

Early detection of tomato leaf diseases based on deep learning techniques

Mohammed Hussein Najim¹, Salwa Khalid Abdulateef¹, Abbas Hanon Alasadi²

¹Department of Computer Science, College of Computer Science and Mathematics, Tikrit University, Tikrit, Iraq

²Department of Computer Science, College of Computer Science and Information Technology, University of Basrah, Basrah, Iraq

Article Info

Article history:

Received Jan 31, 2023

Revised Jun 8, 2023

Accepted Jun 12, 2023

Keywords:

Convolution neural network

Deep learning

Detection

Plant disease

Tomato leaf disease

ABSTRACT

Tomato leaf diseases are a big issue for producers, and finding a single method to combat them is tough. Deep learning techniques, notably convolutional neural networks (CNNs), show promise in recognizing early indicators of illness, which can help producers avoid costly concerns in the future. In this study, we present a CNN-based model for the early identification of tomato leaf diseases to preserve output and boost yield. We used a dataset from the plantvillage database with 11,000 photos from 10 distinct disease categories to train our model. Our CNN was trained on this dataset, and the suggested model obtained an astounding 96% accuracy rate. This shows that our method has the potential to be efficient in detecting tomato leaf diseases early on, therefore assisting producers in managing and reducing disease outbreaks and, as a result, resulting in higher crop yields.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Salwa Khalid Abdulateef

Department of computer science, College of Computer Science and Mathematics, Tikrit University
Tikrit, Iraq

Email: Khalid.salwa@tu.edu.iq

1. INTRODUCTION

The agricultural sector is considered the most productive among the various sectors in most countries, with agriculture playing a significant role in the country's global economy [1], [2]. After the oil sector, the Iraqi agricultural sector is the second largest contributor to the country's gross domestic product (GDP), employing more than 20% of the country's workforce [3]. One of the most cultivated and produced crops is the tomato crop [4], widely cultivated in fields and farms; out of every ten farmers, nine grow tomatoes on their farms [5]. The sixth-most abundant vegetable in the world is tomatoes, according to the food and agriculture organization (FAO), with an average tomato production of more than 180 million tons per year worldwide [6], [7] and per capita consumption of more than 20 kg per year [8], which represents about 15% of the total vegetable consumption [9].

Tomato is the most common vegetable in Iraq, with about 771,000 tons produced across the country, with high production in Karbala [10], Basrah [11], and Najaf [12]. The main reason for the decline in crop production and the increase in losses is pests and diseases, as tons of crops are lost annually [13]. Therefore, early detection of these diseases is urgent to avoid massive losses and increase yields. The traditional methods of detecting plant diseases need diagnostic experts [14], are expensive in terms of time and effort, and need special tools in addition to a large percentage of error in diagnosis and cannot be performed with the naked eye [15], [16]. Recently, due to the advancement of computer technology, particularly artificial intelligence techniques, it has been able to detect plant diseases early utilizing images to differentiate between diseases, an automated process for fields and farms [17], [18]. This research article presents a new model that can detect and classify the most common diseases in tomato leaves without needing diagnostic experts, thus helping farmers improve production

and increase profits. This presented model is based on convolution neural networks (CNN), a high-performance deep learning network commonly utilized in crop disease identification [19], [20].

The following is how the paper is structured: Section 2 reviews the literature on current approaches. Section 3 covers the recommended technique, model, and measures followed to achieve the desired outcomes. Section 4 discusses the results, analyzes the proposed methodology, and compares it with previous models. Section 5 contains the paper's conclusion and scope for future work.

2. LITERATURE SURVEY

Based on machine learning and image processing, numerous researchers are attempting to offer various solutions to the issue of tomato leaf disease detection to produce an accurate automated classification. Researchers have recently begun to use deep learning techniques to produce more accurate results. Several of the most appropriate deep learning-based methods for detecting tomato leaf disease are covered in this section: Tm *et al.* [21] proposed a CNN-based LeNet model to detect and identify ten diseases in tomato crops in 2018. With the least amount of computational resources and the easiest technology (LeNet), the suggested work addresses the issue of tomato leaf disease identification. The photos in the PlantVillage dataset are scaled down to 60*60 resolution to expedite training and make model training computationally possible. The accuracy of the suggested system ranges from 94% to 95% on average.

Kumar *et al.* [22] analyzed visual geometry group (VGG) Net, LeNet, ResNet50, and Xception from scratch to classify nine different tomato leaf diseases using a pair of PlantVillage datasets, the ImageNet dataset in 2019. The findings demonstrate that, out of all the investigated architectures, the fine-tuned VGGNet achieves the enormous accuracy and lowest loss, with a test accuracy of 99.25%. The key drawback of the suggested approach is that training takes a lot of time and requires expensive hardware equipment.

Ashok *et al.* [15] introduced a new CNN-based model for detecting tomato plant leaf disease that has been developed utilizing open-source algorithms, segmentation, and clustering-based image processing techniques to improve the image. A Gaussian filter is used during preprocessing to remove blur and reduce noise. The feature is extracted using discrete wavelet transform (DWT) and grey-level co-occurrence matrix (GLCM). An accuracy level of 98% was attained using the suggested strategy.

The study of Zhao *et al.* [9] has developed a deep convolutional neural network with a multi-domain of feature extraction to diagnose tomato leaf diseases. The model is trained to distinguish between healthy and diseased tomato leaf images by integrating attention density and residual mass depth. The experiment findings demonstrated that the suggested model achieves an average identification accuracy of 96.81% on the tomato leaf diseases dataset, outperforming prior deep-learning experiments utilizing the most common PlantVillage dataset.

Benoso *et al.* [23] demonstrated developing applications and computational systems for disease detection in tomato plants in 2020. The authors used texture and color statistical descriptors to extract traits and different classifiers to identify diseases. Three diseases can be detected using the methods proposed in this article (mosaic virus, late blight, and septoria leaf spot) on tomato leaves. The plant village image dataset has 160 images for each disease, classified by K-nearest neighbors (K-NN), random forest, support vector machine (SVM), artificial neural network (ANN), and naive bayes. Random forest has the highest accuracy of 90.7%.

According to Agarwal *et al.* [5] created a CNN model with three max-pooling layers, three convolution layers, and two fully connected layers. The model's effectiveness was evaluated using various criteria, including training, validation, test accuracy, and the number of trainable parameters. The plant village dataset has nine disease classes and classes with healthy images. The proposed model has a 91.2% average accuracy. In addition, Four CNN architectures, including VGG-19, ResNet, VGG-16, and Inception V3, were proposed by Ahmad *et al.* [24] in 2020 to recognize and categorize tomato leaf diseases. The two data sets used in this study were self-collected field data and a lab-based dataset called Plant village. On the field-based dataset, these models don't perform well in the laboratory-based dataset using only four tomato leaf disease classes (2,364 images). On both datasets, Inception V3 is found to have the best performance. The highest accuracy of 93.40% was achieved in the laboratory-based data set.

Based on the recently created efficientNet CNN model, Chowdhury *et al.* [25] built a model of a deep convolutional neural network to segment the leaf images from the background. The authors used two segmentation models, a U-net and a modified U-net. In addition, a different class of healthy and diseased leaves was used for classifications. The findings demonstrated that the model performed better than several contemporary deep learning techniques, with an accuracy of 99.89%.

Two convolutional neural networks (CNN), GoogLeNet and VGG16, were used by Kibriya *et al.* [26] to classify tomato leaf diseases. By utilizing a deep learning strategy, the suggested work seeks to identify the ideal answer to the issue of tomato leaf disease detection. On the plant village dataset of 10,735 leaf pictures, VGG16 achieved an accuracy of 98.8% and GoogLeNet a score of 99.23%. In this study, the tomato plant's leaves contained just three infections, which were identified and categorized.

3. PROPOSED METHODOLOGY

3.1. Dataset

The pictures of the tomato leaf disease were collected from the publicly published plantvillage [27] database in Kaggle, which contains 11,000 images of ten different classes, nine different disease classes, and a health class. The database was divided into 10,000 training images for each class, 1,000 images, and 1,000 for testing for each class 100 images. The size of all the images is 256×256 , and the format is jpg. This class is the most common tomato leaf disease: mosaic virus, target spot, late blight, septoria leaf spot, yellow leaf curl virus, leaf mold, two-spotted spider mite, early blight, and bacterial spot, as shown in Figure 1. The dataset contained noise-free images, so noise removal was not required as a preprocessing step. The images have been downsized to a resolution of 128×128 pixels to expedite the training process.

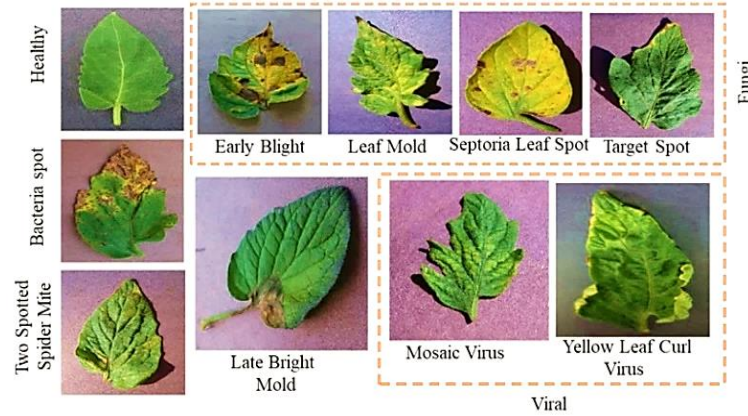


Figure 1. Classes sample image of the dataset

3.2. Model evaluation

To evaluate the model, the accuracy test was used to measure the accuracy of the submitted model, as accuracy determines the classifier's capacity to make a precise diagnosis. The accuracy equation is shown in (1). Also, Precision (Pre), F_score, and Recall (Rec) were used to measure the performance of our presented model, as these are the most popular and efficient measures in the implementation of models. These measures are explained in the following equations:

$$Ac = \frac{(P1+N1)}{(P1+N1+F+F2)} \quad (1)$$

$$Pre = \frac{P1}{P1+F} \quad (2)$$

$$F_{score} = 2 * \frac{Pre*Rec}{Pre+Rec} \quad (3)$$

$$Rec = \frac{P1}{P1+N2} \quad (4)$$

P1 is a number of true positives (the outcome is positive and expected to be positive).

N2- refers to the number of false negatives (despite predictions to the contrary, a positive result occurs).

F - denotes the number of false positives (the result is negative and was predicted to be negative.)

N1- defines the number of true negatives (despite being predicted to be positive, the result is negative)

3.3. Model

The submitted CNN model has been implemented in Colab, an integrated software environment. The proposed CNN model consists of two parts, feature extraction, and classification. For feature extraction, three convolution layers are followed by three max-pooling layers; the first convolution layer contains 64 filters, the second layer contains 32 filters, and the third layer contains 16 filters. All the filters have 3×3 sizes. The rectified linear units (ReLU) activation function is present in these three layers, with a max pooling layer of size 2×2 . The second part of this proposed model is the classification part. Our model used a fully connected

layer containing 500 neurons, followed by the softmax activation function. Figure 2 illustrates our proposed model.

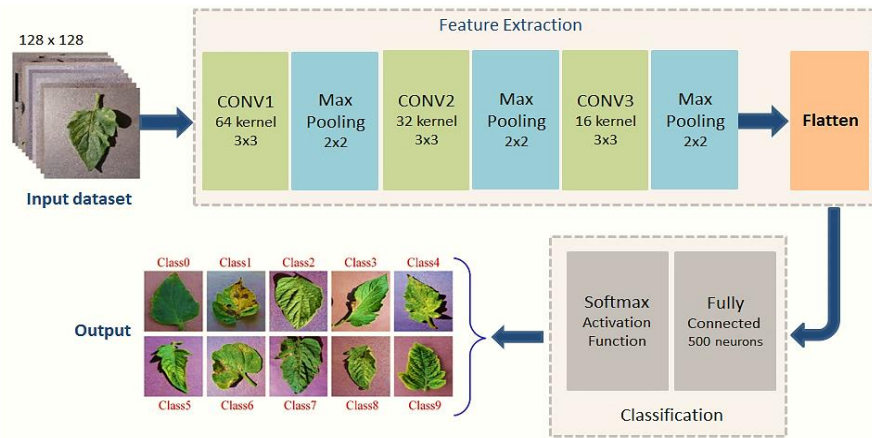


Figure 2. The structure of tomato leaf disease diagnosis

The optimization was carried out with an Adam optimizer, and a loss function was categorical cross entropy whose (5) is shown as follows:

$$E(W, b) = - \sum_{i=1}^m y_i \log(p_i) \quad (5)$$

where (p) is the expected probability and (y) is the goal probability, and (m) is the number of classes. A dropout (0.5) was used to avoid overfitting and data augmentation techniques such as image zooming, vertical cropping, horizontal flip, and rescaling. The model was trained for 40 epochs with 64 batch sizes. The hyperparameters for the model are described in Table 1.

Table 1. Hyperparameters for our model

Hyperparameter	Description
Conv layer	3
Max pool layer	3
Dropout rate	0.5
Activation function	ReLU
Batch size	64
Number of epoch	40

4. RESULTS AND DISCUSSION

The obtained results prove the possibility of the presented model, based on CNN, its efficiency in detecting and classifying tomato leaf diseases. By training the model on the PlantVillage dataset, the highest accuracy value in 40 epochs was obtained, with an average of 96% and a loss magnitude of 0.1. The test accuracy was 92%. The value of the precision was obtained at 91%, the value of the recall was 92%, and the value of the F-score was 92. Table 2 shows the most important results obtained.

Table 2. Model results

Model	Dataset	Accuracy	Loss	Precision	Recall	F1-Score
Special Model (CNN)	PlantVillage	96%	0.1	91%	92%	92%

The proposed model is distinguished from the traditional models, such as VGG16 and ResNet, in its speed and does not require ample storage space compared to the traditional CNN models. The accuracy rating for the training and testing process in conjunction with the epoch number is presented in Figure 3. The loss values for the training and testing processes are displayed in Figure 4.

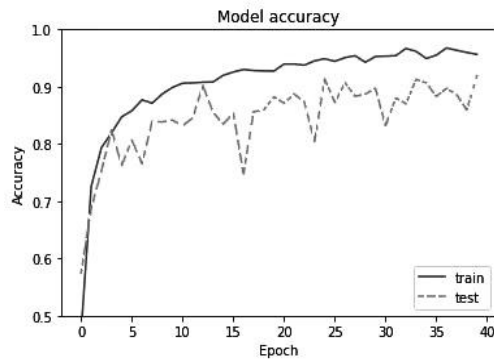


Figure 3. Model accuracy

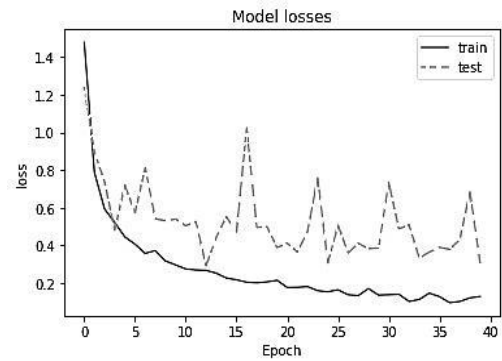


Figure 4. Model losses

The suggested model compares with previous CNN models that use conventional structures, such as LeNet, Resnet, and Alexnet, which all use the same database. The results showed that our model is superior in training and testing accuracy. Our model uses fewer layers (4 layers only) than traditional CNN structures, which may reach a hundred layers. Hence, the suggested model is faster in implementation and does not need large storage space, as it has an area of 23 MB, while the traditional structures reach more than 100 MB. Tables 3 summarize the results of comparing our presented model with some previous models. Figure 5 illustrates the confusion matrix for the test data.

Table 3. Compared to previous models

Paper	Model	No. of layer	Dataset	No. of images	No. of classes	Accuracy
Prajwala <i>et al.</i> [21]	LeNet	5	PlantVillage	18,160	10	94%
Shengyi Zhao <i>et al.</i> [9]	Resnet50	50	PlantVillage	18,340	10	96%
Mohit Agarwal <i>et al.</i> [5]	CNN	8	PlantVillage	10,000	10	91%
Iftikhar Ahmad <i>et al.</i> [24]	Inceptionv3	42	PlantVillage	2,364	6	93%
Our Model	Special model	4	PlantVillage	11,000	10	96%

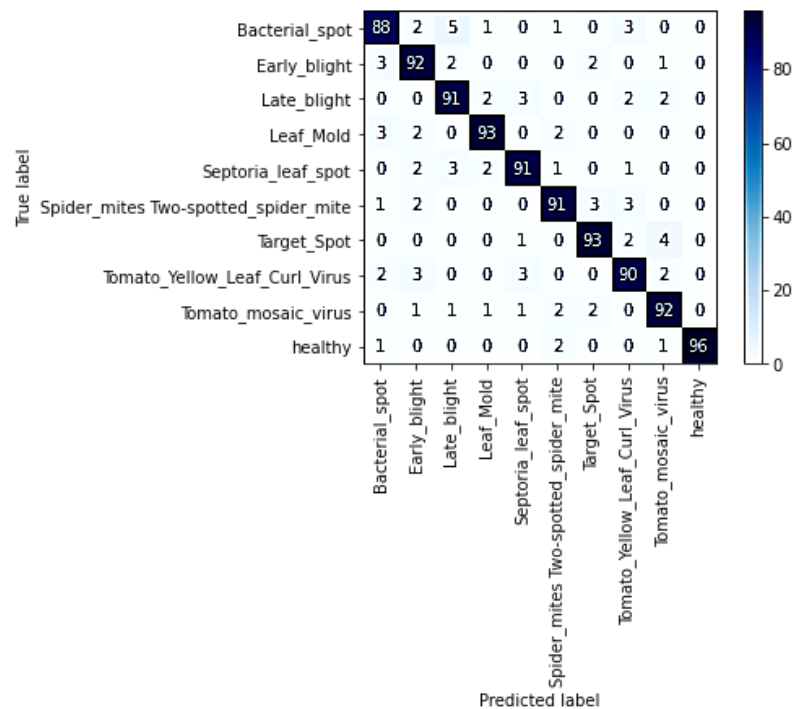


Figure 5. Confusion matrix on the test data

5. CONCLUSION

The tomato crop is one of the most abundant and famous vegetables in most countries. Tomato leaf diseases cause considerable losses in production, as millions of tons are lost annually. This proposed model based on CNN offers the possibility of early detection of tomato leaf diseases, thus contributing to maintaining production and increasing yield. The most common diseases were classified using the PlantVillage database, which contains 11,000 images of ten categories. The presented model also shows its superiority over the traditional methods, as it obtained an accuracy of 96%. Future work will improve the model's accuracy, test it on other crops, and convert it into a mobile application that can be used easily.




REFERENCES

- [1] M. Altalak, M. A. Uddin, A. Alajmi, and A. Rizg, "A Hybrid Approach for the Detection and Classification of Tomato Leaf Diseases," *Applied Sciences (Switzerland)*, vol. 12, no. 16, p. 8182, Aug. 2022, doi: 10.3390/app12168182.
- [2] A. H. H. Alasadi, E. Q. Anduljalil, and A. H. Khaleel, "Leaf Recognition based on Neural Network Feed- Forward and Support Vector Machine Classifiers," vol. 6, no. February, pp. 92–99, Feb. 2017.
- [3] *Agricultural value chain study in Iraq*. FAO, 2021.
- [4] S. Z. M. Zaki, M. A. Zulkifley, M. Mohd Stofa, N. A. M. Kamari, and N. A. Mohamed, "Classification of tomato leaf diseases using mobilenet v2," *IAES International Journal of Artificial Intelligence*, vol. 9, no. 2, pp. 290–296, Jun. 2020, doi: 10.11591/ijai.v9.i2.pp290-296.
- [5] M. Agarwal, A. Singh, S. Arjaria, A. Sinha, and S. Gupta, "ToLeD: Tomato Leaf Disease Detection using Convolution Neural Network," *Procedia Computer Science*, vol. 167, pp. 293–301, 2020, doi: 10.1016/j.procs.2020.03.225.
- [6] N. Khasawneh, E. Faouri, and M. Fraiwan, "Automatic Detection of Tomato Diseases Using Deep Transfer Learning," *Applied Sciences (Switzerland)*, vol. 12, no. 17, p. 8467, Aug. 2022, doi: 10.3390/app12178467.
- [7] A. Bhujel, N. E. Kim, E. Arulmozhi, J. K. Basak, and H. T. Kim, "A Lightweight Attention-Based Convolutional Neural Networks for Tomato Leaf Disease Classification," *Agriculture (Switzerland)*, vol. 12, no. 2, p. 228, Feb. 2022, doi: 10.3390/agriculture12020228.
- [8] I. Adeoye, O. Aderibigbe, I. Amao, F. Egbekunle, and I. Bala, "Tomato Products' Market Potential and Consumer Preference in Ibadan, Nigeria," *Scientific Papers Series : Management, Economic Engineering in Agriculture and Rural Development*, vol. 17, no. 4, pp. 9–16, 2017.
- [9] S. Zhao, Y. Peng, J. Liu, and S. Wu, "Tomato leaf disease diagnosis based on improved convolution neural network by attention module," *Agriculture (Switzerland)*, vol. 11, no. 7, p. 651, Jul. 2021, doi: 10.3390/agriculture11070651.
- [10] M. A. Jasim and J. M. Al-Tuwaijari, "Plant Leaf Diseases Detection and Classification Using Image Processing and Deep Learning Techniques," in *Proceedings of the 2020 International Conference on Computer Science and Software Engineering, CSASE 2020*, Apr. 2020, pp. 259–265, doi: 10.1109/CSASE48920.2020.9142097.
- [11] M. Hussein and A. H. Abbas, "Plant Leaf Disease Detection Using Support Vector Machine," *Al-Mustansiriyah Journal of Science*, vol. 30, no. 1, pp. 105–110, Aug. 2019, doi: 10.23851/mjs.v30i1.487.
- [12] R. A. Al-ani, M. A. Adhab, S. A. H. Hamad, and S. N. H. Diwan, "Tomato yellow leaf curl virus (TYLCV), identification, virus vector relationship, strains characterization and a suggestion for its control with plant extracts in Iraq," *African Journal of Agricultural Research*, vol. 6, no. 22, pp. 5149–5155, Nov. 2011.
- [13] R. I. Hasan, S. M. Yusuf, and L. Alzubaidi, "Review of the state of the art of deep learning for plant diseases: A broad analysis and discussion," *Plants*, vol. 9, no. 10, pp. 1–25, Oct. 2020, doi: 10.3390/plants9101302.
- [14] N. Sabri, N. S. Kassim, S. Ibrahim, R. Roslan, N. N. A. Mangshor, and Z. Ibrahim, "Nutrient deficiency detection in maize (*Zea mays* L.) leaves using image processing," *IAES International Journal of Artificial Intelligence*, vol. 9, no. 2, pp. 304–309, Jun. 2020, doi: 10.11591/ijai.v9.i2.pp304-309.
- [15] S. Ashok, G. Kishore, V. Rajesh, S. Suchitra, S. G. G. Sophia, and B. Pavithra, "Tomato Leaf Disease Detection Using Deep Learning Techniques," in *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, Jun. 2020, pp. 979–983, doi: 10.1109/icces48766.2020.9137986.
- [16] P. Dhandapani and A. Varadarajan, "Multi-Channel Convolutional Neural Network for Prediction of Leaf Disease and Soil Properties," *International Journal of Intelligent Engineering and Systems*, vol. 15, no. 1, pp. 318–328, Feb. 2022, doi: 10.22266/IJIES2022.0228.29.
- [17] R. Karthik, M. Hariharan, S. Anand, P. Mathikshara, A. Johnson, and R. Menaka, "Attention embedded residual CNN for disease detection in tomato leaves," *Applied Soft Computing Journal*, vol. 86, p. 105933, Jan. 2020, doi: 10.1016/j.asoc.2019.105933.
- [18] E. K. Elsayed and M. Aly, "Hybrid between ontology and quantum particle swarm optimization for segmenting noisy plant disease image," *International Journal of Intelligent Engineering and Systems*, vol. 12, no. 5, pp. 299–311, Oct. 2019, doi: 10.22266/ijies2019.1031.30.
- [19] G. B. Singh, R. Rani, N. Sharma, and D. Kakkar, "Identification of Tomato Leaf Diseases Using Deep Convolutional Neural Networks," *International Journal of Agricultural and Environmental Information Systems*, vol. 12, no. 4, pp. 1–22, Sep. 2021, doi: 10.4018/ijaeis.20211001.0a3.
- [20] R. M. J. Al-Akkam and M. S. M. Altaei, "Plants Leaf Diseases Detection Using Deep Learning," *Iraqi Journal of Science*, vol. 63, no. 2, pp. 801–816, Feb. 2022, doi: 10.24996/ijis.2022.63.2.34.
- [21] P. Tm, A. Pranathi, K. Saiashritha, N. B. Chittaragi, and S. G. Koolagudi, "Tomato Leaf Disease Detection Using Convolutional Neural Networks," *2018 11th International Conference on Contemporary Computing, IC3 2018*, 2018, doi: 10.1109/IC3.2018.8530532.
- [22] A. Kumar and M. Vani, "Image Based Tomato Leaf Disease Detection," Jul. 2019, doi: 10.1109/ICCNT45670.2019.8944692.
- [23] B. Luna-Benoso, J. C. Martinez-Perales, and J. Cortes-Galicia, "Tomato disease detection by means of pattern recognition," *International Journal of Computing and Optimization*, vol. 7, no. 1, pp. 35–45, 2020, doi: 10.12988/ijco.2020.9823.
- [24] I. Ahmad, M. Hamid, S. Yousaf, S. T. Shah, and M. O. Ahmad, "Optimizing pretrained convolutional neural networks for tomato leaf disease detection," *Complexity*, vol. 2020, pp. 1–6, Sep. 2020, doi: 10.1155/2020/8812019.
- [25] M. E. H. Chowdhury *et al.*, "Automatic and Reliable Leaf Disease Detection Using Deep Learning Techniques," *AgriEngineering*, vol. 3, no. 2, pp. 294–312, May 2021, doi: 10.3390/agriengineering3020020.




- [26] H. Kibriya, R. Rafique, W. Ahmad, and S. M. Adnan, "Tomato Leaf Disease Detection Using Convolution Neural Network," in *Proceedings of 18th International Bhurban Conference on Applied Sciences and Technologies, IBCAST 2021*, Jan. 2021, pp. 346–351, doi: 10.1109/IBCAST51254.2021.9393311.
- [27] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," *Computers and Electronics in Agriculture*, vol. 161, pp. 272–279, Jun. 2019, doi: 10.1016/j.compag.2018.03.032.

BIOGRAPHIES OF AUTHORS






Mohammed Hussein Najm    holds a B.Sc. degree in Computer Science from Tikrit University in 2018. He is a master's student in computer science and mathematics at Tikrit University. His research interests AI and image processing. He can be contacted at email: mohammed.h.najm22ms@st.tu.edu.iq.



Salwa Khalid Abdulateef    received a B.Sc. degree in Computer Science from Basra University, Iraq. An MS degree in Computer Science from Basra University, Iraq. Currently, she is working lecturer in the Department of computer science at Tikrit University. Her research interests include computer vision, image processing, information hiding, and machine learning. She can be contacted at email: khalid.salwa@tu.edu.iq.



Abbas H. Hassin Alasadi    is a Professor at the Information Technology College at Basra University. He received his Ph.D. from the School of Engineering and Computer Science / Harbin Institute of Technology, China. He spent more than ten years as Associate Professor at different Universities abroad in his current position. His research interests include Medical Image processing, Biometrics, Information retrieval, and Human-computer interaction. His research work has been published in various international journals and conferences. Dr. Abbas is an active reviewer in many computer science and software engineering journals. He is one of ACIT, UJCS, SIVP, IJIMAI, and IJPRAI members. He can be contacted at email: abbas.hassin@uobasrah.edu.iq.