower leaves.

2. Problem statement

*CITA 2026 - Land, Smart Cities, and Sustainable Development, 25-26 June 2026, Cotonou, Benin

[1,*] Corresponding author. alphonseazoma@gmail.com (A. Alphonse)

† These authors contributed equally.

zansboris4@gmail.com (Z. Boris); didavia@gmail.com (K.B.D. Audace); agnimoanalihonou@gmail.com (A. A. Constant) toupou1976@gmail.com (J. O. Toupouvoguy); mkokouassogba@yahoo.fr (M. K. ASSOGBA)

0009-0004-6328-0799 (M. K. ASSOGBA)

The tomato is a rich and widely cultivated food crop [1]. It is the most nutritious crop in the world; its production and cultivation have a major influence on the agricultural economy [2], and its demand continues to grow [3]. According to statistics, smallholder farmers harvest only 80% of agricultural production due to pests and diseases. In agriculture, plant diseases are a major threat [4]. Various tomato diseases can affect the fruit, leaves, roots, and stems of the plant [5]. Common plant diseases are caused by fungi, nematodes, bacteria, and viruses, which cause spots on the stems or leaves, black or brown lesions, yellowing of the lower leaves, black spots, and death of the lower leaves.

3. Related work

Current advances in computer science have led to machine learning (ML) and artificial intelligence (AI) models that contribute to the automated detection of tomato leaf and fruit diseases through computerized methods of observing tomato crops [9]. Machine learning-based methods have improved disease recognition by leveraging digital image processing for leaf and fruit disease classification. Deep learning (DL) with neural networks (NN) improves accuracy through efficient feature extraction. Given the important role of tomatoes, improving disease detection and classification is crucial for sustainable agriculture in the face of increasing demand [10]. The authors of [11] used YOLOv8 as the fundamental structure. The Ultralytics Hub platform offers optimal parameters for training YOLOv5 methods. and YOLOv8. The authors of [12] propose TomatoDet by integrating a self-attention mechanism from Swin-DDETR. Furthermore, the dynamic activation function Meta-ACON of the backbone network enhances the system's ability to identify relevant disease features. The authors of [13] proposed a convolutional neural network (CNN) model in different regions. The integration of image processing and computer vision with deep learning models has enabled significant progress in these regions. The authors of [14] present a technique based on improved YOLOv7. In [15], a novel technique using the Yolov8 structure in deep learning models is proposed. The authors of [16] investigated the performance of InceptionResNetV2 and Xception. The authors of [17] propose a deep learning classification based on ResNet-50. The authors of [18] present seven Bayesian-optimized hybrid models integrating a custom CNN with conventional machine learning techniques. In [19], a deep learning approach using CNNs, integrating several feature extraction methods and gray wolf optimization (GWO), is introduced for increased accuracy on various plant leaves. In [20], a multi-objective hybrid optimization model based on the fruit fly and relying on a simulated annealing optimized SVM is presented. The authors of [21] present an automated technique using a two-stream deep learning approach with machine learning classifiers. The authors of [22] use a deep learning method integrating various classifiers, including Random Forest (RF), Inception V3, DenseNet, ResNet50, Xception, and MobileNet. In [23], an optimized CNN methodology is presented. Existing studies exhibit limited adaptability to diverse datasets and complicate computations. Furthermore, existing models struggle to detect small objects and generalize, thus highlighting the need for more robust models to improve classification accuracy.

4. Background

For this work, we propose a new method for characterizing textures in digital images based on wavelet decomposition, which is non-destructive and allows for localized multi-resolution analysis in both the spatial and frequency domains. This characteristic is crucial for capturing the subtle alterations in leaf texture that precede the appearance of visible symptoms, such as changes in surface roughness, disturbances in cell structure, and local variations in pigmentation.

5. Data source

5.1. Study data

We used macroscopic images extracted from the "Plant Village" database obtained from the Kaggle website, containing three hundred (300) images of healthy tomato leaves and three hundred (300) images of infected tomato leaves of the same size at an early stage, whose characteristics are given in Table 1. Each image was saved as a JPEG file in RGB color format. Figure 1 shows some of the images. We wrote and implemented our algorithms using Matlab R2024b.

Table 1
Composition of the image database

Dataset	Source and link	Images number	Classes
Tomato Leaves	https://www.kaggle.com	300	Healthy Tomato Leaves
		300	Infected tomato leaves

(a) (b) (c) (d)

Figure 1: Healthy Tomato Leaves: (a), (b); infected tomato leaves :(c), (d)

5.2. Study design

The approach used in the proposed AGGD-MM-SVM model is illustrated in the following figure 2

6. Methods

6.1. Data preprocessing

In order to effectively use image data to train our detection model, several essential preprocessing steps are necessary:

6.1.1. Image resizing

Consider a source image J with original dimensions (H, W) (height × width) of pixels (XJ, YJ) resized to an image I with dimensions (256, 256) of pixels (XI, YI):

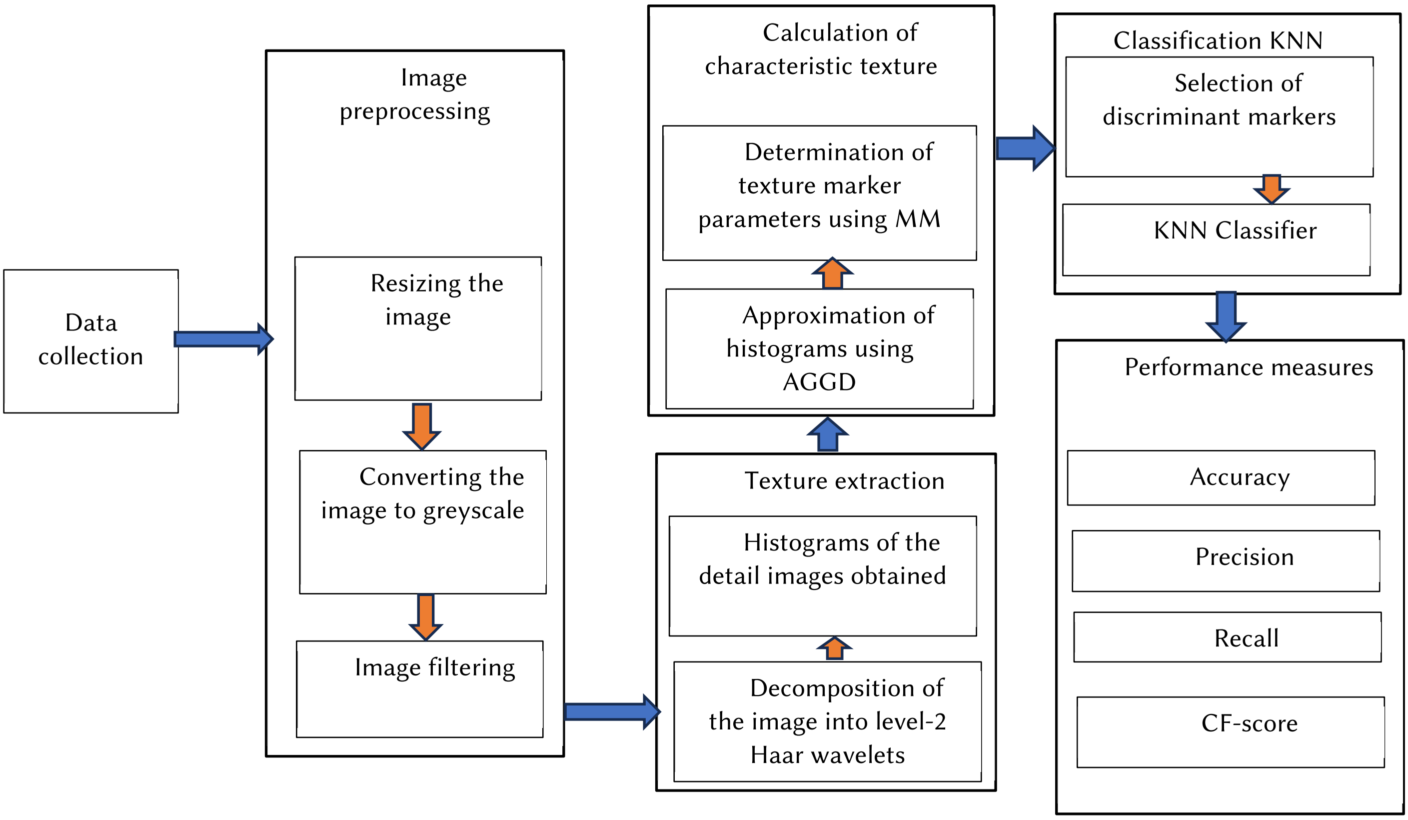

Figure 2: Framework for proposed system

$$X_J = \frac{X_I . W}{256} \qquad (1)$$

$$Y_J = \frac{Y_I . H}{256} \qquad (2)$$

6.1.2. Converting the image to greyscale

$$I_{gris} = 0{,}299 . R(x, y) + 0{,}587 . G(x, y) + 0{,}114 . B(x, y) \qquad (3)$$

With R, G, B representing the values of the red, green, and blue channels (0 to 256)

6.1.3. Image filtering

To reduce noise due to outliers in isolated pixels, we apply the median filter defined by formula (4) to these grayscale images:

$$h(x, y) = \text{median}\{I(m, n) \ / \ (m, n) \in S(x, y)\} \quad (4)$$

Where h represents the median filter, I the grayscale image to be filtered, and S(x, y) a 3x3 neighborhood of (x, y).

6.1.4. Image normalization

$$I_{norm}(x, y) = \frac{I(x, y) - I_{min}}{I_{max} - I_{min}} \qquad (5)$$

With I(x, y): value of the source pixel (grayscale)
$I_{min} = \min I(x, y)$: minimum value in the image
$I_{max} = \max I(x, y)$: maximum value in the image

6.2. Image texture extraction

6.2.1. Haar Wavelet Decomposition of the Normalized Image

The wavelet transform decomposes an input signal x(t) into a series of wavelet functions $\psi(a,b)(t)$ (formula 6) which are derived from a given parent function $\psi(t)$ by dilation and translation operations (formula 7).

$$C_{a,b} = \int_{-\infty}^{+\infty} X(t)\, \psi_{a,b}(t) dt \qquad (6)$$

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t - b}{a}\right) \qquad a \neq 0 \quad (7)$$

With a being the scale factor and b the translation factor, the resulting coefficients Ca,b [24] contains information about the signal x(t) studied at different scales. In the case of images, the wavelet transform is applied using a bank of low-pass g and high-pass h filters, first row by row, then column by column. Four images are then generated at each level. In the present work, the filtered and normalized image of the tomato leaves is wavelet decomposed into two levels with the Haar mother wavelet according to the scheme [25] (Fig. 3). The images Ih2, Iv2, and Id2 represent the high-frequency detail components of level two and characterize the image texture in the horizontal, vertical, and diagonal directions [26].

6.2.2. Histograms of detail images Ih2, Iv2, and Id2

For each of the H×W-sized images Ih2, Iv2, and Id2, the histogram h(k) is:

$$h(k) = \sum_{x=0}^{w-1} \sum_{y=0}^{H-1} \delta[I(x, y) - k] \qquad (8)$$

with k: intensity level (generally from 0 to 1 for a normalized image)
δ: Kronecker function (δ(0) =1, δ(k)=0 for k≠0)

6.3. Characterization of detail images Ih2, Iv2, and Id2

6.3.1. Approximation of histograms using AGGD

The histograms of each detail image contain information about the textures in the digital images [26]. However, they are often noisy because they contain outliers or missing values. They must then be approximated by a parametric distribution. The histogram distribution of a detail component follows a Generalized Asymmetric Gaussian (AGGD) distribution with respect to a zero mean [27]. In order to adapt it to the specific characteristics of diseases on tomato leaves, we introduce a non-zero mean μ to obtain a generalized Gaussian distribution that is asymmetric with respect to a non-zero mean:

$$P(x, \alpha_L, \alpha_R, \beta) = \frac{\beta}{(\alpha_L + \alpha_R)\Gamma\left(\frac{1}{\beta}\right)} e^{-\left(\frac{|x-\mu|}{\alpha_L}\right)^{\beta}} \quad \text{si } x < \mu \qquad (9)$$

$$P(x, \alpha_L, \alpha_R, \beta) = \frac{\beta}{(\alpha_L + \alpha_R)\Gamma\left(\frac{1}{\beta}\right)} e^{-\left(\frac{|x-\mu|}{\alpha_R}\right)^{\beta}} \quad \text{si } x \geq \mu \qquad (10)$$

Where αL and αR are the left and right scaling parameters respectively, β is the shape parameter, and x is the gray level of a pixel (i,j) in the detail image.

6.3.2. Calculation of the characteristic texture parameters αL, αR, and β

These parameters are calculated using the method of moments [27]:

$$\alpha_L = \left(\frac{\Gamma\left(\frac{1}{\beta}\right) \sum_{\substack{i=1 \\ x_i<\mu}}^{N} |x_i|^2}{N_L \Gamma\left(\frac{3}{\beta}\right)} \right)^{1/2} \qquad (11)$$

$$\alpha_R = \left(\frac{\Gamma\left(\frac{1}{\beta}\right) \sum_{\substack{i=1 \\ x_i\geq\mu}}^{N} |x_i|^2}{N_R \Gamma\left(\frac{3}{\beta}\right)} \right)^{1/2} \qquad (12)$$

μ the average grey level of the pixels (i,j) of the detail image:

$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad (13)$$

$$\gamma = \left(\frac{\frac{1}{N_L} \sum_{\substack{i=1 \\ x_i<\mu}}^{N} |x_i|^2}{\frac{1}{N_R} \sum_{\substack{i=1 \\ x_i\geq\mu}}^{N} |x_i|^2} \right)^{1/2} \qquad (14)$$

$$N_L = \text{card}(x_i, x_i < \mu) \qquad (15)$$

$$N_R = \mathrm{card}(x_i, x_i \geq \mu) \qquad (16)$$

$$\beta = GGR^{-1}\left[\frac{(1+\gamma)(1+\gamma^3)\left(\sum_{1=1}^{N}|x_i|\right)^2}{N(1+\gamma^2)^2\sum_{i=1}^{N}|x_i|^2}\right] \qquad (17)$$

The parameter β obtained is a characteristic of the texture in the detailed images obtained. Thus, for the healthy tomato, we have the diagonal, horizontal, and vertical textures characterized respectively by βsd, βsh, and βsv, and for the infected tomato, the diagonal, horizontal, and vertical textures characterized respectively by βid, βih, and βiv.

6.4. Classification by KNN

6.4.1. Selection of discriminatory parameters

The calculated horizontal and vertical texture characteristics (βsh, βsv) were used to identify the healthy tomato, and the diagonal and vertical texture characteristics (βid, βiv) were used to identify the infected tomato.

6.4.2. Classification by KNN

To discriminate between healthy and infected tomatoes, the nearest neighbor classification method was used. The dataset was divided into two parts, with 70% allocated to training and 30% to testing. Standard statistical measures widely used in machine learning classification, including recall, accuracy, precision, and F-score, were used to evaluate the performance of the proposed AGGD-MM-KNN model.

7. Results

The result in Figure 3 demonstrates that the selected parameters are relevant and allow the model to perform better discrimination between healthy tomatoes and infected tomatoes.

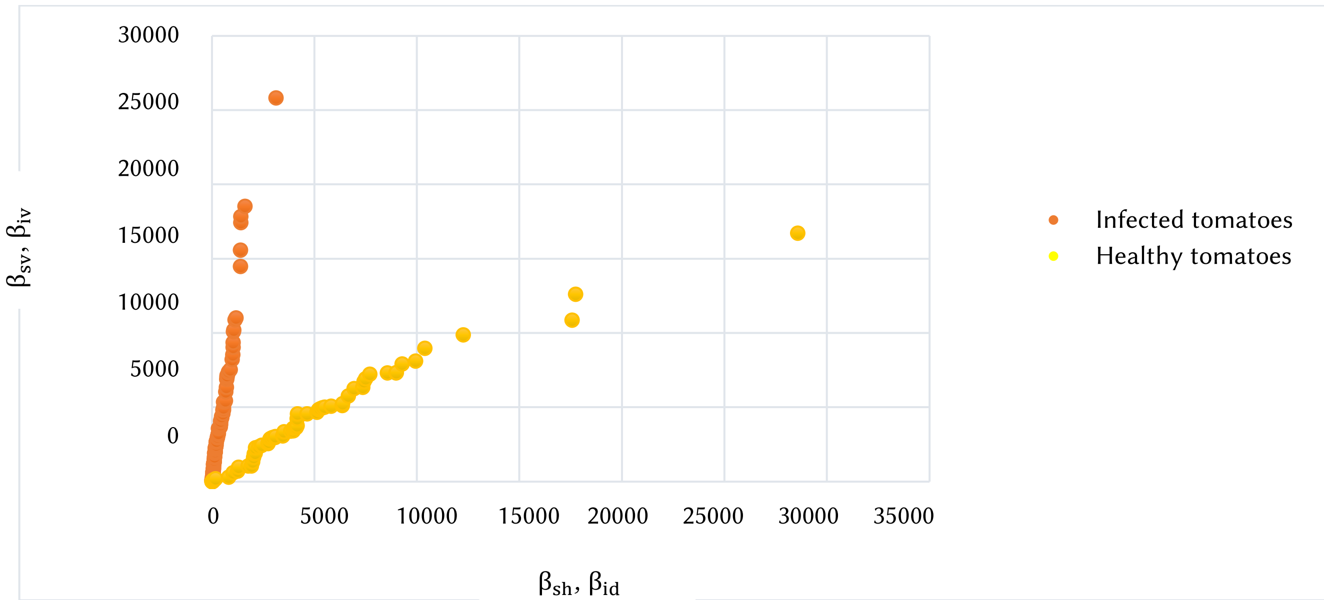

Figure 3: Discrimination between healthy tomatoes and infected tomatoes

This result is confirmed by the cross-validation curve in Figure 4, which provides an accuracy of 99.29% for k=1. The other performance values are calculated from the confusion matrix and summarized in the table 2.

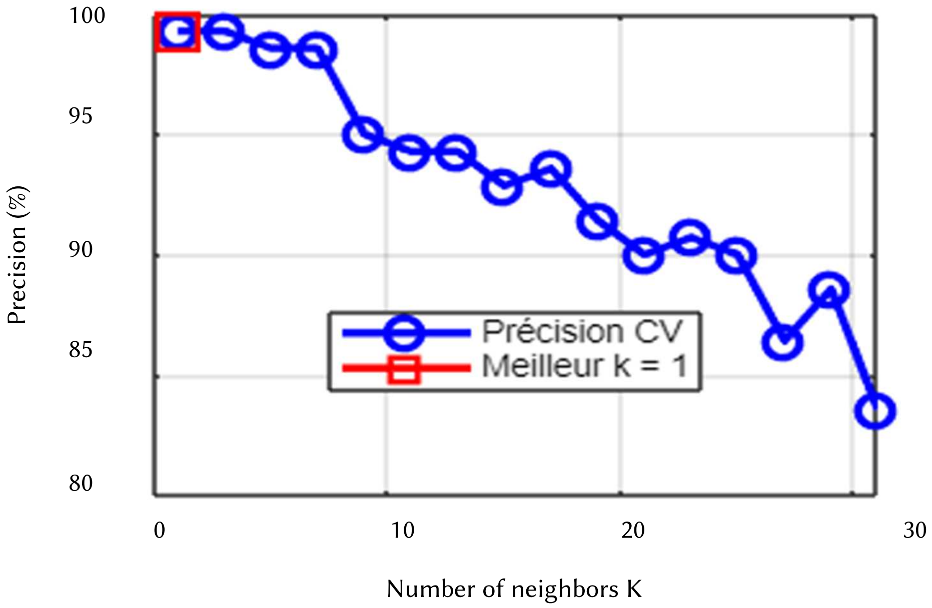

Figure 4: cross-validation curve

Table 2
Performances of the AGGD-MM-KNN model

Performances of the AGGD-MM-KNN model			
Accuracy	F-Score	Recall	Precision
99,29%	100%	100%	99,29%

The performance of the proposed AGGD-MM-KNN model is compared with that of recently developed models and shown in Figure 5, then recorded in Table 3.

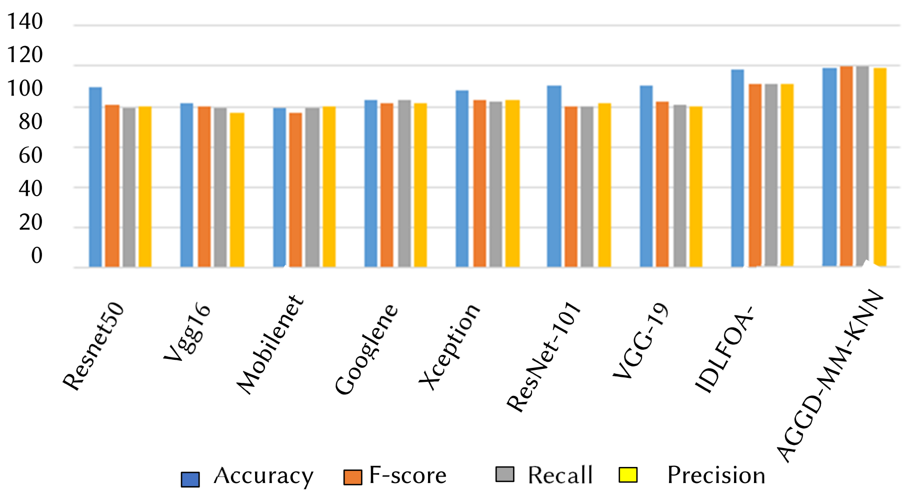

Figure 5: Comparison of models performances

Table 3
Comparison of models performances

Method	Accuracy (%)	F-score (%)	Recall (%)	Precision (%)
Resnet50	89,65	81	79	80
Vgg16	81,75	80	79	77
Mobilenet	79,2	77	79	80
Googlenet	82,81	82	83	82
Xception	88,16	83,19	82,14	83,25
ResNet-101	90,13	80,04	80,13	81,95
VGG-19	90,42	82,43	80,47	80,39
IDLFOA-DCTLFD	98,02	90,98	91,16	90,96
AGGD-MM-KNN	**99,29**	**100**	**100**	**99,29**

8. Conclusion

In this work, we proposed a new model to characterize wavelet coefficient distributions for the early detection of tomato diseases: the Asymmetric Generalized Gaussian Distribution (AGGD) with a non-zero mean, combined with the method of moments (MM) followed by the KNN classification (AGGD-MM-KNN). Experimental results obtained using images of healthy and infected tomatoes showed that the AGGD-MM-KNN method is more suitable than the IDLFOA-DCTLFD method for modeling wavelet coefficient distributions and is therefore more effective in the early detection of diseases. In the tests, we used the Haar wavelet transform; however, this model can be used with other types of wavelets transforms. We believe that using the AGGD-MM-KNN model instead of the IDLFOA-DCTLFD model in early detection applications in image processing will significantly improve performance.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] A. Chug, A. Bhatia, A. P. Singh, and D. Singh, "A novel framework for image-based plant disease detection using hybrid deep learning approach," Soft Computing, vol. 27, no. 18, pp. 13613–13638, Sep.

[2] H. M. Zayani et al., "Deep Learning for Tomato Disease Detection with YOLOv8," Engineering, Technology & Applied Science Research, vol.14, no. 2, pp. 13584–13591, Apr. 2024, https://doi.org/10.48084/etasr.7064.

[3] M. S. Alzahrani and F. W. Alsaade, "Transform and Deep Learning Algorithms for the Early Detection and Recognition of Tomato Leaf Disease," Agronomy, vol. 13, no. 5, May 2023, Art. no. 1184, https://doi.org/10.3390/agronomy13051184.

[4] A. Wang et al., "NVW-YOLOv8s: An improved YOLOv8s network for real-time detection and segmentation of tomato fruits at different ripeness stages," Computers and Electronics in Agriculture, vol. 219, Apr. 2024, Art. no. 108833, https://doi.org/10.1016/j.compag.2024. 108833.

[5] M. M. Elsharkawy et al., "Systemic resistance induction of tomato plants against tomato mosaic virus by microalgae," Egyptian Journal of Biological Pest Control, vol. 32, no. 1, Apr. 2022, Art. no. 37, https://doi.org/10.1186/s41938-022-00538-2.

[6] N. Aishwarya, N. G. Praveena, S. Priyanka, and J. Pramod, "Smart farming for detection and identification of tomato plant diseases using light weight deep neural network," Multimedia

Tools and Applications, vol. 82, no. 12, pp. 18799–18810, May 2023, https://doi.org/10.1007/s11042-022-14272-2.

[7] V. Shwetha, A. Bhagwat, and V. Laxmi, "LeafSpotNet: A deep learning framework for detecting leaf spot disease in jasmine plants," Artificial Intelligence in Agriculture, vol. 12, pp. 1–18, Jun. 2024, https://doi.org/10.1016/j.aiia.2024.02.002.

[8] A. Gatla et al., "Optimizing Edge AI for Tomato Leaf Disease Identification," Engineering, Technology & Applied Science Research, vol. 14, no. 4, pp. 16061–16068, Aug. 2024, https://doi.org/10.48084/etasr.7802.

[9] R. Rajamohanan and B. C. Latha, "An Optimized YOLO v5 Model for Tomato Leaf Disease Classification with Field Dataset," Engineering, Technology & Applied Science Research, vol. 13, no. 6, pp. 12033– 12038, Dec. 2023, https://doi.org/10.48084/etasr.6377.

[10] R. Ahmed and E. H. Abd-Elkawy, "Improved Tomato Disease Detection with YOLOv5 and YOLOv8," Engineering, Technology & Applied Science Research, vol. 14, no. 3, pp. 13922–13928, Jun. 2024, https://doi.org/10.48084/etasr.7262.

[11] A. Abdullah, G. A. Amran, S. M. A. Tahmid, A. Alabrah, A. A. ALBakhrani, and A. Ali, "A Deep-Learning-Based Model for the Detection of Diseased Tomato Leaves," Agronomy, vol. 14, no. 7, Jul. 2024, Art. no. 1593, https://doi.org/10.3390/agronomy14071593.

[12] X. Wang and J. Liu, "An efficient deep learning model for tomato disease detection," Plant Methods, vol. 20, no. 1, May 2024, Art. no. 61, https://doi.org/10.1186/s13007-024-01188-1.

[13] M. B. Yildiz, M. F. Hafif, E. K. Koksoy, and R. Kurşun, "Classification of Diseases in Tomato Leaves Using Deep Learning Methods," Intelligent Methods In Engineering Sciences, vol. 3, no. 1, pp. 22–36, Mar. 2024, https://doi.org/10.58190/imiens.2024.84.

[14] M. Umar, S. Altaf, S. Ahmad, H. Mahmoud, A. S. N. Mohamed, and R. Ayub, "Precision Agriculture Through Deep Learning: Tomato Plant Multiple Diseases Recognition with CNN and Improved YOLOv7," IEEE Access, vol. 12, pp. 49167–49183, Jan. 2024, https://doi.org/10.1109/ACCESS.2024.3383154.

[15] X. Guo, "Automatic detection of tomato leaf disease using an adopted deep learning algorithm," Journal of Intelligent & Fuzzy Systems, vol.46, no. 4, pp. 7909–7921, Jan. 2024, https://doi.org/10.3233/JIFS-236905.

[16] P. K. Nalli, D. P. Garapati, M. V. Subbarao, E. Katta, A. S. Krishna, and N. Deevi, "Comparative Study of Deep Learning Techniques for Detecting Tomato Plant Leaf Diseases Using Transfer Learning," in International Conference on Distributed Computing and Optimization Techniques, Bengaluru, India, Mar. 2024, pp. 1–7, https://doi.org/10.1109/ICDCOT61034.2024.10516123

[17] L. Upadhyay and A. Saxena, "Evaluation of Enhanced Resnet-50 Based Deep Learning Classifier for Tomato Leaf Disease Detection and Classification," Journal of Electrical Systems, vol. 20, no. 3s, pp. 2270–2282, Mar. 2024, https://doi.org/10.52783/jes.1856.

[18] B. Khan et al., "Bayesian optimized multimodal deep hybrid learning approach for tomato leaf disease classification," Scientific Reports, vol.14, no. 1, Sep. 2024, Art. no. 21525, https://doi.org/10.1038/s41598-024-72237-x.

[19] A. N. Jabbar and H. Koyuncu, "Deep learning and grey wolf optimization technique for plant disease detection: a novel methodology for improved agricultural health," Traitement du Signal, vol. 40, no. 5, pp. 1961–1972, 2023, https://doi.org/10.18280/ts.400515.

[20] E. Gangadevi, R. S. Rani, R. K. Dhanaraj, and A. Nayyar, "Spot-out fruit fly algorithm with simulated annealing optimized SVM for detecting tomato plant diseases," Neural Computing and Applications, vol. 36, no.8, pp. 4349–4375, Mar. 2024, https://doi.org/10.1007/s00521-023-09295-1.

[21] U. Zahra, M. A. Khan, M. Alhaisoni, A. Alasiry, M. Marzougui, and A. Masood, "An Integrated Framework of Two-Stream Deep LearningModels Optimal Information Fusion for Fruits Disease Recognition," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 3038–3052, 2024, https://doi.org/10.1109/JSTARS.2023.3339297.

[22] A. Singh, S. Kumar, and D. Choudhury, "Tomato Leaf Disease Prediction Based on Deep Learning Techniques," in International Conference on Computation of Artificial Intelligence & Machine Learning, Jaipur, India, Jan. 2024, pp. 357–375, https://doi.org/10.1007/978-3-031-71481-8_28.
[23] T. Soujanya, S. Padmaja, D. Ramesh, Shabana, and S. Mohmmad, "An Optimized Deep Learning Technique for Enhance Disease Identification in Tomato Leaf to Promote Sustainable Agriculture," in International Conference on Evolutionary Algorithms and Soft Computing Techniques, Bengaluru, India, Oct. 2023, pp. 1–5, https://doi.org/10.1109/EASCT59475.2023.10392310.
[24] Grossman A., Morlet J., 1984, Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape, SIAM J. of Math. vol. 15, no. 4, pp. 723-736.
[25] Van de Wouwer G., Scheunders P., Van Dyck D., 1999, Statistical texture characterization from discrete wavelet representations. IEEE Transactions on Image Processing, vol. 8, no. 4, pages 592–598.
[26] Lefebvre A. 2011, Contribution de la texture pour l'analyse d'images à très haute résolution spatiale: application a la détection de changement en milieu périurbain. Géographie. Thèse de Doctorat de l'Université de Rennes 2; Université Européenne de Bretagne, p 274.
[27] Lasmar N. 2012, Modélisation stochastique pour l'analyse d'images texturées: Approches Bayesiennes pour la caractérisation dans le domaine des transformées. Thèse de Doctorat de l'Université de Bordeaux I, p 216.
[28] K. Sundaramoorthi, Mari Kamarasan, "Optimization Algorithm for Early Detection and Classification of Tomato Leaf and Fruit Diseases", Engineering, Technology & Applied Science Research, November 2024, DOI: https://doi.org/10.48084/etasr.9216