ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

HARNESSING MACHINE LEARNING TECHNIQUES TO DIAGNOSE TOMATO PLANT DISEASES

S. Aruna

Department of Computer Application, Sona College of Technology Salem, TN, India

R. Abinaya

Department of Computer Applications, Sona College of Technology Salem, TN, India

A. Vanithasri

Department of Computer Applications, Sona College of Technology, Salem, TN, India

A. Vasanthakumar

Department of Computer Applications, Sona College of Technology, Salem, TN, India

ABSTRACT—

Agriculture is basic in the development of any nation and also contributes to economic stability. Tomato production constitutes a significant aspect of agriculture in Tamil Nadu and India. It is facing yield and quality issues. However, disease in crops lowers the health of tomato leaves and the productivity of a tomato plant. This study has focused on an approach to develop a Convolutional Neural Network (CNN) model improved with data augmentation for detecting diseases in tomato leaves. It identifies and classifies multiple diseases on tomato leaves accurately at 89% with over 35 epochs during training. All validation metrics support the strength and effectiveness of the model: AUC score, precision, and recall. The software solution also serves both detection and practical application by suggesting appropriate chemicals for recognized diseases such as early blight, septoria leaf spot, and powdery mildew. This function makes it easier to manage the disease as a whole and a loss on crops is not so severe. It was trained on images of tomato leaves, some of which were obtained from the Plant Village repository in order to have variation in the datasets to make them practical for use in the real world. The research underlines the possibility of technology integration in agricultural practices and proposes an effective method of focusing on the prevention of disease outbreak and its link to appropriate management activities. This adds crop productivity but also promotes sustainable agricultural practices which enhances economic and environmental stability.

Keywords— Convolutional Neural Network, Data Augmentation, Tomato Leaf Disease Detection.

1. INTRODUCTION

iagnosis plays a very important role in crop production and management. Tomatoes are among the most cultivated crops, and they play an important role in the economy. Timely and correct identification of leaf diseases is very useful. Plants show general symptoms of leaf diseases, and early treatment saves money and enhances productivity. This is especially important in tomato production because it maximizes output, reduces losses, and enhances overall crop quality. But traditional mechanical diagnosis has been the farmer's mainstay, which requires skill and time is not always readily available today. Image processing-based automatic systems and machine learning developments provide a promising alternative. These technologies allow for rapid and accurate disease detection without special skills, thus enhancing modern agriculture. The work deals with the detection of diseases in tomato leaves through the use of Convolutional Neural Networks (CNN) as part of automation. CNN models are more suited for image classification, and thus they are the most suitable models to be utilized for disease pattern recognition in tomato leaves. The model will respond less to

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

the changes in the dataset as the model becomes stronger, leading to better accuracy in the classification of diseases. As they are widespread and play a significant role in tomato crop health, three of the most notable tomato leaf diseases, i.e., powdery mildew, early blight, and septoria leaf spot, are under consideration here. It is a product of Alternaria solani fungus and creates brown, concentric rings on older leaves, defoliates leaves, creates weaker plants, and smaller fruit sets. Septoria leaf spot is a leaf-spotting disease that creates dark, small, round spots with grey centres on leaves, resulting in premature leaf drop. It grows fast under wet conditions kills photosynthesis and drastically reduces yields. It is supported by warm and moist weather conditions, therefore, it is critical to identify it early. Powdery Mildew is a disease caused by fungi and appears in the form of white powdery spots on the leaf's surface, preventing photosynthesis and plant health. It infects rapidly, particularly under hot dry weather, and affects the overall growth of tomatoes as well as tomato quality. The model trained with CNN diagnoses such diseases and gives actionable advice, including pesticide recommendations, to farmers to help them control the disease. The model incorporates data augmentation methods and metrics like accuracy, Area Under the Curve (AUC) score, precision, and recall for checking successful identification of disease. Application of such technology in agriculture can increase productivity by a significant amount, decrease loss, and ensure healthy growth of plants. The model is trained for 35 epochs and has an accuracy of 89%, which represents its performance under real agricultural scenarios. The system enables sustainable agriculture and effective monitoring of crop health through the incorporation of disease classification with prescriptive treatment recommendations.

2. LITERATURE REVIEW

While preserving data privacy and localization, a federated learning framework can provide cooperative training. [1] Traditional post-processing in disease detection lacks precision and interpretability, hindering effective disease management. The research methodology integrates IoT, deep learning, and Vector Autoregressive Moving-Average with eXogenous regressors (“VARMAx)-CNN-GAN Integration” mechanisms for early detection and management of tomato leaf diseases. This approach uses a fusion of CNNs, Generative Adversarial Networks (GANs), and autoencoders for accurate disease classification and saliency mapping for better disease localization and severity estimation. Aiming at the complex background of the early period of tomato diseases and pests image objects in the natural environment, an improved object detection algorithm based on You Only Look Once Version 3 (YOLOv3) for early real-time detection of tomato diseases and pests was proposed. [2] In order to preserve the excellent quality and receptive field and enhance the capacity for small object recognition, an expanded convolution component is employed in place of the convolution layer in the backbone network to target the complex backdrop of tomato diseases and insect images under natural conditions. Second, the detection problem of mutually obscured tomato disease and pest objects is resolved in the detection network by retaining the obscured objects based on the size of the candidate box intersection ratio (an IOU) and linear distortion level of confidence predicted by multiple grids.

[3] An image vision-based method for classification and identification of diseases of tomato leaves was described by N. Z. Enxu Zhang et al. To make up for the unbalanced dataset, oversampling is used to expand the dataset and a piecewise linear conversion method is used to enhance the disease feature information in the images. To construct a lightweight model referred to as the Linear Discriminant Analysis Modified Network (LDAMNet), this paper then introduces the Digital to Analogue Converter (DAC) Block, a dual attention mechanism convolutional block. To enhance the algorithm's feature acquisition ability, the DAC Block innovatively handles the channel information and spatial information of input images with Hybrid Channel Attention (HCA) and Coordinate Attention (CSA), respectively. [4] Miss Tejswini S. Danwadkar et al., By applying technology and machine learning algorithms, we suggested a successful method for identifying plant diseases. Among

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

the most devastating losses to crop farms is the transmission of disease from infected plants to all other healthy plants. Additionally, if not detected in their initial stages, these diseases can have the power to affect the entire organization and spread like fire. Today, detection methods of plant diseases assist us in detecting plant diseases in a large area of crops in a cost-effective way and assist us in detecting disease plants in their very initial phases.

[5] Tomato leaves with various illnesses will be examined by D. N. C. Adnan Pipawala et al. Using a CNN-based architecture called VGG16, research has been conducted on a dataset of tomato leaves that comprises eleven classes, helping to provide accuracy. The photos used to feed the model are retrieved from two dataset sources: Plant Village and Tomato Leaf Diseases. [6] The development and optimization of lightweight CNN-(RNN) Recurrent Neural Network models for effective image-based diagnosis of tomato plant diseases is the main objective of this study. The need to develop effective and user-friendly disease detection techniques in agricultural environments, especially in resource-limited regions, is the driving force behind this aim. To fulfill this aim, we will focus on a few major areas. Firstly, this study explores various CNN and RNN hybrids to extract geographical and sequential information from plant images.

[7] A. R.-M. Antonio et al., suggest a convolutional neural network-based model to detect and classify tomato leaf diseases from a public dataset and augment it with other images captured in the fields of the nation. Generative adversarial networks were employed to prevent overfitting by generating samples with the same features as the training set. The outcomes indicate that the model proposed is highly efficient in the detection and classification of tomato leaf diseases. [8] The aim of the research is to classify among 10 different categories of tomato plant leaves using the proposed novel TomConv model which deploys an improved Convolutional Neural Network (CNN). For this purpose, the publicly available dataset called Plant Village comprising more than 16000 images of tomato leaves, both diseased and healthy was used for experimentation purposes. The tomato leaf images were pre-processed to reduce the size to 150×150 dimensions. It constitutes four layered CNN followed by a max pooling layer. The model splits the corpus into training and validation datasets in an 80:20 ratio, and is trained under 105 epochs for tomato leaf images.

B. H. Mohamed Bouni, et al., utilized the [9] pre-trained deep neural networks and automation, which are key to this technique. Digital image processing can be employed to detect plant disease. Deep learning has significantly improved digital image processing in recent years and has overtaken traditional methods. This paper detects tomato leaf disease through a deep convolutional neural network (CNN) and transfer learning. CNN's core consists of AlexNet, ResNet, VGG-16, and DenseNet. Adam and Rms Prop optimization techniques analyze these networks' comparative performance, showing that the DenseNet model with the Rms Prop optimization method has the highest results. To diagnose tomato leaf diseases, this research [10] proposes a new method that integrates the vision transformer (ViT) and deep neural networks (DNN). In consideration of the model's performance and interpretability, this research seeks to build a system for the diagnosis of tomato leaf diseases. In order to make its extraction of significant picture data more efficient, the new method incorporates L1-norm attention with a better multi-head self-attention mechanism. ViT's learning ability of images as a sequence of patches combines with DNN's classification and interpretation abilities.

Deep learning has been explored in several studies for detecting plant diseases. YOLOv3 and GAN-based models have made better accuracy rates. VGG16 was reported to have excellent performance but with high computational costs. Our effort, with an accuracy of 89%, is a compromise between computational expense and real-world relevance using data augmentation and hyperparameter tuning. This work differs from current research in complex models as it focuses more on usability and practicality.

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

3. PROPOSED METHODOLOGY

The proposed study makes use of a dataset obtained from Kaggle, consisting of 2,800 publicly available images of tomato leaves, divided into three different disease classes and one healthy class. In this experiment, the interest is in four individual classes: Early Blight, Septoria Leaf Spot, Powdery Mildew, and Healthy, with each class having 700 images for the balanced dataset. This balanced allocation allows for serious training and testing of the model, resulting in very accurate identification of diseases for all chosen classes.

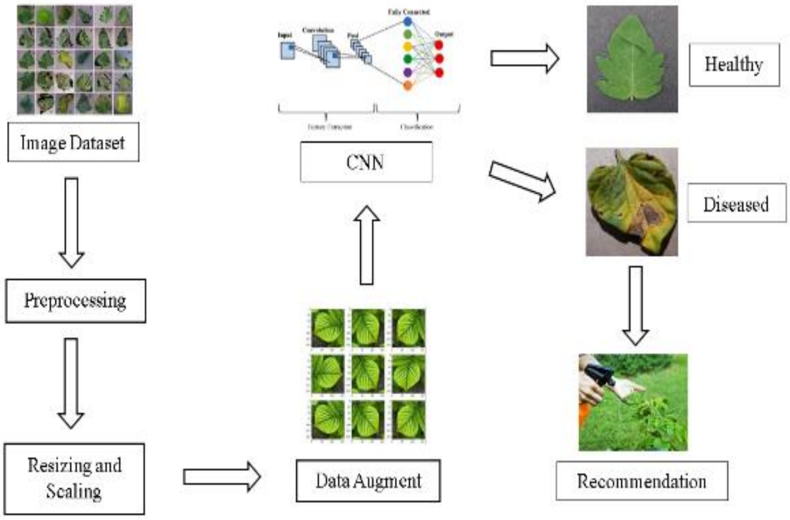

Figure 1. Tomato Disease Diagnose Process

The process starts with a set of tomato leaf pictures, both healthy and infected ones. These images are the basic input for model training and testing. The data is varied in order to guarantee correct disease recognition. Correct labelling is performed in order to place images into various classes.

The gathered images are cleaned, noise-reduced, and normalized. This is to improve image quality and eliminate unwanted distortions. It is aimed at ensuring the model has good-quality inputs to analyse. Preprocessing further entails transforming images to a processing-ready format. The preprocessing is done by using the Sequential model for resizing and rescaling input images prior to feeding them into a neural network.

- **tf.keras.layers.Resizing(224, 224) :** This resizes the input images to a fixed 224×224 pixel size. Images in datasets are usually of different sizes, and resizing them makes them all of the same size, which is what the proper training and best performance of the model requires.
- **tf.keras.layers.Rescaling(1.0/255) :** Rescales the pixel values from their native range of [0, 255] to [0, 1] to normalize them. Normalization stabilizes learning, improves convergence, and increases model accuracy.

3.1 Data Augmentation

3.1.1 Random Contrast Adjustment

Contrast Adjustment alters the contrast between light and dark regions of an image. Algebraically, contrast adjustment can be represented as:

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

$I' = \alpha (I - \mu) + \mu$

3.1.2. Random Flipping (Horizontal & Vertical)

Flipping changes the image to a mirror reflection along the vertical or horizontal axis. The change can be shown as:

For horizontal flipping:

$I'(x, y) = I(W - x - 1, y)$

For vertical flipping:

$$I'(x, y) = I(x, H - y - 1)$$

3.1.3. Random Zooming

Zooming makes an image larger or smaller by scaling it. The transformation is provided by:

$$I'(x, y) = I(S_x . x, S_y . y)$$

3.1.4. Random Rotation

Rotation changes an image by turning it about its centre. The coordinates of the new pixels (x', y') after rotation by an angle θ are given by:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} \begin{bmatrix} x - x_c \\ y - y_c \end{bmatrix} + \begin{bmatrix} x_c \\ y_c \end{bmatrix}$$

3.2 CNN

CNN is structured so that images can be categorized into four groups. It consists of layers of preprocessing, convolution, dense, and output.

Preprocessing - Generalisation is enhanced through the use of random augmentations like flipping, zooming, rotation, and contrast change besides rescaling and normalising pixel values in preprocessing layers.

Convolution - ReLU activation and max pooling are used following each convolution in order to reduce spatial dimensions because the convolutional layers collect image features with a 3×3 filter starting with 120 filters and lowering to 60.

Flatten - To prepare the data for the dense layers, the flattening layer converts the 2D feature maps to a 1D array.

Dense - Each of the 800, 300, and 100 neurons in the model's dense layers uses ReLU activation. At rates of 40%, 30%, and 20%, dropout layers are used to control model complexity and avoid overfitting.

Output - Four neurons with softmax activation make up the output layer, which provides the probability distribution over the four potential classes.

4. RESULT AND DISCUSSION

Model performance was also compared to training and validation accuracy over 35 epochs with steady improvement. Accuracy trends tell us about effective learning and are a generalization measure.

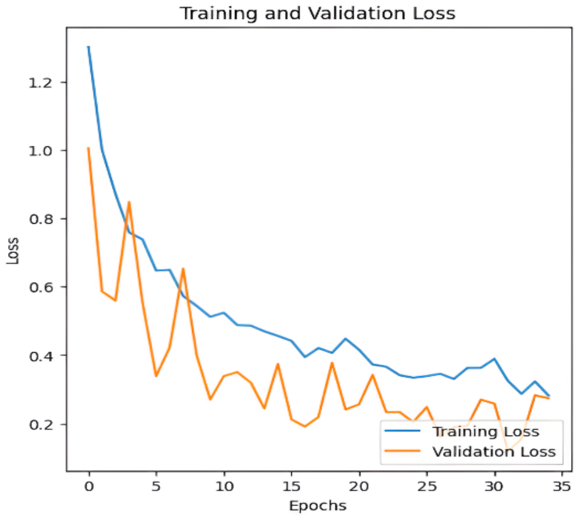

Figure 2. Training and validation accuracy

The Figure 2 shows training and validation accuracy at 35 epochs which increases as the model learns. Validation accuracy fluctuates early on but eventually reaches a stable level of 89%, slightly above training accuracy. The model generalizes very well with minimal overfitting.

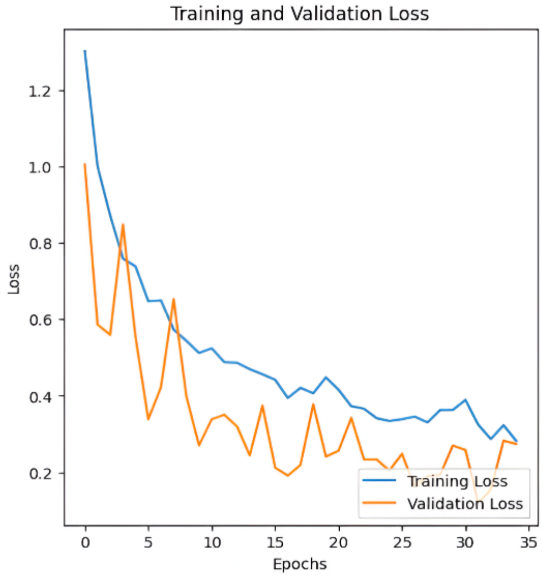

Figure 3. Training and validation Loss

The figure 3 shows Training and validation loss curves go down for 35 epochs, indicating a good learning. The validation loss is oscillating but is less than the training loss, indicating good generalization. This trend overall also shows that the model is learning and not overfitting the dataset.

	precision	recall	f1-score	support
Septoria_leaf_spot	0.93	0.71	0.81	140
Early_blight	0.76	0.97	0.85	140
powdery_mildew	0.99	0.89	0.94	140
healthy	0.94	0.99	0.97	140
accuracy			0.89	560
macro avg	0.91	0.89	0.89	560
weighted avg	0.91	0.89	0.89	560

Figure 4. Classification Report of the Model

A classification report (figure 4) shows F1 score, recall, and precision for four classes, with an overall accuracy of 89%. The best F1-scores are of both the classes of "Healthy" and "powdery mildew", which suggest better performance in classification. The macro and weighted averages suggest fair performance for this model across all classes.

The AUC (figure 5) scores of these scores in the graph represents how well the model is able to distinguish classes. The highest AUC of 0.99 belongs to the "healthy" class, followed by "powdery mildew" and "Early blight" (0.95 and 0.93), while the lowest value (0.85) is given for "Septoria leaf spot." Meantime, higher AUC values indicate better classification capabilities for the respective classes.

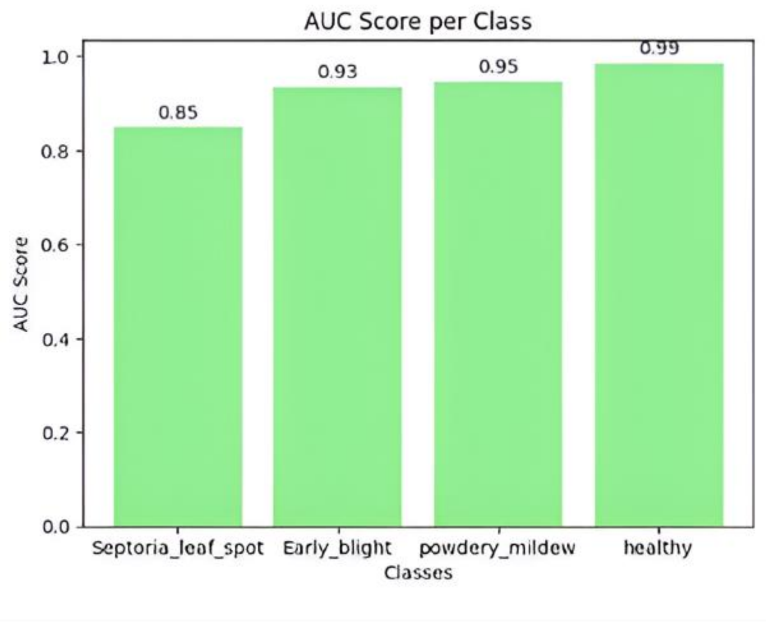

Figure 5. AUC Score per Class

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

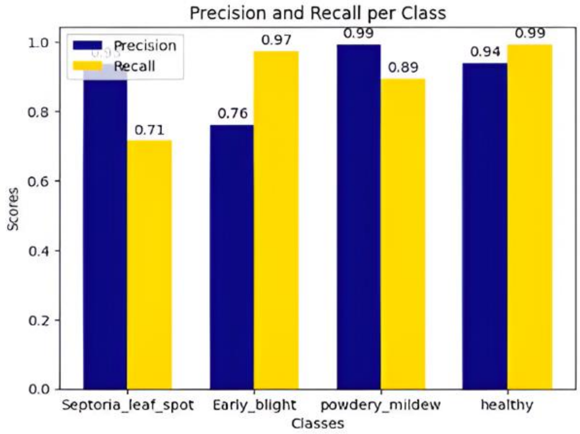

Figure 6. Precision and Recall per Class

In the above graph, precision and recall for each class compared — blue for precision and yellow for recall. Powdery mildew" and "healthy" classes achieve the highest precision (0.99 and 0.94) and recall (0.89 and 0.99), showing very good classification performance. Early blight has high recall (0.97) but lower precision (0.76), and Septoria leaf spot has the lowest recall (0.71) suggesting that it is misclassified more.

Precision (Positive Predictive Value):

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall (Sensitivity or True Positive Rate):

$$\text{Recall} = \frac{TP}{TP+FN}$$

Act Label: Early_blight
Pred Label: Early_blight
Use Chlorothalonil or Copper-based fungicides. Apply every 7-10 days.

Figure 7. Early Blight Disease Affected

In the figure 7 the model correctly classified the leaf as having Early Blight and it suggests using Chlorothalonil or Copper-based fungicides every 7-10 days for treatment.

Figure 8. Powdery mildew Disease Affected

The above figure 8 shows the model correctly identified the leaf as having Powdery Mildew and recommends using sulfur-based fungicides or potassium bicarbonate every 7-10 days for treatment.

Figure 9. Septoria leaf spot Disease Affected

In the figure 9 it shows that the leaf is affected by the Septoria Leaf spot and also it suggests using Mancozeb or copper-based fungicides at the first sign of symptoms, with weekly reapplication.

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

Figure 10. Healthy Leaf

The above image shows that the model correctly classified the leaf as Healthy and no pesticide recommendation is needed.

5. CONCLUSION AND FUTURE SCOPE

This work successfully classifies various tomato leaf diseases with machine learning methods and reports an accuracy of 89% using our CNN model. The work points out the merits of various methods of classification and proves that the proposed method is robust in distinguishing disease types. Besides classification, the model provides suitable pesticides as recommendations, thereby being a handy tool for effective disease management.

Further extensions to the proposed system would be development of real time image capturing system which would allow farmers to use camera to capture image of the tomato leaves on real time and achieve real time detection of disease. Mobile apps and IoT-based smart farming solutions allow for near-real-time analysis, making disease detection easier and more accessible.

6. ACKNOWLEDGEMENT

We would like to thank our guide, Prof. ARUNA S., with the deepest gratitude for her constructive feedback, valuable guidance, and encouragement in the learning process of this report. Her assistance has contributed significantly to our success.

We are also highly grateful to our Department Head, Prof. PADMA., for her provision of the required facilities and assistance in completing this work successfully. We also thank the staff of the Master of Computer Applications Department sincerely for their considerable help.

REFERENCES

6. S. R. K. Vengaiah, "Tomato leaf disease detection and management using," Journal of King Saud University - Science, p. 9, 2024.
7. J. L. Xuewei Wang, "An efficient deep learning model for tomato disease detection," Plant Methods, p. 18, 2024.
8. N. Z. Enxu Zhang, "A lightweight dual- attention network for tomato leaf disease identification," Frontiers in Plant Science,

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

9. p. 18, 2024.
10. D. P. S. Miss Tejswini S. Danwadkar, "Tomato Plant Leaf Disease Detection Using Convolutional Neural Network," International Journal on Recent and Innovation Trends in Computing and Communication, p. 9, 2024.
11. D. N. C. Adnan Pipawala, "An Enhanced Tomato Plant Disease Detection And Classification Methodology Using Cnn," International Journal of Advanced Research in Computer and Communication Engineering, p. 7, 2024.
12. I. A. Masoud Shakiba, "Tomato disease detection with lightweight recurrent and convolutional deep learning models for sustainable and smart agriculture," Frontiers in Sustainability, p. 6, 2024.
13. A. R.-M. Antonio Guerrero-Ibañez, "Monitoring Tomato Leaf Disease through Convolutional Neural Networks," Electronics, p. 15, 2023.
14. K. K. Preeti Basera, "TomConv: An Improved CNN Model for Diagnosis of Diseases in," ScienceDirect, p. 9, 2023.
15. B. H. Mohamed Bouni, "Impact of Pretrained Deep Neural Networks for Tomato Leaf Disease Prediction," Journal of Electrical and Computer Engineering, p. 11, 2023.
16. H. K. Mamta Tiwari, "Tomato Disease Detection Using Vision Transformer with Residual L1-Norm," International Journal of Intelligent Engineering and Systems, p. 10, 2023.
17. Suji Priya, J., Mohana Priya, H., Iyswarya, M., & Kiruthika, K. (2024, June). Prediction of Pathogen Causing Rice Plant Disease and Recommendation using Enhanced Machine Learning Technique. In 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC) (pp. 643-648). IEEE.
18. Thamizhamuthu, R., S. Parameswari, Suhana Nafais A, Pankaj Negi, A. Prakash, and Badugu Suresh. "AI-Based Detection and Prediction of Magnetic Interference in Drone Magnetometer Systems Using Multi-Sensor Data with Edge Computing." (Nov 15,2024) (2024).