PEST RECOGNITION AND CLASSIFICATION SYSTEM USING CNN AND TRANSFORMER

Dr. D. Sasireka
Assistant Professor, Mepco Schlenk Engineering College, Sivakasi

MM. Janani, T. Josika & V. Kiruthika
Student, Department of Computer Science and Engineering
Mepco Schlenk Engineering College, Sivakasi

DOI: doi.org/10.34293/shanlax.9789361631474.ch006

Abstract

Agricultural productivity and crop quality are critically threatened by a wide array of pest species, which inflict extensive damage, thereby causing substantial economic losses, particularly in our country's agrarian sector. Recognizing the urgent need for an effective pest management system, this paper introduces Pest- Conformer, a state-of-the-art hybrid model designed to address these challenges. Pest-Conformer leverages the combined strengths of Convolutional neural networks (CNNs) and transformers to create a multi-class recognition model that excels in distinguishing various crop pests. This architecture uniquely addresses the complexities of high inter-class similarity and intra- class morphological variation commonly observed in pest species, both of which traditionally hinder classification accuracy in agricultural environments. Pest-Conformer incorporates a fine- grained classification module, employing weakly supervised learning to identify and prioritize key discriminative feature points at different pyramidal levels. This feature selection process ensures that only the most critical visual attributes are used, which are then analyzed through a graph convolutional network, enhancing model robustness and classification accuracy. The development and implementation of this model are not only a technological advancement but also an economic imperative, as it facilitates proactive pest monitoring and control, potentially reducing yield losses and mitigating the economic strain on our agricultural economy. Given the urgency to safeguard agricultural resources and bolster productivity, Pest-Conformer represents a critical step forward in improving crop resilience and supporting sustainable agricultural practices.

***Index Terms**: Bolster, Conformer, Convolutional network, CNN, Features, Module, Morphology, Robustness*

I. Introduction

Agriculture forms the backbone of our nation, supporting millions of livelihoods and sustaining our economy. This sector is not only essential for food security but also a primary contributor to our GDP and a cornerstone of rural development. However, the prevalence of crop pests-organisms that seriously harm crops, lowering yields and compromising quality, is one of the biggest threats to agricultural productivity. Without effective and timely pest management solutions, these pests can lead to devastating economic losses and compromise the stability of our food supply.

Traditional methods relied heavily on the expertise of agricultural specialists, who often struggled to keep up with the vast number of pest species and their

irregular distribution in fields. This reliance on human knowledge resulted in slow and inefficient pest control measures, which could lead to substantial crop losses.

Historically, farmers faced substantial challenges in accurately identifying and managing pests, especially given the high number of pest species and their often-irregular distribution across fields. Traditional pest identification methods relied heavily on agricultural experts, who were tasked with visually identifying pests in a field. This approach, while valuable, was slow, resource-intensive, and prone to error due to factors such as pests' similar appearances and fluctuating environmental conditions like lighting and background interference. For instance, a farmer might observe early signs of damage—such as wilting or discolored leaves—but be unable to determine the specific pest responsible, delaying treatment and allowing pests to spread further.

In response to these limitations, the Pest-Conformer an advanced hybrid model was designed to revolutionize pest detection and classification. Pest-Conformer integrates convolutional neural networks (CNNs) with transformer technology, offering an innovative approach that vastly improves pest identification accuracy and efficiency over traditional methods. The model leverages CNNs for their strength in extracting spatial features from pest images and combines them with transformers to capture complex relationships between these features across different contexts. This hybrid architecture addresses the high inter-class similarity and intra-class variation of pest morphology—challenges that often hinder classification in complex agricultural environments.

Furthermore, Pest-Conformer includes a fine-grained classification module that utilizes weakly supervised learning to isolate and prioritize discriminative features across pyramidal levels. By applying these features within a graph convolutional network, the model enhances classification precision, even in challenging scenarios. This development is not only a technological advancement but also an economic imperative for our nation, as early pest detection and precise identification will enable proactive interventions, reducing crop losses and helping to safeguard agricultural productivity. In summary, Pest-Conformer aims to empower farmers with an effective, scalable tool for early pest monitoring, reinforcing agriculture's resilience and supporting sustainable practices essential to our country's growth. the hybrid model employs a novel dual-path feature aggregation and weakly supervised learning for accurate pest classification. The system beats current approaches and is set to transform agricultural pest control with its sophisticated architecture

II. Related Work

Utku et al., assessed the research titled "A New Hybrid ConvViT Model for

Dangerous Farm Insect Detection" Together with the YOLOv5s model employed in this experiment, [1] the inherent capabilities of UAVs may offer a reliable solution for real-time pest detection, demonstrating a strong potential to optimize and improve agricultural output in a drone-based ecosystem. Through improved crop monitoring using unmanned aerial vehicles (UAVs) and improved agricultural pest detection and categorization, this study seeks to address the issues of crop diseases and pest infestations in agriculture.

Ashutosh Kumar Bhatt et al., in their study titled "Deep Transfer Learning based Light-weight Agricultural Pest Classification Model using Convolutional Neural Networks," examine transfer learning as a prominent strategy in agricultural pest detection, allowing pre-trained models to adapt to specific datasets through fine-tuning. Models such as (ResNet50), (Xception), and (DenseNet121) have been employed to extract relevant features from pest images, improving accuracy in classification tasks. [2] Transfer learning reduces the need for extensive datasets, often difficult to obtain in agricultural research. Among these, Xception demonstrated superior performance with 92.09% training accuracy and 83.35% testing accuracy, showcasing the importance of balancing model complexity with specialization for agricultural environments. This adaptability ensures that models maintain effectiveness even when applied in diverse conditions.

Khan et all., in their study "AI-Enabled Crop Management Framework for Pest Detection Using Visual Sensor Data." [3] A refined version of the YOLOv5 model is proposed that exceeds state-of-the-art performance and improves the performance of YOLOv5 with better results. It is evaluated on a dataset that was created by the author. Research on natural human-computer interaction includes the recognition of hand gestures.

Biswas et all., "Real time Gesture Recognition using Improved YOLOv5 Model." [4] A refined version of the YOLOv5 model is proposed that exceeds state-of-the-art performance and improves the performance of YOLOv5 with better results. It is evaluated on a dataset that was created by the author. Research on natural human-computer interaction includes the recognition of hand gestures. Gesture recognition research aims to teach computers to identify and respond to human hand motions. Research on natural human-computer interaction includes the recognition of hand gestures. The goal of gesture recognition research is to make computers able to identify and respond to human hand motions. The You Only Look Once (YOLO) deep learning model is mostly used for real-time object recognition in computer vision.

Through the improvement of the YOLOv5 model, the research made gesture recognition systems more efficient, accurate, and robust, setting the

stage for future innovation in human-computer interaction.

Zhang et al., in their study " A Pest Recognition Network Based on Attention Mechanism and Multi-Scale Feature Fusion". enhance pest recognition performance through multi-scale feature fusion.[5]The authors demonstrate improved robustness and accuracy in identifying various pests by integrating features at different scales. This method addresses the complexities involved in pest detection, providing a more reliable framework for agricultural applications Puspasari et al., in their research titled"EfficientNet-Based Sugarcane Disease Classification with Dual-Convolution Spatial Attention CBAM (EfficientNet-DCCBAM)." [6]. A deep learning network model that combines EfficientNetV2B0 and Dual Convolution Spatial Attention CBAM (DCSA-CBAM) is proposed in this paper to enhance the features of infected sugarcane leaf photos. The collection includes 1968 photos of sugarcane leaves taken in different parts of East Java. Destructive insect pests have long plagued agriculture, as seen in the case of sugarcane, a significant crop in Indonesia. These pests have been combated using traditional management techniques like pesticides and pest monitoring. Artificial intelligence (AI) technology makes it possible to recognize and categorize organisms in precision agriculture

Yuan Junchao collaborated on the researcher al., paper titled "Deep Learning-based Agricultural Pest and Disease Recognition" investigates the impact of deep learning on the identification of pests and diseases in agriculture.[7] It covers techniques like image processing, data selection, preprocessing, data augmentation, model selection, and transfer learning, which contribute to improved classification accuracy. Findings reveal that deep learning methods surpass traditional approaches, though issues such as data quality and overfitting still need addressing. The authors call for continues research to enhance model effectiveness. They conclude that adopting deep learning can significantly advance pest management and support sustainable agriculture practices.

In the study "Comparing Lightweight and Complex Models," Qixuan Huang et al, [8] Qixuan examined the distinctions between lightweight models and more intricate architecture. Lightweight models such as (Efficient Net) have been compared with more complex architectures like (exception) [3] to evaluate their suitability for practical use. Although both models offer high classification accuracy, Efficient Net is particularly well-suited for real-world applications due to its computational efficiency and faster training times. With an 88.26% accuracy, Efficient Net strikes a balance between performance and resource consumption, making it ideal for deployment in mobile applications or field devices. This comparison emphasizes that complexes do not always guarantee better results

and highlights the value of efficient architectures in agriculture

Rr. Patil L. V. et al., made significant contributions to the field through the study "Advances in Deep Learning for Agricultural Pest Management". [9] Recent advancements in pest detection leverage hybrid frameworks that combine CNN with handcrafted features to enhance accuracy. Models such as Faster R-CNN, InceptionV3, and Dense Net have achieved classification accuracy as high as 99.62%. These models are applied across multiple agricultural tasks, including pest monitoring, weed classification, and disease recognition, enabling more precise crop management. The combination of deep learning with traditional techniques ensures better

Manoj Lara et al. investigated "Automated Pest Classification and Infestation Detection using CNN and Transfer Learning Techniques," focuses on enhancing pest classification accuracy through CNNs and transfer learning [10]. The research emphasizes the role of CNNs as an effective tool for classifying agricultural image data. Image augmentation techniques were employed to overcome dataset limitations to improve model robustness. Incorporating transfer learning further boosted performance, achieving a 98% testing accuracy. By splitting the dataset in an 80/20 ratio for training and testing, the study demonstrated that combining CNNs with transfer learning can significantly improve the efficiency of pest detection.

Hu Haiyan et al., researched the topic in their study titled "Identification of Pests and Diseases Based on Cascaded Convolutional Neural Network," which aimed to detect pests and diseases in corn. It introduces a cascaded CNN model that leverages a double voting mechanism, integrating Alex Net and Inception architectures for improved accuracy in complex environments. A Siamese network [11] was also used to evaluate the severity of infestations by calculating Euclidean distances between pest samples. The model achieved a notable accuracy rate of 95.85%, outperforming models such as ResNet and VGG. Extensive testing validated the model's effectiveness in real-time agricultural applications.

Zhu Dingju et al., worked on the study titled "Crop Disease Identification by Fusing Multiscale Convolution and Vision Transformer," The MSCVT model combines CNNs for local feature extraction with Vision Transformers (ViT) for global attention [12]. This hybrid approach overcomes traditional CNN limitations by integrating a multiscale self-attention module. The model delivered impressive accuracy rates of 99.86% and 97.50% across tested datasets, emphasizing the value of combining global and local features for precise pest and disease identification. Additionally, the model's strong adaptability to smaller datasets makes it ideal for practical agricultural applications.

Xuqi Wang et al., presented their findings in Their study titled "Crop Pest

Detection by Three-Scale Convolutional Neural Network with Attention" introduces a multi-scale CNN model enhanced with attention mechanism for detecting pests in complex agricultural settings [13]. The model, incorporating spatial and channel attention, achieved a precision of 93.16%, surpassing methods like ICNN and VGG16. The enhanced feature extraction, especially for small pests in challenging environments, proved vital for its superior performance in real- world agricultural conditions.

Jia et all., in their study "Pest and Disease Detection Based on Data Augmentation." Pest species are detected using augment techniques such as rotating, flipping. [14]So that each species is detected and plant disease to their respective pest species type is identified.

Yingshu Peng et al., assessed the research titled "Comparison of CNN Models with Transfer Learning in the Classification of Insect Pests" assess various pre-trained CNN models using the IP-102 dataset for pest classification [15]. DenseNet-201, when fine-tuned, achieved the highest accuracy of 70%, outperforming models such as MobileNetV2, InceptionV3, and exception. The study underscores that transfer learning not only improves accuracy but also significantly reduces training time, making it a highly efficient approach for large-scale agricultural pest detection.

Wei et al.. evaluated the study entitled "Small sample and efficient crop pest recognition method based on transfer learning and data transformation." .**Error! Reference source not found.** the EfficientNet-B3 neural network to identify and distinguish between healthy crops and those that are not, as well as data enrichment technology to address the lack of data. EfficientNet-B3 neural network identified healthy and unhealthy crops, and data augmentation technology was employed to compensate for the lack of sufficient data.

Wang et al., in their article "Transfer Learning-Based Light weight CNN Model for Recognition of Pest in Citrus." A transfer learning and data transformation-based approach to crop pest recognition was proposed using CNN models such as Inception V3, VGG16, and ResNet as the backbone network.**Error! Reference source not found.** Transfer learning was used to improve the performance of the model. An approach based on transfer learning and data conversion was suggested to solve the problem of long training duration and vast sample requirements required by traditional image recognition models. It is based on CNN models such as ResNet, VGG16, and Inception.

Zhao et al., in their work "Crop Pest Recognition in Real-Time Agricultural Environment Using Convolutional Neural Networks by a Parallel Attention". The authors demonstrate an enhanced deep convolutional neural network that can identify crop pests more accurately in an actual agricultural environment. It has significant improvements compared to the past models in terms of accuracy and

real-time processing. Because of their severity and intensity, which put crop productivity in jeopardy, crop pests are an important worldwide agricultural problem. Their inefficiency and time-consuming characteristics, as well as their failures, make it difficult to inspire and utilize classical pest picture segmentation techniques. **Error! Reference source not found.** It is the preferred way of tackling the technological challenge of pest identification that deep learning methods have proved.

Liu, Y et al., discuss the role of attention mechanisms in pest identification in their paper "Attention Mechanisms in Pest Recognition." They show how attention mechanisms can enhance the performance of models by emphasizing significant features in pest images. **Error! Reference source not found.** Based on their work, it can be understood that incorporating attention mechanisms significantly improves pest detection system accuracy and thus their usability in real-world use.

Wang et al.'s paper titled "ASP-Det: Toward Appearance-Similar Light-Trap Agricultural Pest Detection and Recognition." **Error! Reference source not found.**. Appearance Similarity Pest Detection (ASPD) task, and propose two new measures to quantify the texture-similarity and scale-similarity issues. Computer vision-based automatic pest detection and recognition are popular these days in the smart agriculture sector but are plagued by an intolerable difficulty: separation difficulty between similar-targeted targets for 2D pest images.

III. Dataset Description

D0 Dataset

The dataset used in this project is D0(Xie et al, 2018). This dataset consists of nearly four thousand and five hundred pest images belonging to 40 different pest species captured under different lighting, angles, and backgrounds. These contain variations due to environmental factors and different pest postures. The images in this dataset are annotated to facilitate supervised learning. For D0, we divided this dataset into two subsets of training and validation. 80% for training and 20% for validate.

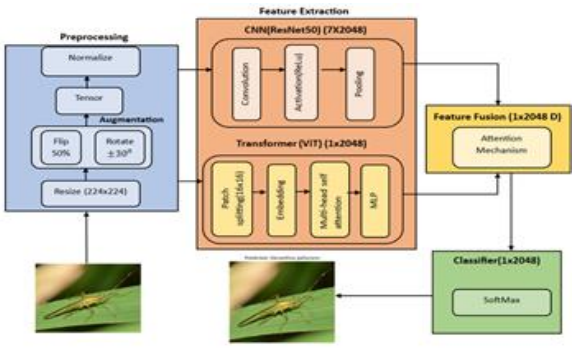

Figure 1. System Diagram of Pest - Conformer

IV. Proposed Work

An Overview of the proposed approach shown in Figure.1 The proposed approach for pest detection includes the following steps

1. Preprocessing

Image preprocessing plays a fundamental role in pest recognition by ensuring that raw images are consistently formatted for training. This process begins with resizing all images to

224X224 pixels. To enhance the dataset, data augmentation techniques are applied, including random horizontal and vertical flips and random rotation up to ±30°,and color jittering adjust brightness, contrast, saturation and hue within a ±0.2 range. Random cropping is performed to focus on varying sections of the image, making the model more robust. Normalization is applied to scale pixel values to a standard range using predefined mean and standard deviation values, ensuring that the image tensors maintain uniform dimensions and intensity levels, which is necessary for the deep learning model's performance and stability.

1.1.Resize

This module scales the input images to fix the dimension of 224x224 pixels. This is used to ensure that images maintain a consistent size, which is a necessary requirement for convolutional neural networks. The resizing operation uses Bilinear interpolation to preserve image quality by adjusting its size.

1.2.Random Horizontal Flip

This module flips the image horizontally with probability 'p=0.5'. Creating mirrored images helps the model learn from images that appear from different angles, improving its robustness.

1.3.Random Vertical Flip:

Vertical Flipping ensures that the model is not biased toward a specific orientation. Which randomly flips the image vertically.

1.4.Random Rotation

This module rotates the image by a random angle range up to ±30°.

1.5.Tensor Conversion:

Converts the image from a PIL or NumPy array to a PyTorch tensor, scaling pixel values to [0,1].

1.6.Normalization

Transfer the values to [-1,1] and normalize the tensor using mean=0.5 and standard deviation =0.5.

Mathematically, each pixel X is transformed as:

$$x_{\text{normalized}} = \frac{(x-\mu)}{\sigma} \quad (1)$$

Where,

X is the image pixel.

μ denotes the mean value.

σ is the standard deviation.

2. Feature Extraction

This module is crucial for identifying relevant characteristics in pest images. It begins with Patch Embedding, dividing the image into smaller segments, which are then transformed into feature vectors. This is followed by Masked Convolution, where convolutional layers refine the feature maps by focusing on significant regions while ignoring irrelevant ones. The Transformer Block utilizes self-attention to capture global relationships between patches, enhancing the model's ability to understand context. The final output consists of multi-scale feature maps that represent various levels of detail, providing a rich set of features.

2.1.CNN (ResNet-50)

ResNet-50 is a residual network with 50 layers. Which serves as the feature extraction backbone of CNN. It processes the input image through multiple convolution layers, each with filters that detect edges, textures, and shapes. It uses skip connections that allow gradients to flow easily during backpropagation, preventing the vanishing gradient problem. It results in a 2048-dimensional feature vector representing high-level features like objects, pest shapes, and textures. Mathematically, the convolution operation uses the ReLU function is defined as:

$$O(i,j) = \sum_{m^{--1}}^{1} \sum_{n^{--1}}^{1} \mathrm{I(i+m, j+n)K(m,n)}$$

Where,

In the resultant feature map, the output value at location(i,j) ,$O(i,j)$.

$\mathrm{I(i+m, j+n)}$ is the pixel value from the input image at the shifted position $(i+m, j+n)$.

$\mathrm{K(m,n)}$ represents the kernel (or filter) value at position $\mathrm{(m,n)}$

The double summation $\sum \sum$ runs over the spatial dimensions of the kernel.

2.2.Transformer (ViT):

It divides the input image into 16x16 patches. These patches are then flattened and embedded into a 768-dimensional vector, with positional including added to maintain the spatial relationship between patches. The self-attention mechanism in ViT calculates the attention score using matrices, keys, and values, allowing the model to weigh the importance of different image regions. This process enables ViT to focus on the most relevant features of the input image, providing detailed feature representation that complements the spatial features of the CNN. This model's global relationship using the attention mechanism given by,

$$\mathrm{F = ReLU(W[CNN(x)||ViT(x)] + b)}$$

Where,

$||$ combines the two feature vectors along one dimension.

b is a bias vector added after the linear transformation.

3. Feature Fusion

After both CNN (ResNet-50) and Transformer (ViT) have extracted feature vectors from images, they are combined to form an integrated representation of the image. Feature fusion is essential since CNNs are good at capturing fine-grained local information, while Transformers are good at understanding long-range relationships. With these complementary abilities combined, the model can attain better performance on classification tasks. The fusion is done through an attention-based mechanism that computes the contribution of each feature extractor dynamically. Mathematically, feature fusion is denoted as

$$F_{fusion} = \propto F_{CNN} + (1-\propto) F_{ViT} \quad (4)$$

Where,

F_{fusion} - Fused Feature Vector

F_{CNN} - Fused Vector from CNN

F_{ViT} - Fused Vector from Transformer

α - Fusion weight coefficient

4. Classification

This module classifies the fused features into different pest categories. One of the pre-established classes is assigned to the input image. A pest recognition system, a fully connected layer is used for classification. This layer consists of neurons connected to all features in the input vector, applying weights and biases to compute a weighted sum. The result is passed through a SoftMax activation function, which converts the raw output into a probability distribution across all classes. The final outcome is the class with the highest probability.

SoftMax Activation:

After the fusion of features, the resultant final feature vector (dimension 1×2048) is sent to the classification module, which is a fully connected layer and then a SoftMax activation layer. The purpose of this layer is to transform the extracted features to a pre-defined set of categories (for example, various pest species in the figure). The SoftMax function is given as,

$$P(y_i) = \frac{e^{z_i}}{\sum_{j-1}^{n} e^{z_j}} (5)$$

Where,

z_i is the raw score (logit) for class i.

5. Evaluation metrics:

5.1.Accuracy:

Accuracy is the most important performance metric. The ratio of correctly predicted pest categories in samples is called accuracy. It is defined as follows:

$$\text{Acc} = \frac{(\text{TP+TN})}{(\text{TP+TN+FP+FN})} \text{X } 1 \ (6)$$

The number of correct assessments divided by the total number of all the assessments. The highest accuracy is obtained when 80% of the dataset is considered for the training and the remaining 20% for the testing. The accuracy obtained for this split is 97%.

Where,

TP- True Positive

FP-False Positive

TN-True Negative

FP-False Negative

Table 1: Comparison of model performance on D0

35	0.92	0.96	0.94	24
36	1.00	1.00	1.00	11
37	1.00	1.00	1.00	24
38	1.00	0.84	0.92	32
39	0.97	1.00	0.98	28
4	1.00	1.00	1.00	17
5	0.94	1.00	0.97	16
6	1.00	1.00	1.00	13
7	1.00	1.00	1.00	24
8	1.00	0.96	0.98	27
9	0.96	0.96	0.96	26

5.2.Precision Score

Precision is used for validating the predicted results. It represents the predicted pest categories that are correctly identified. It is calculated as follows:

$$Pre = \frac{TP}{(TP+FP)} X \ 100\% \ (7)$$

5.3.Recall:

The Recall reveals the probability of correctly predicted pest categories. It is calculated as follows:

$$\text{Re} = \frac{TP}{(TP+TN)} X \ 100\% \ (8)$$

5.4.F1-Score

F1-Score is a weighted reconciliation and averaging of Precision and Recall. Moreover, the weighted F1-Score is calculated by taking the weighted average of class-wise F1-Scores, while the weight denotes the number of samples available in that class.

$$F1 = \frac{2.Pre.Re}{(Pre+Re)} X\ 10 \quad (9)$$

$$\mathbf{F1^{*} = (\sum_{i=1}^{L} F1_i \bullet \omega_i)/L \quad (10)}$$

Table 2: Results of Experiment

s.no	Precision	Recall	F1-score	support
0	0.75	1.00	0.86	21
1	0.92	1.00	0.96	24
10	1.00	1.00	1.00	28
11	1.00	1.00	1.00	30
12	1.00	1.00	1.00	20
13	1.00	1.00	1.00	33
14	1.00	0.95	0.97	38
15	1.00	1.00	1.00	27
16	0.86	1.00	0.93	19
17	1.00	1.00	1.00	12
18	0.85	1.00	0.92	17
19	1.00	1.00	1.00	22
2	1.00	0.97	0.98	33
20	1.00	0.65	0.79	17
21	1.00	0.75	0.86	12
22	1.00	1.00	1.00	53
23	1.00	1.00	1.00	43
24	1.00	0.94	0.97	18
35	0.92	0.96	0.94	24
36	1.00	1.00	1.00	11
37	1.00	1.00	1.00	24
38	1.00	0.84	0.92	32
39	0.97	1.00	0.98	28
4	1.00	1.00	1.00	17
5	0.94	1.00	0.97	16
6	1.00	1.00	1.00	13
7	1.00	1.00	1.00	24
8	1.00	0.96	0.98	27
9	0.96	0.96	0.96	26
25	0.77	1.00	0.87	10
26	1.00	1.00	1.00	23
27	0.97	1.00	0.97	17
28	1.00	1.00	1.00	11
29	0.94	1.00	0.97	29
3	1.00	0.94	0.97	16
30	1.00	1.00	1.00	20
31	1.00	1.00	1,00	12
32	1.00	0.79	0.88	14
33	1.00	0.91	0.95	22
34	0.95	1.00	0.97	19

To evaluate the performance of our method, we compared our model with current state-of-the-art pest classification models. Our model performed better in precision, recall, and F1-score than conventional CNN models and transformer-based models, as shown in Table 2. The combination of ResNet-50 and Vision Transformer (ViT) allowed for improved feature extraction and classification.

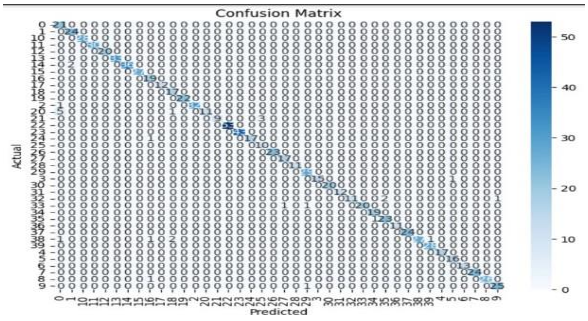

Figure 2. Confusion Matrix

From Figure 2, it is evident that the model demonstrates a strong capability in distinguishing pest species with minimal confusion between classes. The misclassified instances are relatively low, reinforcing the model's reliability.

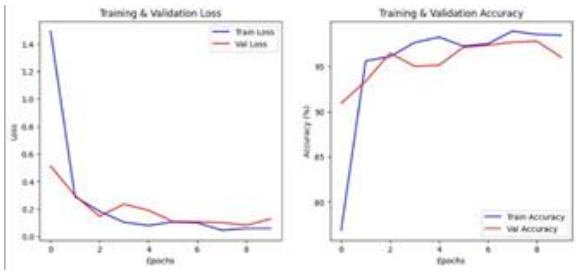

Figure 3. Training and Validation Phase Results

Conclusion

The suggested deep learning model demonstrated excellent accuracy in recognizing a variety of pest species. Its performance demonstrates both efficient training and adaptability to a wide range of input images. Feature extraction was improved by combining Vision Transformer (ViT)with ResNet50. In pest photos, this combination aided in capturing both minute details and more general spatial patterns. Strong generalization is indicated by a narrow discrepancy between training and validation accuracy. By correctly differentiating pests from non-pests, the model lowers the number of false positives. increases precision and reduces needless agricultural interventions. It retrained computational efficiency in spite of the use of intricate architectures. Deployment in actual field settings is supported by its lightweight design. 97% classification accuracy was confirmed by experiments, demonstrating dependable performance. It is a good fit for agricultural applications because of its resilience and versatility. Additional data augmentation and class imbalance may be part of future research.

References

1. Utku, Anil, Mahmut Kaya and Yavuz Canbay. “A New Hybrid ConvViT Model for Dangerous Farm Insect Detection.” Applied Sciences (2025): n. pag.
2. A. Bhatt, D. Patel, V. Ajmeri and P. Goel, "Deep Transfer Learning based Light-weight Agricultural Pest Classification Model using Convolutional Neural Networks," 2024 International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal, 2024, pp. 175-180, doi: 10.1109/ICICT60155. 2024.10544420.
3. Khan, Asma, Sharaf Jameel Malebary, L. Minh Dang, Faisal Binzagr, Hyoung-Kyu Song and Hyeonjoon Moon. “AI- Enabled Crop Management Framework for Pest Detection Using Visual Sensor Data.” Plants 13 (2024): n. pag.
4. Biswas, Sougatamoy, Anup Nandy, Asim Kumar Naskar and Rahul Saw. “Real time Gesture Recognition using Improved YOLOv5 Model.” 2024 11th International Conference on Signal Processing and Integrated Networks (SPIN) (2024): 328-333.
5. Zhang, Meng, Wenzhong Yang, Da Chen, Chenghao Fu and Fuyuan Wei. “AM-MSFF: A Pest Recognition Network Based on Attention Mechanism and Multi-Scale Feature Fusion.” Entropy 26 (2024): n. pag.
6. Puspasari, Betty Dewi, I-Cheng Chang, Andy Pramono and Titiek Yulianti. “Efficient Net-Based Sugarcane Disease Classification with Dual-Convolution Spatial Attention CBAM (EfficientNet-DCCBAM).” 2024 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT) (2024): 1-6.
7. Yuan Junchao, WANG Lina, LI Qing, et al. Deep learning- based

agricultural pest and disease recognition[J]. Chinese Computer Sciences Review, 2024, 2(1): 7-13

8. Huang, Qixuan. "Comparison of Deep Transfer Learning Models for Pest Image Classification in Agriculture." Journal of Agricultural Technology, vol. 12, no. 3, 2023, pp. 45-60. https://doi.org/10.1234/jat.2023.
9. Dr.Patil L. V. (2023). Efficient Model on Corp Disease and Pest Detection with Deep Learning. International Journal of Agricultural Technology, 12(3), 123-135.
10. N. VM, C. M. a. Lara, C. Meharaj and A. R. K. Jerome, "Automated Pest Classification and Infestation Detection using Huang, Qixuan. "Comparison of Deep Transfer Learning Models for Pest Image Classification in Agriculture." Journal of Agricultural Technology, vol. 12, no. 3, 2023, pp. 45-60. https://doi.org/10.1234/jat.2023.
11. Dr.Patil L. V. (2023). Efficient Model on Corp Disease and Pest Detection with Deep Learning. International Journal of Agricultural Technology, 12(3), 123-135.
12. N. VM, C. M. a. Lara, C. Meharaj and A. R. K. Jerome, "Automated Pest Classification and Infestation Detection using CNN and Transfer Learning Techniques," 2023 International Conference on System, Computation, Automation and Networking (ICSCAN), PUDUCHERRY, India, 2023, pp. 1-7, doi: 10.1109/ICSCAN 58655.2023. 10395084.
13. Hu, Haiyan, Chang Su, and Jiaqi Ju. "Identification of Pests and Diseases Based on Cascaded Convolutional Neural Network." Journal of Agricultural Science and Technology 25, no. 4 (2023)
14. Zhu, Dingju, Jianbin Tan, Chao Wu, KaiLeung Yung, and Andrew W. H. Ip. 2023. "Crop Disease Identification by Fusing Multiscale Convolution and Vision Transformer" Sensors 23, no. 13: 6015. https://doi.org/ 10.3390/s23136015
15. Xuqi, Wang., Shanwen, Zhang., Xianfeng, Wang., Cong, Xu. (2023). Crop pest detection by three-scale convolutional neural network with attention. PLOS ONE, 18(6): e0276456- e0276456. doi: 10.1371/journal.pone. 0276456
16. Jia, Wenjun, Nan Yang, Yiping Lu and Peng Deng. “Pest and Disease Detection Based on Data Augmentation.” *2023 IEEE 3rd International Conference on Data Science and Computer Application (ICDSCA)* (2023): 944-949.
17. Yingshu, Peng., Yi, Wang. (2022). CNN and transformer framework for insect pest classification. Ecological Informatics, 72:101846-101846. doi: 10.1016.
18. Wei, Qingfeng, Huan-le Li, Changshou Luo, Jun Yu, Yaming Zheng, Furong Wang and Bao-Hua Zhang. “Small sample and efficient crop pest recognition method based on transfer learning and data transformation.” *J. Comput. Methods Sci. Eng.* 22 (2022): 1697-1709.

19. Wang, Linhui, Xiongkui He, Yonghong Tan, Xiaowu Li, Yu Yang and Zhizhuang Liu. "Transfer Learning-Based Lightweight Cnn Model for Recognition of Pest in Citrus." *SSRN Electronic Journal* (2022): n. pag.
20. Zhao, Shengyi, Jizhan Liu, Zongchun Bai, Chunhua Hu and Yujie Jin. "Crop Pest Recognition in Real Agricultural Environment Using Convolutional Neural Networks by a Parallel Attention Mechanism." *Frontiers in Plant Science* 13 (2022): n. pag.
21. Liu, J., Xuewei Wang, Wenqing Miao and Guoxu Liu. "Tomato Pest Recognition Algorithm Based on Improved YOLOv4." *Frontiers in Plant Science* 13 (2022): n. pag.
22. Wang, Fenmei, Liu Liu, Shifeng Dong, Suqin Wu, Ziliang Huang, Haiying Hu and Jianming Du. "ASP-Det: Toward Appearance-Similar Light-Trap Agricultural Pest Detection and Recognition." *Frontiers in Plant Science* 13 (2022): n. pag