Journal of Intelligent Systems and Internet of Things
Vol. 17, No. 01, PP. 27-38, 2025
DOI: https://doi.org/10.54216/JISIoT.170103

TRIP-CID: Transformer and ResNet Improved Pest Classification and Identification Detection Model for Pesticide Management in Precision Agriculture

R. Kiruthika[1,*], B. Arun kumar[2]

[1]Research scholar, Department of Computer Science and Engineering, Faculty of Engineering, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India

[2]Professor and Head, Department of Artificial Intelligence and Data Science, Faculty of Engineering, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India

Emails: kkmailkiruthi07@gmail.com; arunkumar.oct06@gmail.com

Abstract

In these modern agriculture system crop pests causes major social, economic and environmental issues worldwide. Each pest necessitates an alternative method of control and precise detection has become a very important challenge in agriculture. Deep learning technique shows remarkable results in image identification. Standard pest detection framework might struggle with accuracy due to complicated algorithms and lack of data, and result in incorrect detection, which leads to harm the crop environment. To end this, we developed a novel framework named Transformer and ResNet Improved Pest Classification and Identification Detection (TRIP-CID) for crop pest classification and identification. At first, the pest images are obtained through the benchmark dataset for pre-processing. The Pre-processed images are immediately delivered to the Improved ResNet (IR-Net) and Pyramidal Vision Transformer (PVT) for multi-scale spatial, channel and contextual feature maps extraction within three stages. The extraction feature maps in the two modules are combined to produce a superior feature map. Then refined feature maps was fed to the three distinct Machine Learning (ML) classifiers offered pest detection outcomes. For accurate results, we employ ensemble-voting technique, which outputs effective pest detection result that is vastly used for particle suggestion. Finally, we utilized presented technique for detecting and identify crop pest in 10-pest class for instance larva of laspeyresia pomonella, Euproctis pseudoconspersa strand, Locusta migratoria, acrida cinerea, empoasca flavescens, spodoptera exigue, parasa lepida, chrysochus chinensi, L.pomonella types of insects pests and larva of S. exigua. Additionally, the suggested methodology has shown to provide experts and farmers with quick, efficient assistance in identifying pests, saving money and preventing losses in agricultural output.

Received: November 14, 2024 Revised: January 02, 2025 Accepted: March 05, 2025

Keywords: Deep learning; Pest detection framework; TRIP-CID; Ensemble voting technique; Feature extraction

1. Introduction

Precision agriculture is a present-day farming process, which optimizes agricultural processes over the usage of technology and data analytics. When increasing crop yields it pursuit for to diminish the amount of input resources including water, fertilizer, and pesticides. This tactic differs from conventional farming methods that normally hang on applying resources consistently concluded perfect fields, irrespective of variations in crop health, soil fertility, or moisture conditions [1-2]. The volume of precision agriculture to simplify site-specific management techniques is one of the main recompenses. Farmers can aim specific areas and times while they are most needed instead of applying inputs evenly through a field. For occasion, farmers will adapt planting, irrigation, and fertilizer application rates according to regional conditions by implementing automated systems and GPS-guided equipment [3-4].

Additionally, the increasing resource efficiency and this lessens the impact on the environment by decreasing overflow and chemical leakage into waterways. Moreover, precision agriculture facilitates data-driven decision-making throughout the farming cycle. Predictive analytics and machine learning algorithms can anticipate yields by employing historical data analysis and detect early disease symptoms or nutritional deficits to regulate the best times to harvest [5-7]. Farmers may reduce risks, increase profits, and preserve workable farming techniques by being proactive. When lowering hazards to the environment and public health pesticide management is a revolutionary strategy meant to maximize pesticide application for effective crop protection inside the situation of precision agriculture. Precision agriculture utilizes cutting-edge technologies such as remote sensing, GPS-guided equipment and data analytics to qualify site-specific pesticide application, in divergence to traditional methods, which be contingent on uniform application through entire fields [8-9]. Farmers will diminish pesticide consumption and lessen environmental consequence by accurately targeting regions with insect infestations or crop stress by employing real-time data from sensors, drones, and satellite photography [10-12]. Moreover, precision farming provisions sustainability over decreasing pesticide leaching and runoff, maintaining healthy soil and securing water quality. Upcoming expansions in biological control integration, predictive modelling, and smart monitoring systems offer promise for cumulative pesticide management procedures in precision agriculture and assuring robust and maintainable food production systems worldwide as technology continues to development [13-14].

Recent progresses in deep learning have transformed computer vision, especially with the Transformer and Residual Network (ResNet) architectures that will allows for prominent gains in responsibilities such as object detection and image classification. These developments are currently being employed to innovative and important fields including identification and classification of pests. All-inclusive, pests characterize serious risks to ecosystem health, food security, and agricultural output. Conventional tactics to pest identification regularly hang on labour-intensive, time-consuming, and human error-prone manual examination or simple image processing procedures [15-16]. Moreover, these procedures have trouble to appropriately identifying pests in a variety of environmental settings and with complex antiquities. Deep learning models offer intriguing answers to these problems including Transformers and ResNet. Transformers have established to be operative at capturing contextual information and long-range dependencies [17-18]. Transformers are outstanding in capturing spatial relationships among various picture components while applied to image data. Numerous goals are projected to be proficient by participating Transformer and ResNet models into pest detection frameworks. These models will improve accuracy rates in distinguishing among pest species and developmental phases. Furthermore, they offer improved scalability, empowering the employment of automated monitoring systems through extensive agricultural regions or natural environments [19-20].

Precision agriculture can assistance from enhanced crop pest identification and detection over the use of deep learning. These arrangements have the potential to greatly improve pest management techniques that will raise agricultural yields and promote more sustainable farming practices. They organize this by implementing large volumes of image data and cultured neural network architectures. To overcome from the mentioned issues, we recommend a new approach called Transformer and ResNet Improved Pest Classification and Identification Detection (TRIP-CID). We flinch our method by locating pest images from benchmark datasets that are cautiously pre-processed. Following that, the pre-processed images are concurrently directed into two different modules such as the Pyramidal Vision Transformer (PVT) and the Improved ResNet (IR-Net). These modules extract multi-scale spatial, channel, and contextual feature maps over the three processing stages. Formerly, the feature maps, which were occupied from PVT and IR-Net, are merged to generate improved feature maps that encompass comprehensive information about pests. The improved feature maps are formerly placed into three different machine learning (ML) classifiers to harvest reliable pest identification. Every classifier independently developments the feature maps to yield preliminary pest detection results. Then, we employ an ensemble voting approach to assurance accuracy and reliability, which masses the ML classifier outputs. This ensemble approach enhances the ability of model to appropriately detect pests diagonally a range of environmental circumstances by uniting different viewpoints from the classifiers. Our TRIP-CID model's final output is critical for suggesting targeted pesticide application techniques to supporting with accurate pest detection. Our model pursues to simplify and optimize pest control decisions in precision agriculture settings by fusing cutting-edge deep learning procedures with ensemble learning procedures. At last, this investigates advances the supportable farming procedures by dropping pesticide use and growing crop protection and productivity. The contribution of our work is detailed explained below:

Integration of Improved ResNet and Pyramidal Vision Transformer for Enhanced Feature Extraction: The TRIP-CID model shapes on the strengths of enhanced ResNet (IR-Net) and Pyramidal Vision Transformer (PVT) to harvest stronger multi-scale feature extraction abilities. IR-Net enhances spatial and channel feature extraction in classic ResNet, while PVT effectively gathers multi-scale contextual data by implementing transformer-based designs. The TRIP-CID approach signifies pest images more holistically and precisely by the integration of feature maps from IR-Net and PVT. This fusion enhances accuracy for tasks like detection and classification by manipulating the wide-ranging spatial, channel and contextual information recovered from the dual models. Ensemble of Machine Learning Classifiers for Robust Pest Detection: The TRIP-CID method utilizes three separate Machine Learning (ML) classifiers using modified feature maps to achieve pest identification results. The model captures numerous features of the data that individual classifiers may overlook by applying several classifiers and resulting in a more thorough analysis. The ensemble voting technique is then utilized to merge the outputs of these classifiers, resulting in finished and exact pest detection findings. This ensemble technique considerably improves the robustness and reliability of the results while reducing the possibility of false positives and negatives. Application of Ensemble Voting Method for Pesticide Recommendation: The final addition of the TRIP-CID approach is the use of the ensemble voting approach to acquire reliable pest detection findings. This approach merges the results of numerous classifiers to engender a final judgment and employing the strengths of every classifier. This method is important for practical applications like pesticide recommendations since it guarantees which pest detection findings are precise and dependable. The ensemble voting mechanism promises which the findings are not only accurate but also actionable and subsequent in enhanced pest control tactics and pesticide use.

2. Related Works

Ullah N et.al [21] suggested a radical end-to-end DeepPestNet architecture for the classification and recognition of pests. Eleven learnable layer's structure the presented approach which consisting of three fully connected and eight convolutional layers. He also implemented image rotation and augmentation methods to enlarge the related dataset and measure the DeepPestNet generalizability of the method. Then, he assessed the presented approach structure by employing the well-known crop dataset of Deng. At last, crops pests were detected and classified into 10 classes by utilizing the presented approach. Wang B et.al [22] suggested a crop disease and pest detection approach related on DL from the viewpoint of biological and ecological guard to report the issues of abundant types of crop diseases and pests, dissolute dispersal speed and lengthy physical detection times. Initially, field sampling is employed to gather crops images to generate a dataset. Then, nearest neighbour interpolation is implemented to appearance the image pre-processing. Moreover, the Alex Net network structure is improved. Experimental parameters and different neuron nodes are set by optimizing the entire connection layer. Finally, agricultural disease and pests are detected by employing the Alex net approach. Li W et.al [23] utilized three innovative deep convolutional Neural Network (DCNN) approaches such as Yolov5, mask-RCNN and faster RCNN. Moreover, he formed two of his own coco datasets by implementing the IP102 and Baidu AI insect detection datasets. In addition, he employed the two coco datasets to equivalence these three innovative deep learning approaches. Yolov5 is highly endorsed for the insect pest detection along with the provisional results related on the simple background of the Baidu AI insect detection dataset. Lin S et.al [24] presented a fine-grained pest identification method based on a GPA-Net to maximize the production of agricultural efficacy. Additionally, a multilevel pyramid structure was intended to acquire multiscale spatial features and graphic relations for improving capacity to identify pests and diseases.

Zhang W et.al [25] suggested the lightweight AgriPest-YOLO pest detection approach that targets to accomplish well-balanced model size, accuracy and efficacy for pest identification. Initially, he offered a CLA approach to minimize noise interference and harvest smoother and richer insect features and especially for complexly backgrounded pests. Then, a unique GSPPF is developed that advances the multi scale representation of pest features. Soft NMS is lastly augmented to the forecast layer to increase the concluding predictions of overlying pests. He also assessed his efficiency of approach with retaining a large-scale multi pest image dataset on 25K images and 24 classifications. Khalid S et.al [26] suggested an approach for early pest detection, which creates use of CNN and DL. A dataset comprising images on citrus psylla, red beetles, and thistle caterpillars is exposed to object detection. The input dataset contains 9876 images of each bug in different lighting situations. They were all selected based on these models accomplished in object detection. An android application, which identifies the pests in real time, also integrates the Yolov8 perception. In his learning, he employs new deep learning approaches, analytical techniques, and a workflow for pest detection to efficiently achieve pests in crops. Rodríguez-García M. A et.al [27] presented a preliminary version of a fully functional, semantically improved IPM decision assistance system. The eventual unbiased is to generate a comprehensive agricultural knowledge base by accumulating data from numerous and heterogeneous sources to generate a system which will support farmers to make decisions around handling diseases and pests. Subsequently evaluation in a simulated setting, the pest classifier background yielded a cumulative high accuracy.

Chen J et.al [28] presented the attention mechanism together with a CAM were combined in his planning for acquiring the important pest data of input images to enhance knowledge ability for pest images with muddled circumstances and also the MobileNet-V2 pre-trained on ImageNet is selected. Moreover, two-stage transfer learning and an optimized loss function were employed training process. This category of advanced learning maximizes the detection accuracy of plant pest images by primary supporting model acquire around large-scale features and then progressively turning its emphasis to finer facts. Zheng Z et.al [29] presented electronic noses related on metal oxide semiconductor sensors to identify insect pests and crop diseases. Electronic nose (E-nose) technology has advanced rapidly, imitating the olfactory senses of animals and provided that early disease and pest warnings. Among its many advantages are its non-damage detection, cheap cost, high sensitivity, real-time analysis, ease of use, and portability. When pests give the impression, crops will release either Volatile Organic Compounds (VOCs) to avoid the bugs or publication VOCs to draw in the pests' natural enemies as a defence mechanism. The development of MOS electronic nose technology and its use in the detection of agricultural illnesses and insect pests is predictable to yield important information for crop disease and insect pest prevention. MacDougall S et.al [30] focused on different gas-sensing technologies have been practical over the last ten years to better effectively identify plant illnesses and pests. Among these is the utilization of various gas-sensing technologies to take advantage of volatile organic chemicals emitted by stressed plants. These techniques regularly satisfy these criteria, but they have a number of drawbacks as well, such as the enormous number of variables that can influence the profile of volatile organic compounds released, such as sensor drift and sensitivity to environmental factors and soil nutrient or water availability.

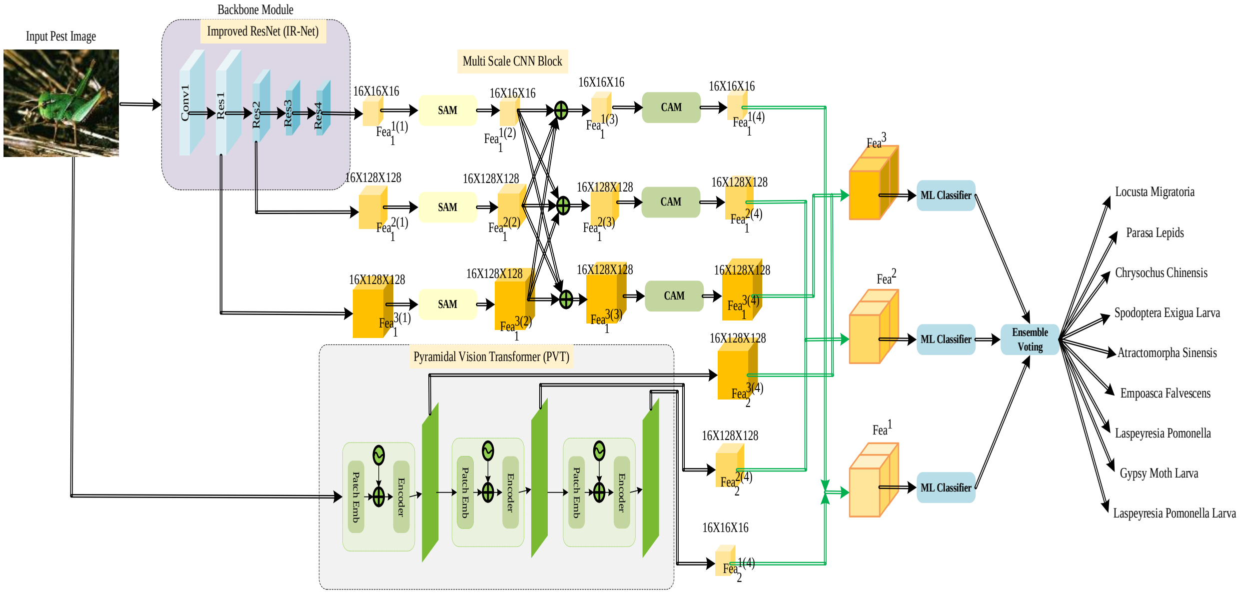

Figure 1. Illustration of TRIP-CID Model

3. Methodology

3.1 TRIP-CID Model

The Deep leaning has become most popular method for image processing and computer science. In this research, we proposed a Transformer and ResNet Improved Pest Classification and Identification Detection (TRIP-CID) method for accurate crop pest's detection and categorization. By utilizing the (Deng.et.al) dataset, we created a 10-class crop pest's classification. The proposed TRIP-CID contains of four fundamental phases, which include data augmentation and image rotations, dataset resizing, image resizing, model training and testing, as shown in figure 1. This model is further evaluated to classify our method performance; we developed a nine-class classification by implementing the proposed TRIP-CID framework utilizing the traditional Kaggle "pest dataset" to demonstrate the model effectiveness. In addition, this proposed framework consists of 11 learnt layers including 8 convolutional layers and 3 FC layers. The particulars of presented model were detailed below.

A. Pre-Processing

(i) Data augmentations

Lack on huge quantity of information for training the deep learning technique is one of the major issues while attempting to utilize DL methods to categorization difficulties and pest identification. Most of the crop pest data is hard and costly to obtain the materials and time. The data augmentation extents the amount of obtainable data with no need of obtaining new data through employing numerous methods to present data, has been proven beneficial in classifying images. Due to the low amount of images dataset, we employ the data augmentation and image rotations method in this research. Furthermore, we also rotated all the image datasets in 90 degrees multiple times. Researcher haphazardly adjust photos within 0.9 and 1.1 to create another image. This technique vigorously creates augmented images between every phase training. The number of images within the training set were drastically increased by utilizing data augmentation technique. By increasing, the amount of training images enhances the efficiency of our deep learning model. Moreover, only the real images in the dataset were utilized for testing which ensure the augmented images was solely for training purposes.

(ii) Image resizing:

In the dataset, the input images are in various sizes, which guarantee consistency and increase the speed of processing. Then, we implemented a specific pre-processing to resize the image input to 226 × 226 pixels related on image input necessities of our method.

(iii) Dataset partitioning:

For every experimentation, the dataset was divided into training and testing sets. More accurately, almost 95% of the dataset was utilized to train the model and 5% was utilized to test the model.

3.2 Pyramidal Vision Transformer (PVT)

We delivered the PR^{im} input to the Pyramid Vision Transformer (PVT). The PVT contains distinctive aspects of transformer and CNN, which could be utilized as direct substitution for the backbones of CNN. The illustration of the PVT architecture is showed in figure 2. In order to achieve higher output resolution, the method can only train on densely partitions images nonetheless; it uses the progressive shrinking pyramid to decrease the calculations of huge feature maps. Moreover, it also implements a Dimensionality-reduction attention (DRA) for additional resource consumption reduction while learning high-resolution features. The detailed equations are provided below:

$$\mathrm{SRA}(R, L, W) = \mathrm{Con}\left(\mathrm{head}_0, \ldots \mathrm{head}_{M_j}\right) X^p \qquad (1)$$

$$\mathrm{head}_m = \mathrm{Att}\ (RX_m^R, \mathrm{SRA}(L)X_m^L, \mathrm{SRA}(W)\ X_m^W) \qquad (2)$$

$$\mathrm{SRA}_{(y)} = N(\mathrm{Res}\ (y, \mathrm{Re}_j)X^t) \qquad (3)$$

$$\mathrm{Att}(R, L, W) = \mathrm{Softmax}\left(\frac{R^L}{\sqrt{e_{head}}}\right) W \qquad (4)$$

Thus Con (.) represents the concatenation operation in $X_i^R \in \mathrm{Re}^{D_j * e_{head}}, X_i^L \in \mathrm{Re}^{D_j * e_{head}}, X_i^W \in \mathrm{Re}^{D_j * e_{head}}$, and $X^p \in \mathrm{Re}^{D_j * D_j}$ denotes the parameters for linear projection. The SRA(.) can reduce the geospatial complexity for the sequence input for instance L or W. The Reshape operation (y, Re_j) includes reshaping the sequence input x in the size sequence $\frac{I_j X_j}{\mathrm{Re}_j^2} * \left(\mathrm{Re}_j^2 D_j\right)$. Thus Re_j represents the layer attention phase j decrease ratio. $X_t \in \mathrm{Re}^{\left(\mathrm{Re}_j^2 D_j\right) * D_j}$ represents the liner projection which decreases the sequence's dimension to D_j. N (.) represents the normalization of layer.

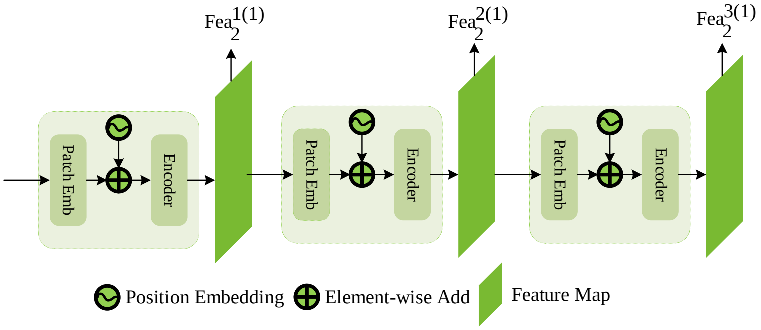

Figure 2. Illustration of Pyramidal Vision Transformer (PVT)

C. Improved ResNet

The pre-processed image Pr^{Im} is provided as an input to the backbone module (IR-Net) for multi scale feature extraction. To be clearer, the $Pr_{(1)}^{Im} \in R^{3 \times He \times We}$ is fed as input to the IR-Net which produce three feature maps $\mathrm{Fea}_j^{(1)}, \mathrm{Fea}_j^{(2)}$, and $\mathrm{Fea}_j^{(3)}$ with varied scales from the backbone. The designed backbone composed of five blocks, which also composed of 5×5 convolutional layers, and four residual blocks. The blocks are denoted as $\mathrm{Con1}, \mathrm{Res2}, \mathrm{Res3}, \mathrm{Res4}$, and $\mathrm{Res5}$ respectively. With the stride 2, the $\mathrm{Res3}$ and $\mathrm{Res4}$ perform down sampling operation. Overall, the three feature maps $\mathrm{Fea}_j^{(1)}, \mathrm{Fea}_j^{(2)}$, and $\mathrm{Fea}_j^{(3)}$ are extracted from $\mathrm{Res2}, \mathrm{Res3}$, and $\mathrm{Res5}$ respectively. The extracted features are provided to the Spatial Attention Module (SAM). The SAM intelligently process the important information in the feature map. To be clearer, the element wise multiplication is performed on every channel of $\mathrm{Fea}_j^{i(1)}$ using 2D spatial attention $N_s(\mathrm{Fea}_j^{i(1)}) \in R^{He \times We}$. Features that are more important are provided with higher weights based on its position. By this way, the SAM robustly highlights the crucial pest information by diminishing the irrelevant regions in the pest image. In order to attain spatial attention $N_s(\mathrm{Fea}_j^{i(1)})$, we introduce maxpooling and average pooling operations in the axes of channel and amalgamate the pooling results to generate $N_s(\mathrm{Fea}_j^{i(1)})$. The formulation of spatial attention with respect to $\mathrm{Fea}_j^{i(1)}$ can be formulated as,

$$\mathrm{N_s}\left(\mathrm{Fea}_{\mathrm{j}}^{\mathrm{i(1)}}\right) = \sigma(\mathrm{f}(\left[\mathrm{APool}\left(\mathrm{Fea}_{\mathrm{j}}^{\mathrm{i(1)}}\right); \mathrm{Mpool}\left(\mathrm{Fea}_{\mathrm{j}}^{\mathrm{i(1)}}\right)\right]) \qquad (5)$$

From the above equation, sigmoid function is denoted as σ, APool and Mpool denotes the average and maxpooling operations respectively, and $\mathrm{f}(.)$ denotes the convolution operation. So that, the obtained feature map from SAM can be denoted as follows,

$$\mathrm{Fea}_{\mathrm{j}}^{\mathrm{i(2)}} = \mathrm{Fea}_{\mathrm{j}}^{\mathrm{i(1)}} \otimes \mathrm{N_s}(\mathrm{Fea}_{\mathrm{j}}^{\mathrm{i(1)}}) \qquad (6)$$

From the above equation, element wise multiplication can be denoted as $\otimes$. For every feature map channel, this research lime lights the weight matric of the $\mathrm{N_s}$ in the dimension of channel during element wise multiplication.

The extracted features maps are provided for feature exchange to resolve the temporal and domain gap issues. Let us consider the feature maps $\mathrm{Fea}_1^{\mathrm{i(1)}}$ and $\mathrm{Fea}_2^{\mathrm{i(1)}}$ in the spatial dimension can be formulated as,

$$\mathrm{Fea}_1^{\mathrm{i(1)}}(\mathrm{ba, ch, he, wi}) = \begin{cases} \mathrm{Fea}_1^{\mathrm{i(1)}}(\mathrm{ba, ch, he, wi}), \mathrm{H}(\mathrm{ba, ch, he, wi}) = 1 \\ \mathrm{Fea}_{2^{2-\mathrm{j}}}^{\mathrm{i(1)}}(\mathrm{ba, ch, he, wi}), \mathrm{H}(\mathrm{ba, ch, he, wi}) = 0 \end{cases} \qquad (7)$$

From the above equation, channel, width, height, and batch of the dimensions are denoted by ch, wi, he, and ba respectively. The feature exchange decision operation to perform feature exchange at the specific position can be denoted as H. More distinctively, the feature exchange is performed among $\mathrm{Fea}_1^{1(1)}$ & $\mathrm{Fea}_2^{1(1)}$, $\mathrm{Fea}_1^{2(1)}$ & $\mathrm{Fea}_2^{2(1)}$, and $\mathrm{Fea}_1^{3(1)}$ & $\mathrm{Fea}_2^{3(1)}$. We perform both the channel and spatial feature exchange for the large- and small-scale features with higher spatial resolution.

$\mathrm{Fea}_{\mathrm{i}}^{\mathrm{j(2)}}$ was merged by another two features of image $\mathrm{A_j}$ within various scales sequentially through selection to create the fused feature $\mathrm{Fea}_{\mathrm{i}}^{\mathrm{j(3)}}$. Furthermore, the fused feature feed to the transformer module and CAM to create the feature map $\mathrm{Fea}_{\mathrm{i}}^{\mathrm{j(4)}}$. This transformer modal uses an encoder and decoder blocks. The proposed CD network utilizes several self-configuring transformer modules. This work utilizes the SAM and transformer framework to represent both spatial and global context data. This CAM method transmit the context data through highlighting the transmits correlated to the variations. We define the CAM in a detailed manner below: the numerous features provide similar channel attention $\mathrm{N_d}$ for image $\mathrm{Y_j}$. In order to determine the channel attention, initially we combine the feature maps of similar scale of multiple branches through employing element-wise summation and apply max pooling together with the spatial dimension of the combined results. Furthermore, we implement an element-wise summation once again to combine the multiscale outcomes of the max-pooling operation and permits the combination outcomes by a multi-layer perception (MLP) to acquire the attention channel $\mathrm{Y_j}$. This MLP contains a full convolution layer with a sigmoid activation function and a full convolution layer with a ReLU activation function. The detailed process is formulated in the following. Let $\mathrm{U}(\mathrm{Fea}_{\mathrm{i}}^{\mathrm{j(3)}})$ represent the feature map gathered through transmitting $\mathrm{Fea}_{\mathrm{i}}^{\mathrm{j(3)}}$ within the transformer method. The max pooling result of the combined feature of $\mathrm{U}(\mathrm{Fea}_1^{\mathrm{j(3)}})$ and $\mathrm{U}(\mathrm{Fea}_2^{\mathrm{j(3)}})$ can be denoted as:

$$\mathrm{N_i} = \mathrm{MaxP}(\ \mathrm{U}(\mathrm{Fea}_1^{\mathrm{j(3)}}) \oplus \mathrm{U}(\mathrm{Fea}_2^{\mathrm{j(3)}})), \qquad (8)$$

Thus $\oplus$ represents element-wise summation, s represents the radio reduction channels. The attention map channels are:

$$\mathrm{N_i} = \mathrm{noq}(\mathrm{N_1} + \mathrm{N_2} + \mathrm{N_3}) \qquad (9)$$

$$= \alpha(\mathrm{X_2}(\mathrm{ReLU}(\mathrm{X_1}(\mathrm{X_1} + \mathrm{X_2} + \mathrm{X_3})))), \qquad (10)$$

Thus, $\mathrm{X_1} \in \mathcal{S}^{\mathrm{d/s \times d}}$ and $\mathrm{X_2} \in \mathcal{S}^{\mathrm{d \times d/s}}$. And, the feature map $\mathrm{Fea}_{\mathrm{i}}^{\mathrm{j(4)}}$ is attained by the CAM as follows:

$$\mathrm{Fea}_{\mathrm{i}}^{\mathrm{j(4)}} = \mathrm{UFea}_{\mathrm{i}}^{\mathrm{j(3)}} \otimes \mathrm{N_d}. \qquad (11)$$

4. Result and Analysis

We suggest an entire instantaneous of the data from several experiments assumed to evaluate the efficiency of our TRIP-CID model. The sector too delivers supplementary data on the database applied in this learning. Deng 1018 is implemented for calculating an effectiveness of method. We appraised its performance against the publicly accessible "Pest Dataset" to guarantee which our model is generalizable. Table 1 designates the hyper-parameters, which were selected. We qualified our TRIP-CID model by employing SGD. The background is qualified on 80 epochs to distinguish and categorize pests that taking into contemplation over-fitting.

Table 1: Hyper-parameters

Parameter	Value
Shuffle	Each epoch
Optimization algorithm	SGDM
Learning rate	0.002
Maximum Epochs	85
Verbose	False
Validation Frequency	35
Train size	0.9
Test size	0.3

We used the dataset published to calculate the effectiveness of the presented TRIP-CID architecture. It entails of 10 dispersed pest types, which are most commonly realized in tea plants and other plants through Europe and Central Asia. More surely, the dataset's 10 pest images comprise Euproctis pseudoconspersa strand, Locusta migratoria, acrida cinerea, empoasca flavescens, larva of laspeyresia pomonella, spodoptera exigue, parasa lepida, chrysochus chinensi, L.pomonella. Figure 3 illustrations an illustrative depiction of every pest kind. The Deng data encompasses 562 pest pictures. Every type of pests has 40 to 70 images. Figure 4 presentations instance images subsequent rotations. Each by 90 degrees we exchanged the dataset images twice. Table 2 also embraces the number of images for every pest group subsequent rotation. The pest photographs in the collection, which included photos shot with an SLR camera, were gathered from Mendeley and other online sources. The remaining photos were from Dave's Garden, IPM Images, Insert Images, and other sources. The collection includes RGB images at varying resolutions. The example photographs' backgrounds, lighting, posture, and size differ greatly from one another.

Images from Pest Dataset (a) Locus Migratoria, (b) Parasa Lepida, (c) Gypsy Moth Larva, (d) Empoasca Falvscens, Exigua, (e) Spodoptera Exigua, (f) Chrysochus Chinensis, (g) Laspeyresia Pomonella Larva, (h) Spodoptera Exigua Larva, (i) Atractomorpha Sinensis, (j) Laspeyresia Pomonella

Figure 3. Glimpse of Pest dataset

Resized Data from Pest Dataset

Figure 4. Illustration of Resized Data from Pest Dataset

Table 2: Pest category in pest datasets

Pests' category	Without rotations	After rotation
Parasa lepida	60	179
Atractomorpha sinesis	64	188
Locusta migratoria	74	217
Chrysochus chinesis	51	152
Spodopetra exigua larva	58	169
Empoasca flavescens	41	122
Gypsy moth larva	41	122
Laspetresia pomonella	66	197
Spodopetra exigua	70	203
Laspetresia pomonella larva	52	152
Total No 'of images	577	1701

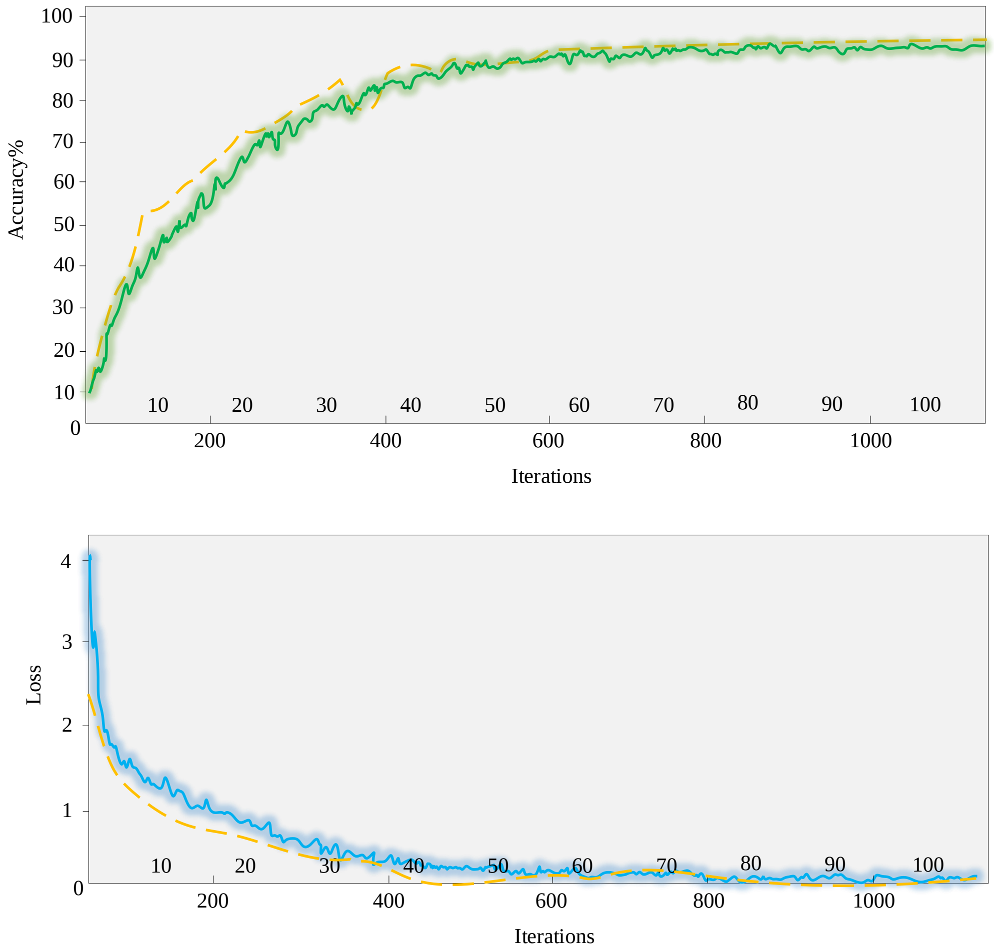

Figure 5. Comparison of accuracy and loss

The method is suggested presentation is evaluated using the following metrics: F1_score (Fs), sensitivity (recall) (Spy), specificity (Spy), accuracy (Acc), precision (Pn), and specificity (Spy). The ratio of the total number of positive photos detected (either correctly or wrongly) by the method to the number of correctly detected or categorized images defines the recommended accuracy model. Specificity is determined by comparing the number of negative photos that are correctly identified (normal) to the total number of negative images in the collection. Lastly, F1_score aggregates recall and accuracy to get a weighted average. The following methods are used to estimate the measures:

$$\mathrm{Acc} = (\mathrm{TNE} + \mathrm{TPE})/\mathrm{TSs} \tag{12}$$

$$\mathrm{Pn} = \frac{\mathrm{TPE}}{\mathrm{TPE+FPE}} \tag{13}$$

$$\mathrm{Ssy} = \frac{\mathrm{TPE}}{\mathrm{TPE+FNE}} \tag{14}$$

$$\mathrm{Spy} = \frac{\mathrm{TNE}}{\mathrm{TNE+FPE}} \tag{15}$$

$$\mathrm{Fs} = 2.\frac{\mathrm{Pn\ x\ Ssy}}{\mathrm{Pn+\ Ssy}} \tag{16}$$

Here, TPE, TNE FPE, FNE, FPE and TSs signifies True positive, True negative, False positive, False Negative and Total samples. Utilizing a machine equipped with an Intel (R) Core (TM) i4-5100U CPU and 6GB of RAM, the suggested methodology is examined and verified. The hardware and software requirements for the TRIP-CID technique that is being discussed are shown in Table 3. Images are downsized to accommodate the down sampled input images required by the models. 10% of the photos were used for testing, while 90% of them were used for training. To train and test the suggested TRIP-CID technique and develop modern models, all studies use training and testing sets with the same experimental circumstances for pest detection and classification as those above. Presentation of the proposed TRIP-CID background for multiclass pest classification is evaluated through a series of studies.

Table 3: Software and Hardware resource

Software/Hardware	Content
CPU	Intel(R) Core (TM) i5-11400H
Operation system	Windows 11
RAM	8 GB
ROM	500 GB
Programming language	MATLAB
MATLAB version	R2020a

Comparing the proposed TRIP-CID method for pest detection and classification to other available DL models is the main objective of this investigation. We evaluated the TRIP-CID framework's classification performance against three pre-trained deep learning models, including Inception Net, VGG-19, and DenseNet. Using a TL configuration, the frameworks are trained on a vast number of images from the ImageNet database. All pre-trained networks are capable of classifying images into a thousand distinct categories. To fine-tune the models, we used the same experimental setup as the TRIP-CID model that was previously described. After rotating the original photos, we used all 1686 crop pest photographs from Deng et al. for our investigation. Additionally, 168 pest photos are used for model testing and 1518 pest images are used for model training. The total number of pest photos utilized for training and validation in each category is then shown. Table 4 shows that Inception Net, VGG-19, and DenseNet performed the worst across all performance procedures when compared to the TRIP-CID model that was provided. The fact that Inception Net, VGG-19, and DenseNet formed the same classification scores for every performance criterion is noteworthy. We discovered that the recommended TRIP-CID framework performed better than the other DL frameworks based on the responses. While the RAF assesses each convolutional layer in the Inception Net, VGG-19, and DenseNet models, their accuracy is lower than that of the recommended model. The Relu function sets all neurons' negative values (x < 0) to zero. It is not certain that every neuron contributing to the dying Relu will be active at all times. In this case, the optimization strategy fails, and the model does not acquire. The reason the fading ReLU problem is problematic is that it leads to a significant percentage of the network becoming idle over time. The proposed TRIP-CID model reports the dying relu issue by using the LR activation function instead of the RAF. Consequently, the TRIP-CID model retains all neurons engaged even when proposed with negative values. Moreover, the proposed model applies BN following each convolutional layer. Even if some qualities are more significant numerically than others, their importance is reserved. Importantly, the proposed model will be much unbiased (with respect to higher-value features). Furthermore, when compared to a framework that rejects BN, this approach allows for faster training and higher accuracy.

Table 4: Comparison of Proposed Deep learning Model with Existing Model

Model	**ccuracy**	**Precision**	**Recall**	**F1 Score**
Inception Net	85.6%	87.7%	88.5%	86.3%
VGG-19	87.6%	88.9%	87.6%	90.6%
DenseNet	90.2%	92.6%	95.4%	92.5%
TRIP-CID	99.9%	99.7%	.9%	99.8%

Moreover, we tested presented TRIP-CID compared to remaining advanced pest appreciation and organization approaches to guarantee the suggested superiority of model. We appraised the suggested method to the most recent DL frameworks and providing our results in Table 5. The expressions about the comparison of other existing works with our presented TRIP-CID approach. DL models were implemented to categorize eight distinct species of tomato pests. The composed pest characteristics are united with three ML classifiers based on DL models such as DA, SVM and the KNN approach. Bayesian optimization was implemented to routinely regulate the hyper-parameters. The VGG16 approach outclassed the other models and achieving high accuracy after image augmentation. In CNN + ML models, the ResNet50 with DA framework obtained classification. This comparison also validates the achievement of the suggested TRIP-CID paradigm while compared to other techniques. It is worth revealing which these approaches are more computationally exclusive than the presented approach since they necessitate more multifaceted frameworks that can necessarily lead to overfitting. These data validate the suggested technique's efficacy as well as its extra assistances including enhanced computer efficiency. The suggested TRIP-CID approach only comprises 11 layers, shadowed by BN and the LR activation function, therefore all CNN levels' biases are inactive. Therefore, we can accomplish which the suggested TRIP-CID technique is more efficient and effective for finding and categorizing pest images.

Table 5: Comparison of Other Existing Works with Proposed Model for Accuracy

S.No	**Work**	**Dataset**	**Method**	**Date**	**Accuracy**
1	Agripest-yolo [25]	Pest24	CLA	2022	92.35%
2	KBS-Net [27]	Almond Tree Dataset	IPM	2021	93.45%
3	S-RPN [34]	AgriPest21	RPN	2021	95.20%
4	CPD [37]	Pest24	DCNN	2021	95.90%
5	Faster-PestNet [39]	IP 102	RCNN	2023	96.25%
6	TRIP-CID	Deng et.al	IR-Net, PVT	2024	98.70%

We assessed it by implementing the Pest Dataset, which is a typical freely accessible Kaggle dataset to additional examine and estimate the effectiveness and generalizability of the presented TRIP-CID system. The numbers embrace five agricultural pests such as Armyworm, Grasshopper, Bollworm, Mosquito and Stem borer. The dataset is well adjusted and encompasses sufficient images for model training, and we did not conduct any rotations. However, we employed information augmentation for testing the generalizability of the suggested model. The assortment comprises 350 images of every pest class such as Armyworm, Bollworm, Grasshopper, Mosquito, and Stem borer pests. Moreover, an automated script is implemented to produce pest images from Google by utilizing Selenium and Chrome Driver. In this experimentation, we implemented all 3150 crop pest images from the "Pest Dataset" dataset without rotating the original images and 2520 for model training and the remaining 630 for model testing. The presented framework trained for pest classification in 1380 minutes and 12 seconds by employing the "Pest Dataset" dataset. This time is comparative to the highest number of epochs and iterations per epoch. The total number of iterations in training stage of TRIP-CID is 1760 (22 iterations per epoch) with 80 epochs. We achieved average accuracy, precision, recall, and F1-scores. The accuracy validates the efficacy and generalizability of the presented TRIP-CID model for pest detection and categorization. We directed an experiment to regulate the effectiveness of the suggested approach in classifying every agricultural pest. Table 6 illustrations the accuracy, recall, and F1-score performance of the presented approach for class-wise agricultural pest categorization. Additionally, the conclusions establish that the presented model misclassifies just one pest. We operated a 3x3 filter to extract detailed characteristics. The BN method is implemented in the proposed model's feature map standardizes the inputs to every mini-batch, delivers regularization and decreases generalization error. Correspondingly, the dropout approach is employed in the proposed model's classification unit that permits for regularization by deleting a percentage of the outputs of previous layer, avoiding overfitting and encouraging simplification. These results exemplify the efficiency of suggested pest recognition and classification process.

Table 6: Analysis of performance metrics on 5 pest classes

Classes	Pest Name	No of truth	No of Classified	Accuracy	Precision	Recall	F1-Score
1	Armyworm	37	36	99.9%	99.8%	99.9%	99.7%
2	Bollworm	36	36	99.7%	99.9%	99.6%	99.9%
3	Grasshopper	36	36	99.8%	99.6%	99.7%	99.7%
4	Mosquito	36	36	99.7%	99.9%	99.8%	99.9%
5	Stem borer	36	36	99.6%	99.7%	99.7%	99.8%

5. Conclusion

Crop pest has become most challenging problem in modern agriculture. To enhance identification accuracy, we develop a TRIP-CID technique, which utilizes pest images from benchmark datasets for pre-processing. Then the pre-processed images are fed to the Improved ResNet (IR-Net) and Pyramidal vision transformer modules for multi-scale spatial, channel and contextual feature extraction. These extracted features are fused in a higher feature map and it was analysed through a distinct ML classifier. In addition, an ensemble-voting framework guarantees precise pest detection. Our modules precisely detect and classifies pest across ten dataset images such as Euproctis pseudoconspersa strand, Locusta migratoria, acrida cinerea, empoasca flavescens, larva of laspeyresia pomonella, spodoptera exigue, parasa lepida, chrysochus chinensi, L.pomonella. The TRIP-CID technique shows promising results in assisting experts and farmers in saving costs and preventing crop losses.

References

[1] R. Umapathi, B. Park, S. Sonwal, G. M. Rani, Y. Cho, and Y. S. Huh, "Advances in optical-sensing strategies for the on-site detection of pesticides in agricultural foods," *Trends in Food Science & Technology*, vol. 119, pp. 69-89, 2022.

[2] D. Marković et al., "Prediction of pest insect appearance using sensors and machine learning," *Journal of Soft Computing*, vol. 21, no. 14, p. 4846, 2021.

[3] J. W. Chen et al., "A smartphone-based application for scale pest detection using multiple-object detection methods," *Electronics*, vol. 10, no. 4, p. 372, 2021.

[4] M. A. Ali, R. K. Dhanaraj, and A. Nayyar, "A high performance-oriented AI-enabled IoT-based pest detection system using sound analytics in large agricultural fields," *Microprocessors and Microsystems*, vol. 103, p. 104946, 2023.

[5] X. Wang, J. Liu, and X. Zhu, "Early real-time detection algorithm of tomato diseases and pests in the natural environment," *Plant Methods*, vol. 17, no. 1, p. 43, 2021.

[6] M. Turkoglu, B. Yanikoğlu, and D. Hanbay, "PlantDiseaseNet: Convolutional neural network ensemble for plant disease and pest detection," *Signal, Image and Video Processing*, vol. 16, no. 2, pp. 301-309, 2022.

[7] L. Du et al., "A novel object detection model based on faster R-CNN for spodoptera frugiperda according to feeding trace of corn leaves," *International Journal of Information Technology*, vol. 12, no. 2, p. 248, 2022.

[8] L. Liu et al., "Towards densely clustered tiny pest detection in the wild environment," *Neurocomputing*, vol. 490, pp. 400-412, 2022.

[9] A. A. Shahane and Y. S. Shivay, "Soil health and its improvement through novel agronomic and innovative approaches," *Frontiers in Agronomy*, vol. 3, p. 680456, 2021.

[10] J. Kong et al., "A spatial feature-enhanced attention neural network with high-order pooling representation for application in pest and disease recognition," *Nature*, vol. 12, no. 4, p. 500, 2022.

[11] R. Wang et al., "An automatic system for pest recognition and forecasting," *Pest Management Science*, vol. 78, no. 2, pp. 711-721, 2022.

[12] A. B. Smith, J. C. Doe, and L. K. Johnson, "Deep learning for pest detection using convolutional neural networks," *Agricultural Systems*, vol. 190, p. 103103, 2022, doi: 10.1016/j.agsy.2021.103103.

[13] G. Goodwin et al., "Mosquito species identification using convolutional neural networks with a multitiered ensemble model for novel species detection," *Scientific Reports*, vol. 11, no. 1, p. 13656, 2021.

[14] W. Li et al., "Field detection of tiny pests from sticky trap images using deep learning in agricultural greenhouse," *Computers and Electronics in Agriculture*, vol. 183, p. 106048, 2021.

[15] T. Kasinathan, D. Singaraju, and S. R. Uyyala, "Insect classification and detection in field crops using modern machine learning techniques," *Information Processing in Agriculture*, vol. 8, no. 3, pp. 446-457, 2021.

[16] M. Lippi et al., "A YOLO-based pest detection system for precision agriculture," in *2021 29th Mediterranean Conference on Control and Automation (MED)*, 2021, pp. 342-347, IEEE.

[17] K. Wang et al., "New image dataset and new negative sample judgment method for crop pest recognition based on deep learning models," *Ecological Informatics*, vol. 69, p. 101620, 2022.

[18] J. Xu et al., "Forest pest identification based on a new dataset and convolutional neural network model with enhancement strategy," *Computers and Electronics in Agriculture*, vol. 192, p. 106625, 2022.

[19] A. Albanese, M. Nardello, and D. Brunelli, "Automated pest detection with DNN on the edge for precision agriculture," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 11, no. 3, pp. 458-467, 2021.

[20] M. T. Tran, L. Q. Nguyen, and H. P. Le, "A novel deep learning approach for real-time pest detection in precision agriculture," *Computers and Electronics in Agriculture*, vol. 192, p. 106612, 2023, doi: 10.1016/j.compag.2022.106612.

[21] B. Wang, "Identification of crop diseases and insect pests based on deep learning," *Scientific Programming*, vol. 2022, p. 9179998, 2022.

[22] T. Zhu, X. Li, J. Dong, and J. Liu, "Recommending advanced deep learning models for efficient insect pest detection," *Sensors*, vol. 12, no. 7, p. 1065, 2022.

[23] J. Kong, C. Yang, and C. Zhao, "An effective pyramid neural network based on graph-related attentions structure for fine-grained disease and pest identification in intelligent agriculture," *Agriculture*, vol. 13, no. 3, p. 567, 2023.

[24] W. Zhang, H. Huang, Y. Sun, and X. Wu, "AgriPest-YOLO: A rapid light-trap agricultural pest detection method based on deep learning," *Frontiers in Plant Science*, vol. 13, p. 1079384, 2022.

[25] S. Khalid, H. M. Oqaibi, M. Aqib, and Y. Hafeez, "Small pests detection in field crops using deep learning object detection," *Sustainability*, vol. 15, no. 8, p. 6815, 2023.

[26] Sherubha, "Graph Based Event Measurement for Analyzing Distributed Anomalies in Sensor Networks," *Sādhanā*, vol. 45, p. 212, 2020.

[27] J. Chen, W. Chen, A. Zeb, D. Zhang, and Y. A. Nanehkaran, "Crop pest recognition using attention-embedded lightweight network under field conditions," *Applied Entomology and Zoology*, vol. 56, no. 4, pp. 427-442, 2021.

[28] Z. Zheng and C. Zhang, "Electronic noses based on metal oxide semiconductor sensors for detecting crop diseases and insect pests," *Computers and Electronics in Agriculture*, vol. 197, p. 106988, 2022.

[29] S. MacDougall, F. Bayansal, and A. Ahmadi, "Emerging methods of monitoring volatile organic compounds for detection of plant pests and disease," *Biosensors*, vol. 12, no. 4, p. 239, 2022.

[30] M. E. A. M. Al-Sabti, M. A. Al-Harbi, and H. M. Al-Turki, "AI-based pest monitoring system for precision agriculture," *Agricultural Engineering International: CIGR Journal*, vol. 24, no. 1, pp. 123-135, 2022.