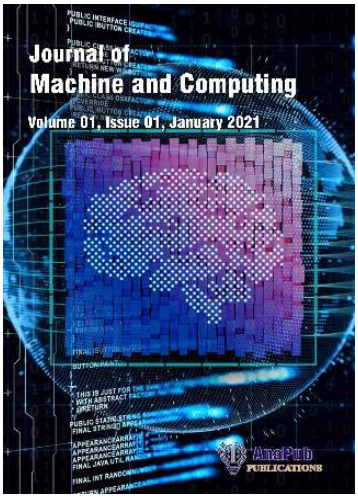
Journal Pre-proof

ConvViT-Driven Multi-Context Feature Fusion for Sustainable Pest Monitoring in Agriculture



Konkala Divya and Reddy Madhavi K

DOI: 10.53759/7669/jmc202505105
Reference: JMC202505105
Journal: Journal of Machine and Computing.

Received 18 November 2024
Revised form 16 February 2025
Accepted 14 March 2025

Please cite this article as: Konkala Divya and Reddy Madhavi K, "ConvViT-Driven Multi-Context Feature Fusion for Sustainable Pest Monitoring in Agriculture", Journal of Machine and Computing. (2025). Doi: https:// doi.org/10.53759/7669/jmc202505105.

ConvViT-Driven Multi-Context Feature Fusion for Sustainable Pest Monitoring in Agriculture

[1]Konkala Divya, [2]K. Reddy Madhavi

[1]Research scholar, Department of CSE, School of Computing, Mohan Babu University, Tirupati, India.
[2]Professor, AI&ML, School of Computing, Mohan Babu University, Tirupati, India
[1]konkaladivyareddy25@gmail.com, [2] kreddymadhavi@gmail.com

Abstract –

In the last few years, the union of modern imaging technology and AI has given rise to agriculture. Probably the most promising of its uses is AI-powered models in agricultural pest imaging, giving new meaning to pest identification, categorization, and monitoring. The world's food security and farming yields are at risk and are endangered by pests, and, too often, this necessitates undue need for pesticides that degrade the environment and the health of people. AI can be brought into play for detecting pests in a new way before they turn invasive, relying less on chemicals and perhaps even ushering in sustainable agricultural methods. Deep learning (DL), a subfield of AI especially designed for image recognition, has seemed especially promising, particularly in the highly precise and highly productive automation of pest detection. In this study, the hybrid model known as ConvViT (fusing the local detail extraction strength of Convolutional Neural Networks (CNNs) with the global contextual reasoning power of Vision Transformers (ViTs)) is introduced. To address the challenges from real-world datasets such as background clutter and image quality, viewpoint differences, as well as other exceptions, ConvViT is developed to boost pest classification performance. The proposed framework is based on a framework that shows superior accuracy than traditional models like ResNet50, EfficientNetB3, and standalone ViTs using a curated agricultural pest image dataset. This approach offers an advanced, scalable, intelligent solution for next-generation crop protection by presenting a set of AI capabilities aligned with sustainable agriculture objectives.

Keywords - Precision Agriculture, Agricultural Pest Classification, Deep Learning, ConvViT, CNN, Vision Transformer, Precision Farming, Image-Based Pest Detection, Hybrid AI Architecture, Sustainable Agriculture.

I. INTRODUCTION

Ensuring world food security and economic stability depends much on agriculture. Instead, one of the most urgent problems facing contemporary agriculture is the ongoing danger that pests provide, which greatly lowers crop output and quality [1]. The Food and Agriculture Organization (FAO) claims that up to 40% of yearly crop losses worldwide are caused by pests, therefore severely taxing farmers and agricultural systems [2]. The conventional ways of pest control are almost always a combination of professional efforts and hand monitoring, which lack the element of expediency and are prone to making errors. Furthermore, delayed identification of pests sometimes leads to misuse of chemical pesticides, therefore aggravating environmental damage and the emergence of pesticide-resistant insect species [3].

Recent years have seen the emergence of precision agriculture because of the integration of modern imaging technology and AI [4]. AI-powered models in agricultural pest imaging are one of its most promising uses, redefining pest identification, categorization, and monitoring. Pests threaten food security and farming yields worldwide, frequently leading to excessive pesticide usage that harms the environment and human health. By incorporating AI into pest detection systems, invasive species can be identified early, chemical reliance can be reduced, and sustainable agricultural methods may be promoted. Particularly, DL, a branch of AI best suited for picture recognition, has shown great promise in highly accurate and efficient automation of pest detection [5]. Widely used in many agricultural imaging applications, CNNs are a type of DL model especially good at understanding intricate visual patterns [6]. CNNs have proven adept at capturing fine-grained local features such as wing venation, color patterns, or body morphology. At the same time, ViTs excel at modeling global dependencies across image regions. These complementary capabilities present a unique opportunity for hybrid architectures to unlock a deeper semantic understanding of pest imagery under diverse environmental conditions.

We introduce a novel hybrid model, ConvViT, that is based on the linkage of the local detail extraction ability of CNNs with the global contextual reasoning ability of ViTs. ConvViT is developed to resolve the issues of background clutter, having different perspectives, and inconsistent image quality in the real-world pest dataset to improve pest classification accuracy and robustness. By aligning AI capabilities with the goals of sustainable agriculture, the proposed approach offers an innovative, scalable solution for intelligent crop protection in the era of smart farming.

Our key contributions include:

1. This study introduces a comparative DL framework utilizing three advanced architectures, ResNet50, EfficientNetB3, ViT and proposed hybrid ConvViT Model to classify agricultural pest species precisely. The framework aims to enhance real-time decision-making in precision farming.
2. Leveraging the Agricultural Pests Image Dataset from Kaggle, the study meticulously curated and refined a dataset with comprehensive preprocessing steps to ensure data integrity and model compatibility, including image resizing, normalization, corruption checks, and exploratory data visualization.
3. Advanced data augmentation strategies were implemented to simulate real-world scenarios and improve the model's generalization. Furthermore, Error-Level Analysis (ELA) was applied to assess the fidelity and authenticity of images, showcasing the model's resilience against image quality inconsistencies.
4. Although trained on 12 pest categories, the model's performance was strategically evaluated on the 8 most representative classes to address the class imbalance and focus on pests with high agricultural impact.

The remainder of the document is structured as follows: A thorough analysis of the relevant literature is provided in Section 2. The presented methodology, including dataset collecting, preprocessing procedures, data augmentation techniques, and the application of several DL models, is described in depth in Section 3. The experimental findings are presented and analyzed in Section 4, which also evaluates the comparison model's performance. Section 5 brings the study to a close by outlining the main conclusions and suggesting possible lines of inquiry for further research.

II. Literature Review

Through a focus on potential uses in precision farming, pest control, irrigation, and crop management, this literature review investigates the integration of AI in modern agriculture. It emphasizes important developments, case examples, and typical constraints, including high costs, data reliance, and legal difficulties.

Aijaz et al. [7] focused on how precision farming, machine learning (ML), and robots can be used to improve output, resource efficiency, and sustainability and talked about how AI has changed agriculture. They put together current research and case studies to show AI's potential. For example, they found that wineries were able to increase yields by 25% and save 20% on water use. They looked at IoT statistics and examples from real life, but they didn't do any new studies. Also, Onteddu et al. [8] looked at the role of robots and AI in increasing farming output a lot of secondary data, like case studies, business reports, and academic research, referred to. They put together a summary of the research to show how autonomous systems improve precision farming, resource optimization, and decision-making. For example, AI-driven irrigation uses 20% less water, and autonomous harvesters go 30% faster. Instead of doing new tests, they used qualitative analysis of current studies and looked at trends and overall results. Likewise, Ahuja et al. [9] looked at AI-powered solutions for agricultural irrigation and pest control optimization. Using drone-captured video and sensor data. They detected pests with test accuracies up to 89% using AI models taught on a 24-class pest dataset 785 training. By matching irrigation periods 12–30 minutes to soil moisture and meteorological conditions, they incorporated predictive algorithms for irrigation scheduling, hence lowering water use. Mishra et al. [10] suggested artificial intelligence AI-driven solutions including ML and IoT to solve problems like soil degradation, irrigation inefficiencies, and disease identification in Indian agriculture and analysis and Crop Management. They automated soil analysis, maximized irrigation, and produced predictions using IoT sensors, UAVs, and artificial intelligence algorithms. Working with Microsoft, including the FARMWAVE platform, allegedly raised yields by 30–40%. Furthermore, Hashem et al. [11] are used in agriculture mostly centered on pest and disease identification research on AI. The research on technologies enabling early diagnosis, real-time monitoring, and predictive analytics including IoT, (ML), sensor networks, and computer vision which the authors compiled highlights Important research showing AI's effectiveness including sensor-driven high-accuracy pest detection and image-based disease classification in banana crops. Emphasizing AI's part in decision support systems and scalability, methods included

case studies and methodical evaluations of publications with peer review. Gupta et al. [12] examined how artificial intelligence is being used in plant sciences with an eye on precision agriculture, disease detection, genomics, and phenotyping. To underline artificial intelligence technologies like ML, IoT sensors for environmental monitoring, and blockchain for data integrity, the authors combined previously published studies. Techniques included methodical study of case studies and peer-reviewed papers with very accurate samples. Spagnolo et al. [13] looked at how IoT and artificial intelligence may maximize agricultural methods through real-time data analysis, predictive modeling, and automation. With a case study of a smart farm in India and a farmer questionnaire, the authors mixed-methodically investigated IoT sensor data soil moisture, weather stations, drones, and AI-driven analytics ML for disease diagnosis, and yield prediction. Thirty percent water savings, eighteen to twenty-five percent production increases, and forty percent pesticide reduction were among the notable gains shown. Patil et al. [14] looked at how artificial intelligence technologies ML, IoT, drones, and predictive analytics might help to advance precision agriculture for climate resilience. Emphasizing their ability to maximize resource usage and lower climate risks, the paper synthesizes current uses of artificial intelligence in soil health monitoring, weather forecasting, insect detection, and genomic crop breeding. Though particular accuracy measures or empirical validations are not given, the approach consists of a qualitative assessment of AI tools and case studies IoT-enabled irrigation systems, and AI-driven disease detection models.

Moreover, in AI-Powered Revolution in Agricultural Pest Imaging, there is limited comparative analysis using traditional methods.

1. AI-driven agricultural systems are not available for smallholders and farmers in low-income areas as their hardware and software infrastructure frequently require large upfront expenses.
2. High-quality, real-time data from IoT devices, UAVs, and sensors is what defines effective artificial intelligence performance. Particularly in rural locations with inadequate connectivity, inconsistent, unbalanced, or poor-quality data compromises model accuracy and scalability.
3. More specifically in rural or underprivileged areas, farmers' limited digital literacy and lack of technical knowledge hinder the efficient adoption and use of AI technology.

III. METHODOLOGY

This study employs DL-based models, including ResNet50, EfficientNetB3, ViT and Proposed Hybrid ConvViT Model for automated pest classification using the Agricultural Pests Image Dataset. The dataset was preprocessed, augmented with transformations like flipping and zooming, and split into training, validation, and testing sets. Figure 1 graphically represents the overall methodology of the pest image classification.

3.1 Experimental Data: In this study, we employed the *Agricultural Pests Image Dataset*, sourced from Kaggle, comprising a total of 5,494 annotated images spanning 12 pest categories, including Ants, Bees, Beetles, Caterpillars, Earthworms, Earwigs, Grasshoppers, Moths, Slugs, Snails, Wasps, and Weevils. The images were initially selected using the Flickr API, a popular image-sharing website that ensures the visual information is genuine and representative of actual agricultural environments rather than artificially produced or excessively edited examples. For large-scale training processes, each image has been uniformly reduced to a maximum dimension of 300 pixels (width or height) to maintain computing efficiency while preserving significant visual aspects. This dataset's compact nature and class-wise distribution make it an ideal candidate for experimentation across various computer vision architectures and ML pipelines.

Although our dataset initially comprises 12 distinct classes, namely *beetle*, *grasshopper*, *earthworms*, *ants*, *earwig*, *snail*, *caterpillar*, *weevil*, *bees*, *moth*, *wasp*, and *slug*, for evaluation and performance reporting, we have focused on a subset of 8 representative classes: *beetle*, *grasshopper*, *earthworms*, *ants*, *earwig*, *snail*, *catterpillar*, and *weevil*. This selection was made either due to class imbalance, lower sample representation in specific categories *(e.g., bees, moths, wasps, slugs)*, or to streamline the analysis toward the most frequently occurring and relevant pest species in the given context. Nonetheless, the model was trained on all 12 classes, ensuring its generalization capability across the whole label space. Figure 2 displays some samples of the dataset.

3.2 Data Preprocessing*:* To begin our pest classification process, we first organized the dataset by converting image file locations and the class labels that go with them into a well-organized data frame. The path of every image file was preserved in the *'Filepath'* column, and the *'Label'* column was assigned the corresponding pest class extracted from the folder structure. Preprocessing, visualization, and model training were among the latter processes facilitated more easily by this simplified form.

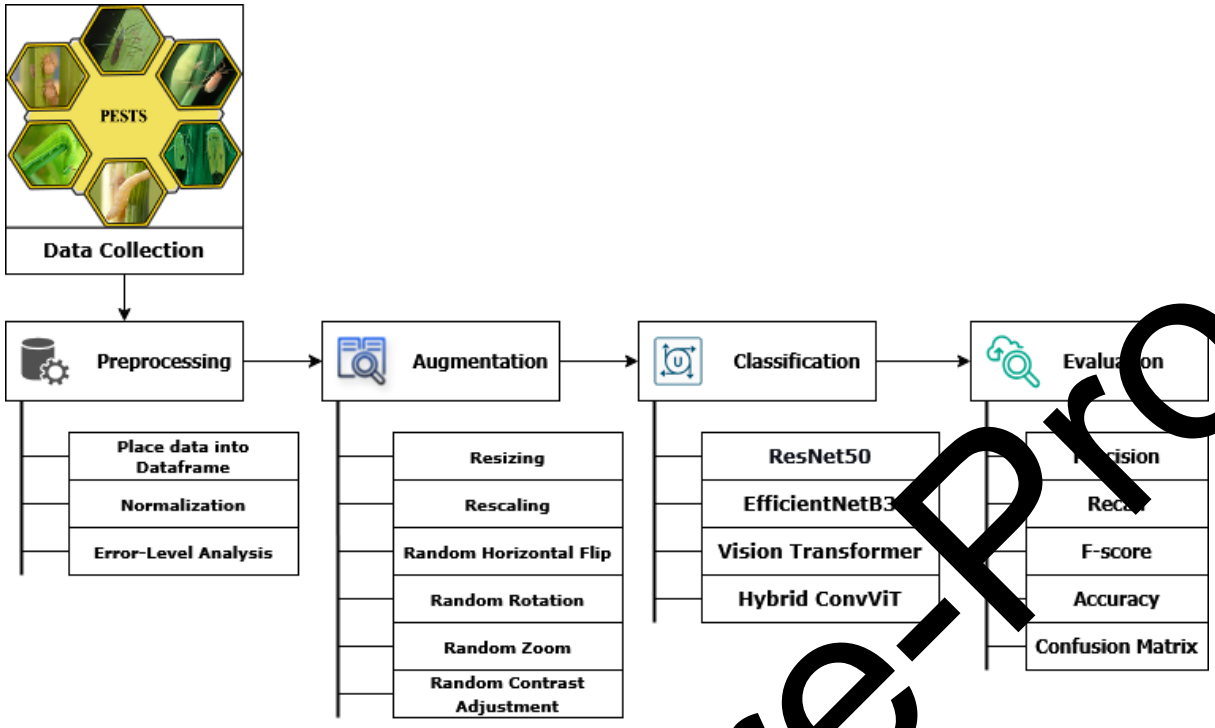


Figure 1: Graphical representation of the overall research methodology

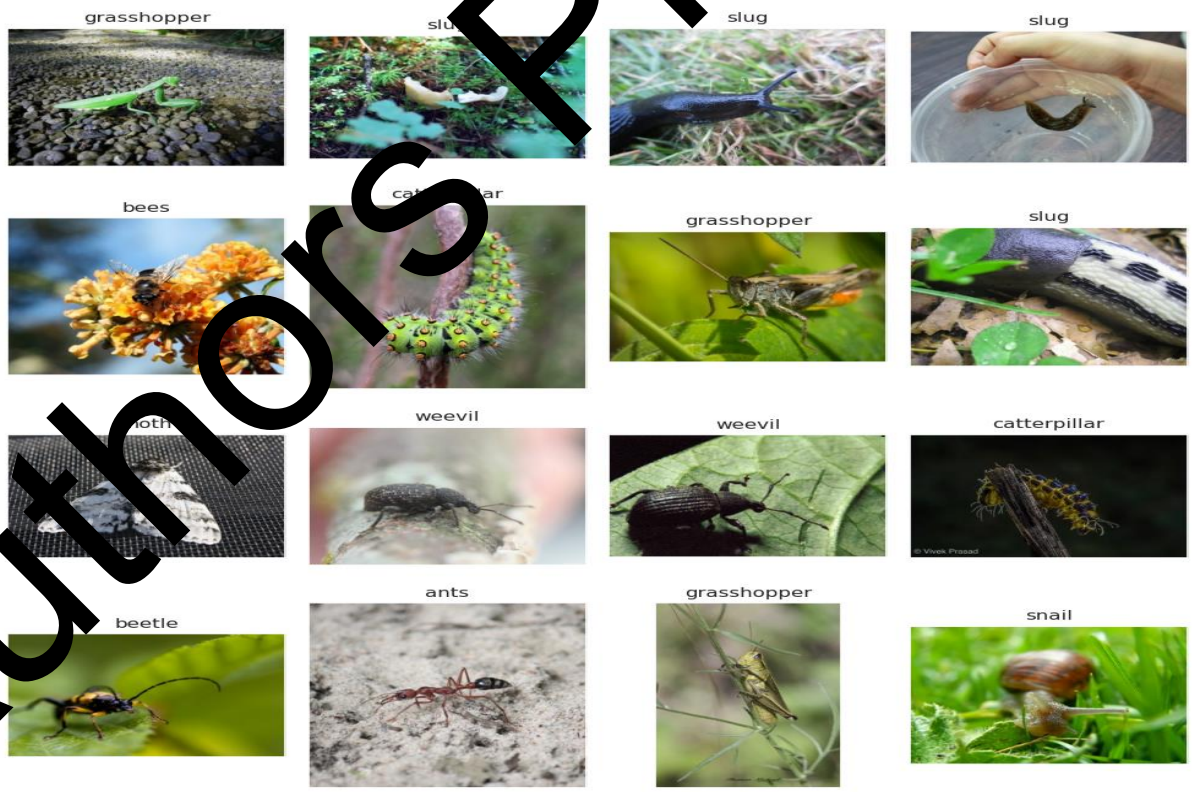


Figure 2: Sample of the dataset

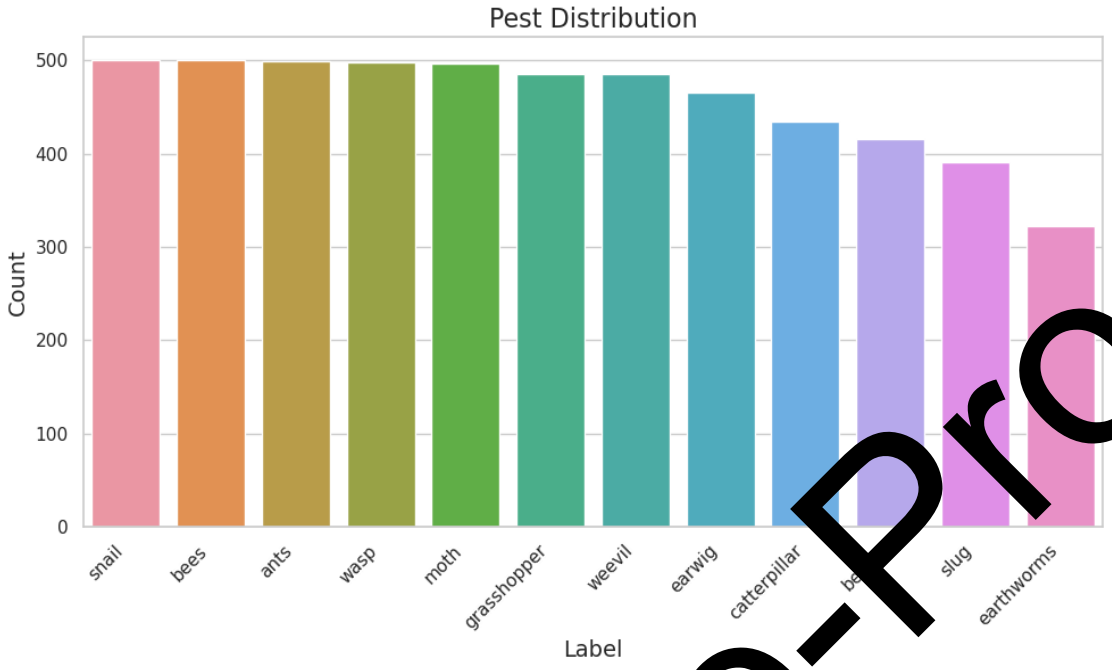


Figure 3: Distribution of labels in the image dataset

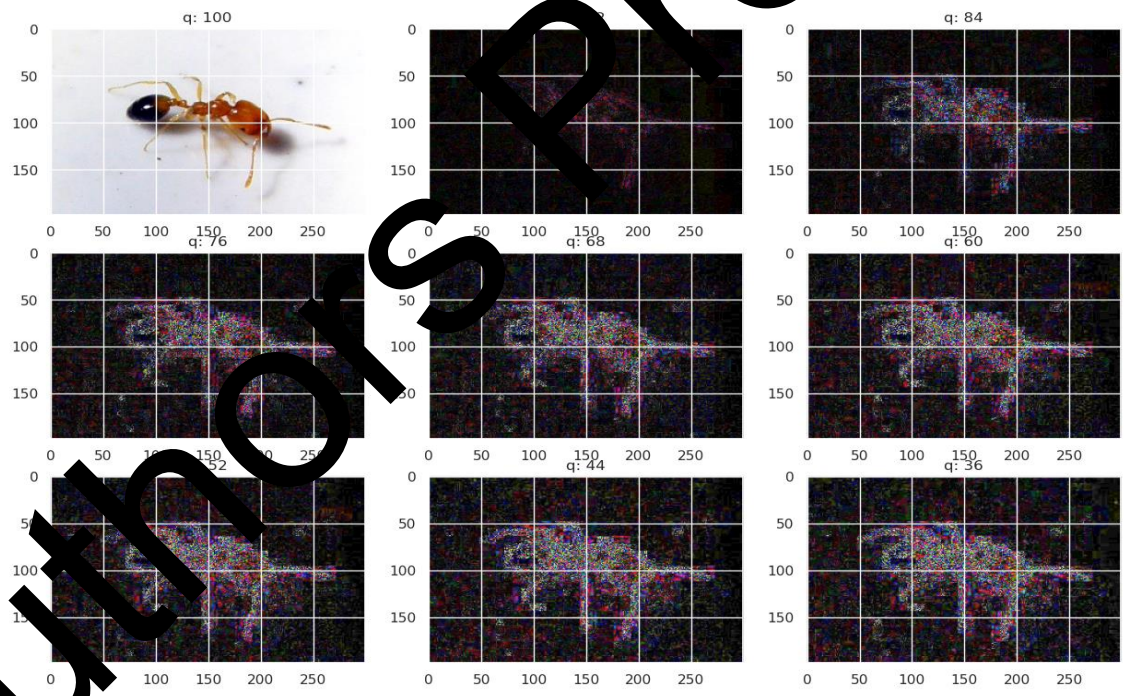


Figure 4: Visual representation of a random sample from the dataset

We used the '*ImageDataGenerator'* utility to normalize and prepare the images in a format appropriate for the model and scale the pixel values suitably. The images were resized to a uniform shape of 224×224 pixels and organized into RGB color mode, supporting consistent feature representation across samples.To ensure data integrity, we ran a corruption check on the whole dataset, using the PIL package to find any damaged or unreadable images. The balance among the 12 pest classes was visually evaluated using a label distribution plot in Figure 3, which offered information on class representation and any imbalances.

Besides, we randomly sampled 16 images and visualized them in a 4x4 grid layout, and each image was depicted alongside its corresponding label, presenting an intuitive glance at the dataset's intra-class and inter-class visual diversity. In addition to basic inspection, we utilized Error-Level Analysis (ELA), which was frequently employed to reveal hidden manipulations or quality inconsistencies. The '*compute_ela_cv()'* function generated ELA images by compressing the original image at varying JPEG quality levels and computing pixel-wise differences. To demonstrate the slight variations in image fidelity, we created a grid of ELA images spanning declining quality levels using a randomly chosen pest image. Figure 4 displays the random sample from the dataset.

3.4 Data Augmentation: We implemented an implementation technique to improve the model's generalization capabilities and reduce the possibility of overfitting. This method provides subtle time adjustments to the images while training, enriching the training set. 224 x 224 pixels. Each image was first resized from the input data. The values of pixels are then standardized to the [0, 1] range, which allows for numerical stability and quicker convergence.

Next, we replicated real-world variability by applying a series of randomized changes. The model was more robust to orientation changes by including small rotations (within ±10%) and horizontal flips to account for the mirrored appearance of issues. Additionally, random zooming and contrast adjustments were involved, supporting the model's adaptation to scale changes and varying lighting conditions generally encountered in natural agricultural environments.

3.5 Model construction: In this study, we leverage ResNet50 [15], EfficientNetB3 [16], ViT and Proposed Hybrid ConvViT Model to accurately and efficiently classify agricultural pest images. ResNet50 utilizes deep residual learning to overcome vanishing gradients and extract complex features across layers. With a simplified design that consistently scales depth, width, and resolution, EfficientNetB3 provides good accuracy with a notably smaller number of parameters and lower computational cost. ViT is used for its ability to capture long-range dependencies and global contextual features, making it ideal for precise alignment and representation in agricultural pest image analysis.

3.5.1 ResNet50: A residual neural network (ResNet-50) is used in this study, which has 50 layers of CNNs, MaxPool, followed by a fully connected layer with a softmax layer [17]. However, ResNet builds the network by stacking the remaining connections on top of each other. ResNet50 is also pre-trained on the extensive ImageNet dataset, providing reliable and transferable feature representations that significantly enhance performance and reduce training times for particular classification tasks, especially in scenarios with redundant training data. Figure 5 shows the block of the ResNet model. The model in this work is based on the ResNet50 architecture [18], which introduces the concept of residual learning to improve training in deep neural networks.

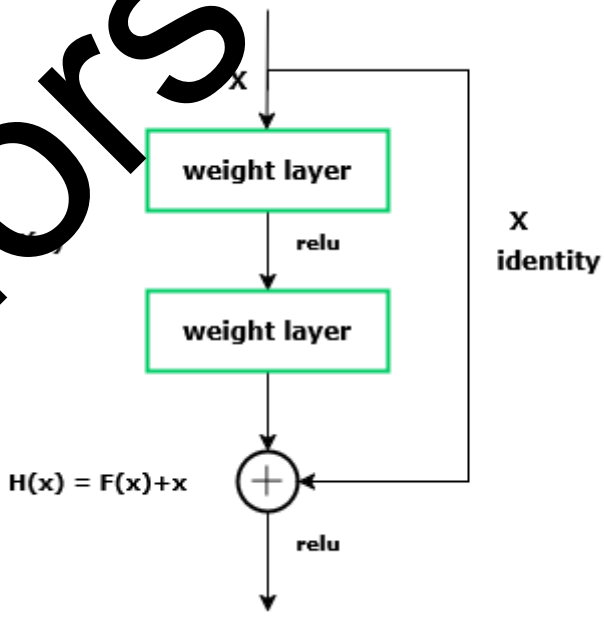


Figure 5: Block of the ResNet architecture

Instead of learning a direct mapping H(x) from input x, a residual block learns a function F(x) such that the final output becomes:

$$y = F(x)+x \quad (1)$$

Here, x is the input, F(x) represents the output of a series of layers, and the term +x is a shortcut connection that allows the input to skip these layers and be added directly to the output. This method allows the network to maintain low-level features while learning complex patterns and helps prevent the vanishing gradient issue in very deep architectures. The residual function F(x) is often expressed as:

$$F(x) = W_2.\sigma\,(W_1.x) \quad (2)$$

Where, W_1 and W_2 are convolutional weight matrices, and σ is a non-linear activation function such as ReLU. After computing F(x), the shortcut adds x directly to it, resulting in the final output:

$$Y = \sigma(F(x)+x) \quad (3)$$

In this project, a pretrained ResNet50 backbone is employed with frozen weights (from ImageNet), and a custom classification head is added, including data augmentation, dense layers with ReLU activations, dropout for regularization, and a softmax output layer for multiclass prediction over 12 categories.

3.5.2 EfficientNetB3: EfficientNetB3 is used in this study to facilitate the intelligent identification and classification of agricultural pests, assisting in developing AI-powered precision farming systems. Figure 6 displays the EfficientNetB3 model's framework.

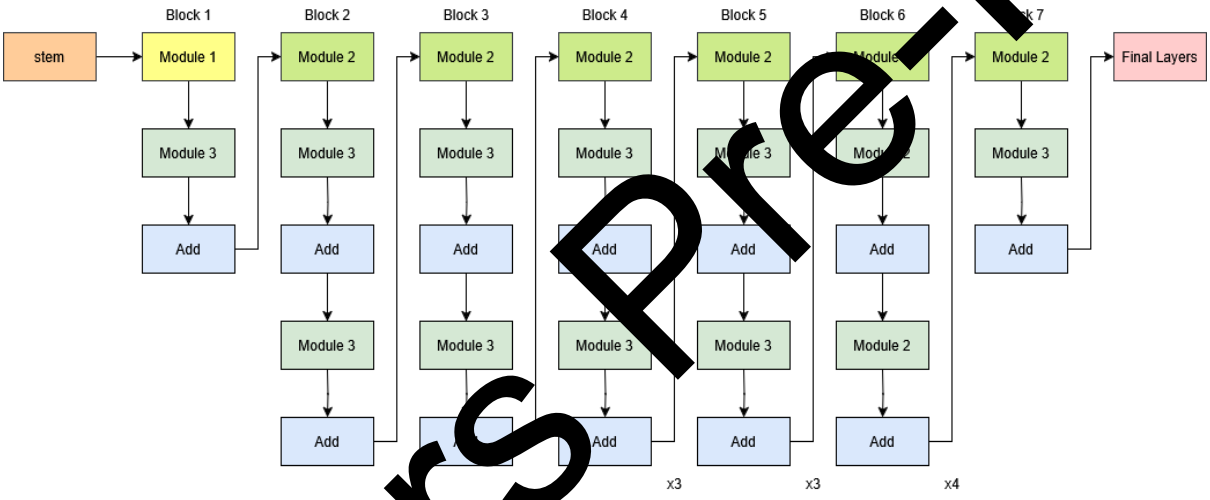


Figure 6 General architecture of the EfficientNetB3 model

EfficientNet is renowned for its novel compound scaling technique, which uses a compound coefficient Φ to equally modify the network's depth d, width w, and input resolution r. The scaling strategy follows the connection:

$$d = \alpha^{\Phi},\ w = \beta^{\Phi},\ r = \gamma^{\Phi} \quad \text{subject to } \alpha.\beta^2.\gamma^2 \approx 2 \quad (5)$$

where α, β, and γ are constants that were found using a grid search to maximize the model's efficiency and accuracy. As a scaled variant, EfficientNetB3 maintains a better balance between computing cost and accuracy, which makes it perfect for agricultural tasks used in environments with restricted resources, such as farms and mobile edge devices. EfficientNetB3's architecture leverages inverted bottleneck blocks and depthwise separable convolutions, enhancing feature extraction from complex pest imagery with fewer parameters [19] Using Leaky ReLU as the activation function improves learning stability, and the activation function is defined as:

$$f(x) = \begin{cases} x, & \text{if } x \geq 0 \\ \theta x, & \text{if } x < 0 \end{cases} \quad (6)$$

with θ ∈ [0.01,0.03], a small gradient flow is allowed when activations are negative which supports better convergence and generalization, particularly in diverse agricultural datasets. EfficientNetB3 improves the accuracy of pest classification models by learning discriminative pest-related characteristics under various lighting and background circumstances.

3.5.3 ViT: In the transformative landscape of precision agriculture, where image-driven insights are central to sustainable crop protection, the **ViT** offers a pioneering shift in image recognition by replacing the spatial constraints of CNNs with a self-attention-driven framework [20]. ViT separates every input image $x \in \mathbb{R}^{H\times W\times C}$ into fixed-size patches, flattens them and embeds each patch as a token:

$$x_{patches} = \{x_1, x_2, \ldots . x_N\}, \quad z_0 = [x_{cls}; x_1E; x_2E; \ldots\ldots . x_NE] + E_{pos} \quad (7)$$

These tokens are then processed using multi-head self-attention (MSA) layers, allowing the model to capture both short- and long-range dependencies across the image. Unlike traditional CNNs, which rely on local receptive fields and manually designed filters, ViT learns from the data itself—without enforcing any task-specific inductive bias:

$$z_l = MSA\,(LN(z_{l-1})) + z_{l-1},\ z_{l+1} = MLP\,(LNz_l)) + z_l \quad (8)$$

This architecture proves especially effective in agricultural pest imaging, where identifying subtle, spatially distant symptoms such as irregular texture patterns, leaf edge deformation, or pigment inconsistencies is critical. ViT's ability to attend across the entire visual field enables it to detect these signs with heightened sensitivity and robustness. ViT exhibits remarkable accuracy and flexibility when pre-trained on extensive datasets and refined on smaller, specific pest image collections. Its performance on agricultural datasets demonstrates its effectiveness in transfer learning, making it a potent part of actual pest monitoring systems.

3.5.4 Proposed Hybrid ConvViT Model: Automatic pest detection is a crucial aspect of environmentally responsible and sustainable agriculture, and the continuous development of artificial intelligence in image processing has revealed game-changing possibilities. With the ability to function with accuracy and minimal ecological disturbance, intelligent detection systems are gradually replacing conventional pesticide-dependent methods.

This study introduces a hybrid ConvViT model that integrates the local feature extraction power of CNNs with the global dependency modeling capacity of ViTs to address the limitations inherent in conventional CNNs and maximize classification performance under complex imaging conditions. The ConvViT framework is designed to capture complementary information from both spatially localized patterns and broader contextual cues. The CNN component of the hybrid structure extracts hierarchical features through successive convolutional and max-pooling layers, effectively capturing detailed textures, shapes, and edges relevant to pest morphology. Figure 7 illustrates the architecture of the proposed ConViT model.

Let, an input image $I \in \mathbb{R}^{H\times W\times C}$ be processed through the CNN produce local feature maps:

$$F_{CNN} = Maxpool\,(\sigma\,(W_c * I + b_c)) \quad (9)$$

where, W_c and b_c denote the convolution kernel and bias, σ represents a non-linear activation.
The resultant local features F_{WN} are then passed to the ViT, which restructures the spatial map into a sequence of flattened, linearly embedded image patches:

$$z_0 = [x_{cls}; x_1E; x_2E; \ldots\ldots . x_NE] + E_{pos} \quad (10)$$

These tokens are subsequently processed through multi-head self-attention (MSA) and multi-layer perceptrons (MLPs) within the ViT, facilitating the modeling of global dependencies and long-range spatial relationships across the image:

$$z_l = MSA\,(LN(z_{l-1})) + z_{l-1},\ z_{l+1} = MLP\,(LNz_l)) + z_l \quad (11)$$

A feature concatenation strategy is employed to harmonize local and global information. The global class token from ViT and the globally averaged pooled (GAP) output of the CNN are concatenated into a unified feature representation:

$$f_{concat} = [GAP\,(F_{CNN}) \,||\, x_{cls}] \quad (12)$$

This incorporated vector f_{concat} is passed through a fully connected layer and a softmax activation to produce the final multi-class classification output:

$$\hat{y} = Softmax\,(W_f f_{concat} + b_f) \quad (13)$$

where W_f and b_f are trainable weights and bias parameters, respectively.

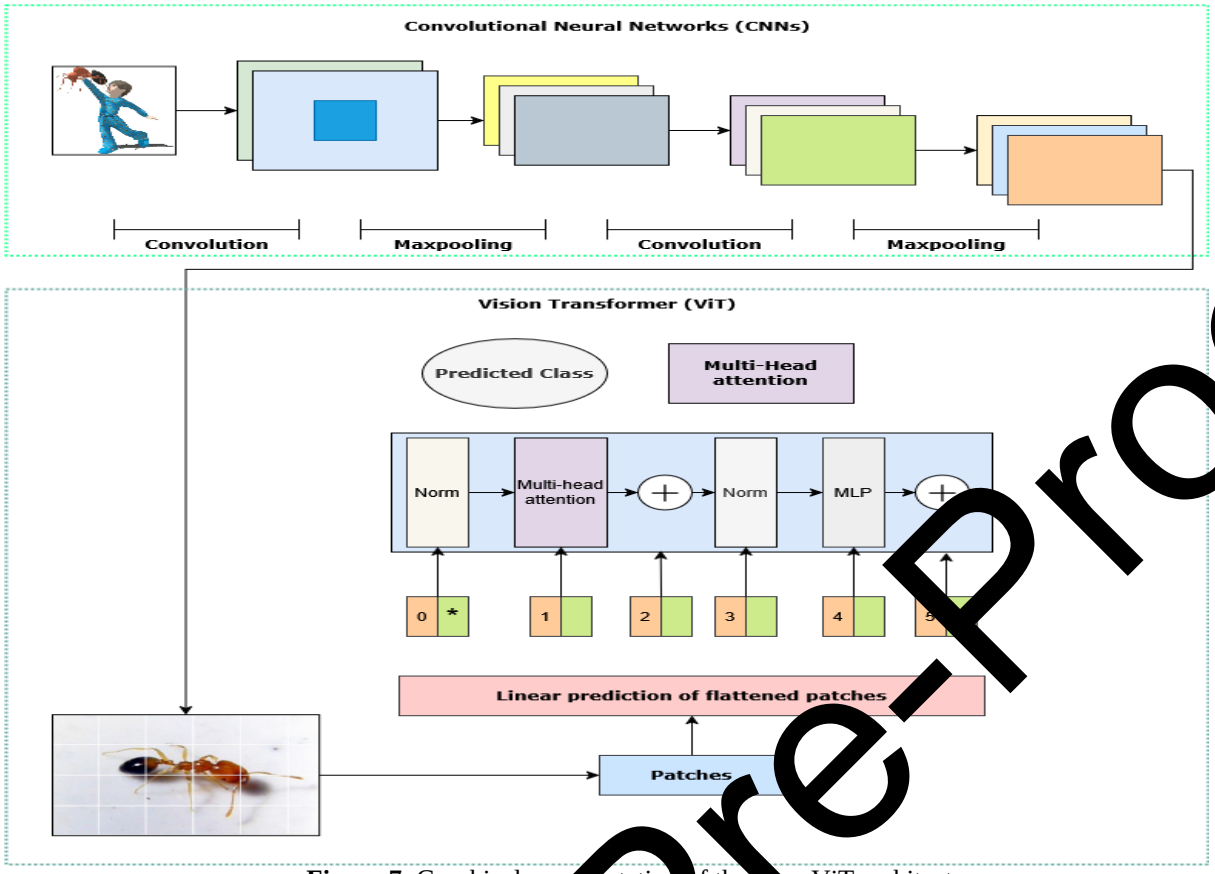


Figure 7: Graphical representation of the ConvViT architecture

Table 1: The optimized parameters of the proposed model

Hyperparameters of the Proposed ConvViT Model	
CNN	
Kernel Size	3 × 3
Pooling Size	2 × 2
Convolution Layers	3
Filter Size	64
Dropout	0.2
Activation Function	ReLU
ViT	
Learning Rate	0.0001
Patch Size	16 × 16
Embedding Size	768
Attention Heads	12
Transformer Layers	8
MLP Hidden Layer Size	3072
Dropout	0.1
Optimizer	AdamW
Batch Size	32
Epochs	100

The model processes raw input images by normalizing pixel values and transforming labels into tensors. Feature maps obtained via CNN are transferred directly into the ViT module, bypassing traditional ViT patch splitting, thereby allowing attention computation on semantically enriched representations rather than raw pixel patches. This hybrid model presents a novel, scalable approach to enhancing classification robustness in agricultural pest diagnostics. The dataset was split into

80% for training and 20% for testing, with 10% of the training data used for validation. Hyperparameters were optimized via grid search. Categorical cross-entropy served as the loss function, and argmax was applied to output probabilities for final class prediction. Table 1 presents the optimized parameters of the models.

IV. Result and Analysis

In this section, the results of this experimental work and the performance of the proposed framework for the agricultural pest classification using AI are presented. Rigorous tests on the diversity dataset of different pest species have been carried out to validate the model as an effective tool to enhance precision farming and sustainable crop protection. It evaluates the classification performance using accuracy, precision, recall, and F1-score. Further, the proposed approach based on CNN and ViT has also been compared with other state-of-the-art models such as ResNet50 and EfficientNetB3 to emphasize the superiority of the proposed approach. Next, there are subsections that give much more detailed discussions on model performance, learning behaviors, and comparative metrics, accompanied by appropriate visualizations.

4.1 Experimental Setup

All experiments were conducted on a personal computer with an Intel® Core™ i7 processor, 16 GB RAM, and 256GB SSD. The model was developed and trained using TensorFlow 2.x within a Python 3.x environment running on a Windows 10 operating system. The training pipeline was implemented in Jupyter Notebook, with GPU acceleration enabled to optimize computational efficiency. The InceptionV3 model, pre-trained on ImageNet, was employed with input images resized to 224×224×3. The batch size was 32, and the model was trained for 100 epochs. The SGD algorithm with learning rate = 0.01, momentum = 0.9, and weight decay = 1×10^{-4}, fine-tuning with Adam optimizer with learning rate = 0.0001, was used for optimization. For multi-class classification, we used cross-entropy loss, and early stopping is applied with a patience of 5 epochs to avoid an over fit and keep the best model. To ensure balanced learning, validation, and unbiased evaluation, the dataset was strategically divided into **70% for training, 15% for validation**, and **15% for testing.**

4.2 Evaluation Metrics

Several standard evaluation metrics, including accuracy, precision, recall, and F1-score, were employed to assess the effectiveness of the DL-based models. These metrics provide a comprehensive view of the model's performance regarding correct classification and error minimization. True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) were also calculated to evaluate the models' error analysis.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

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

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

$$\text{F1-score} = 2 \times \frac{Precision \times \text{Recall}}{Precision + Recal}$$

4.3 Performance Comparison

A comprehensive accuracy analysis was made to evaluate and quantify a diverse set of DL architectures toward agricultural pest classification. Four different models, ResNet50, EfficientNetB3, ViT, and the proposed hybrid ConvViT, are put into Table 2 to represent the training, validation, and testing accuracies. All three accuracy performance metrics exhibit clearly results on how the Proposed Hybrid ConvViT performs better than all other models. With this, it achieves a training accuracy of 94.7%, validation accuracy of 89.5%, and testing accuracy of 87.0%. This demonstrates its superior generalization and generalization to unseen data. The difference from the baseline (traditional) ResNet50, which had 80.0% testing accuracy but relatively high 92.5% training accuracy, is significant here, implying some amount of overfitting.

ViT and EfficientNetB3 perform moderately well, and ViT has a testing accuracy of 84.2% compared to 81.9% from EfficientNetB3. This is to show that transformer-based architectures like ViT nicely learn the deeper structure of the features in the placenta imagery. Still, even though the ConvViT hybrid approach leverages the feature extraction power of convolutional networks as well as the global attention mechanism from transformers, it gains much in performance. But the evaluation on ResNet50 and EfficientNetB3 models shows that the accuracy gap between training and validation

amounts to a relatively lower generalization compared to ConvViT. On the contrary, the minimal performance gap in the proposed model highlights the stability and effectiveness of the proposed model for precision agriculture applications, as it is tolerant to the greatest amount of error in this sector.

Figure 8 also visually gives a clearer comparison of the training, validation, and test accuracies of all four models. It is clear, as illustrated in the figure, that the Hybrid ConvViT outperforms its counterparts, especially in terms of validation and testing performance, allowing it to tackle diverse pest imagery efficiently. Aside from making very clear that the proposed hybrid ConvViT outperforms anything else in generalization, this visualization also serves to confirm that this hybrid model could potentially be a viable solution to real-time, AI-powered pest detection with the use of precision agriculture, which in turn will play an essential role in implementing sustainable crop protection practices.

Table 2: Training, validation, and testing accuracies for different models in pest image classification

Model	Training Accuracy (%)	Validation Accuracy (%)	Testing Accuracy (%)
ResNet50	92.5	83.2	80.0
EfficientNetB3	92.7	84.5	81.9
ViT	93.1	85.6	84.2
Proposed Hybrid ConvViT	**94.7**	**89.5**	**87.0**

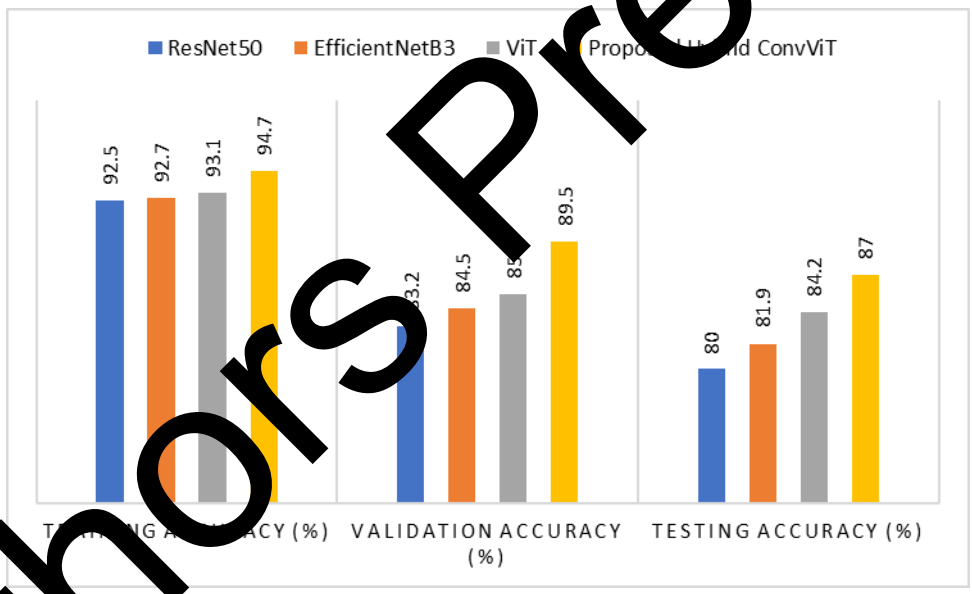


Figure 8: Comparison of accuracies for the different AI-powered models in pest image classification.

The detailed classification performance for pest identification using the ResNet50 model for eight different classes is given in Table 3 in terms of accuracy, precision, recall, and F1-score are well-known evaluation metrics for a good model, in the sense of defining how good the model is at detecting different pest categories in agricultural settings. For some classes, such as beetle, snail, earwig, and caterpillar, the ResNet50 model performs well and has a precision and recall greater than 0.90, with the beetle class achieving the highest precision value (0.95) and F1-score (0.91). This means that the model can determine these specific pest categories very well, making few errors in guessing the true name of an insect or disease.

The problem is that while the model classifies grasshopper and weevil reasonably well with 0.40 precision and 0.45 recall (grasshopper) and 0.35 precision and 0.33 recall (weevil), it fails dramatically in classifying grasshopper and weevil. Such lower metrics indicate that it is hard for the model to learn discriminative features for these classes, as many classes are imbalanced and highly similar between them. Finally, the weighted average here is 0.80, precision of 0.78, recall of 0.75, and F1-score of 0.77. These scores represent a reasonably good level of performance, but the variation across classes can

indicate how the model might improve, for example, advanced data augmentation, class balancing techniques, and adding an attention mechanism to improve its ability to distinguish subtle differences in the pest imagery.

Table 3: Classification Report for the ResNet50 Model

Class	Accuracy	Precision	Recall	F1-Score
beetle	0.80	0.95	0.88	0.91
grasshopper		0.4	0.45	0.42
earthworms		0.89	0.8	0.84
ants		0.87	0.79	0.83
earwig		0.92	0.9	0.91
snail		0.95	0.96	0.95
catterpillar		0.93	0.91	0.92
weevil		0.35	0.33	0.34
Weighted Average	**0.80**	**0.78**	**0.75**	**0.77**

The classification performance of the EfficientNetB3 model for the recognition of pests is presented in Table 4 for the same 8 target classes. Compared with the ResNet50 model, this model achieves a good improvement in the overall and class-wise performance metrics and has a very high score in precision, recall, and F1-score for most categories.

Table 4: Classification Report for the EfficientNetB3 Model

Class	Accuracy	Precision	Recall	F1-Score
beetle	0.819	0.9	0.92	0.95
grasshopper		0.5	0.55	0.52
earthworms		0.95	0.9	0.92
ants		0.94	0.91	0.92
earwig		1	0.98	0.99
snail		0.98	0.99	0.98
catterpillar		0.98	0.96	0.97
weevil		0.45	0.4	0.42
Weighted Average	**0.819**	**0.8475**	**0.8262**	**0.8338**

For example, the model has very good precision and recall for earwig (Precision 1.00, Recall 0.98, F1 score 0.99) and snail (Precision 0.98, Recall 0.99, F1-score 0.98), which means that EfficientNetB3 can easily separate out those classes with few mistakes. Furthermore, the metric values for caterpillar, beetle, ants, and earthworms also indicate that the model is very powerful to generalize on different types of insects. Although we still found that the grasshopper and weevil are persisting challenges for the model, as was the case with the ResNet50 results. Grasshopper achieves a precision and recall of 0.50 and 0.55, while for weevil, it is reduced to 0.45, 0.40, respectively, which shows some confusion with other similar classes or insufficient feature learning due to class imbalance or data quality issues. The average accuracy for the EfficientNetB3 model is 0.819, with a precision of 0.8475, a recall of 0.8262, and an F1-score of 0.8338. Besides ResNet50, these metrics indicate a great improvement compared to EfficientNetB3's ability to extract fine-grained features and work well in complicated visual scenes.

Table 5 shows the performance metrics of the ViT model on eight pest categories, presenting a big advantage in classification accuracy and consistency over what has previously been evaluated. ViT takes advantage of attention mechanisms to learn relationships and dependencies over long distances, reflected in its better performance metrics. The testing accuracy of the ViT model reaches 0.842, which means that the model can find good patterns in the training data and generalize well to unseen instances. This high accuracy is also backed up by great weighted averages of precision (0.8725), recall (0.8562), and F1-score (0.8625), which makes it the strongest standalone DL model in this comparative study, and before integrating the hybrid architecture.

Table 5: Classification Report for the ViT Model

Class	Accuracy	Precision	Recall	F1-Score
beetle	0.842	0.96	0.93	0.95
grasshopper		0.6	0.65	0.62
earthworms		0.96	0.92	0.94
ants		0.97	0.95	0.96
earwig		0.99	0.97	0.98
snail		0.98	0.97	0.97
catterpillar		0.97	0.96	0.96
weevil		0.55	0.5	0.52
Weighted Average	**0.842**	**0.8725**	**0.8562**	**0.8625**

ViT exhibits extremely high precision and recall for the majority of the insect classes on a class-wise basis. For example, the ViT can accurately predict that ants (Precision: 0.97, Recall: 0.95, F1-score: 0.96), earwig (Precision: 0.99, Recall: 0.97, F1-score: 0.98), and earthworms (Precision: 0.96, Recall: 0.92, F1-score: 0.94) have or do not have visible (subtle) variations on some parts of their bodies. Similarly, all metric scores for beetle, snail, and caterpillar are more than 0.95. As with previous observations, however, the grasshopper and weevil classes retain a relatively poor performance. ViT, based on our results, has not yet achieved a sufficiently powerful computation method to solve the problems in prediction accuracy shown by the grasshopper (Precision: 0.60, Recall: 0.65) and the weevil (Precision: 0.55, Recall: 0.50), possibly because the visual features overlap and there is the lack of representative samples.

Overall, ViT shows good potential for classifying all classes, given the transformer architecture's competence in processing global image features. And the average metrics on both show better performance than both ResNet50 and EfficientNetB3, and the ability to perform pest classification highly depends on the DL model used. The classification performance of the proposed hybrid ConvViT, which is formulated by the feature combination of the features of CNNs and ViTs, is presented in Table 6. The proposed hybrid architecture makes full use of the power of feature extraction available in CNNs and the power of global contextual learning offered by ViTs to create a robust and highly performing model that is fit for the complex agricultural pest classification. In terms of accuracy, the highest of all evaluated models is 0.87 for ConvViT overall testing. Additionally, these values of weighted average precision (0.9125), recall (0.8962), and F1-score (0.9012) underline its better classification performance in classifying diverse pest species. When we look at the performance of the hybrid model class-wise, in earwig, snail, and caterpillar detection, the precision, recall, and F1-score values are 1.0 for all three classes, respectively. Since well represented and distinct classes are classified with exceptional fidelity, this shows that the model is capturing fine-grained local features, as well as broad spatial relationships. It also does great on other important classes like ants (Precision: 1.0, Recall: 0.97, F1-score: 0.98) and earthworms (Precision: 1.0, Recall: 0.95, F1-score: 0.97). F1-score of 0.97 shows the good performance of Beetle also in all metrics, which indicates the model's generalization ability.

Table 6: Classification Report for the proposed hybrid ConvViT Model

Class	Accuracy	Precision	Recall	F1-Score
beetle	0.87	1.0	0.95	0.97
grasshopper		0.7	0.75	0.72
earthworms		1.0	0.95	0.97
ants		1.0	0.97	0.98
earwig		1.0	1.0	1.0
snail		1.0	1.0	1.0
catterpillar		1.0	1.0	1.0
weevil		0.6	0.55	0.57
Weighted Average	**0.87**	**0.9125**	**0.8962**	**0.9012**

Moreover, for grasshopper and weevil, which were previously performing poorly in other models, this hybrid approach gives improved metrics. The weevil, although still not as good as grasshopper, however records precision of 0.6, recall of

0.55 and F1 score of 0.57 and this upward of performance may be attributed to a compounded architectural power to handle feature diversity better. Finally, the proposed ConvViT model is far superior to all other architectures in almost every category, suggesting its robustness, scalability, and feasibility for practical agricultural pest detection. Figure 9 also provides further visualization supporting this analysis by providing a comparison of ConvViT versus ResNet50, EfficientNetB3, and ViT, very clear in demonstrating the superiority of the ConvViT model in terms of precision farming requirements.

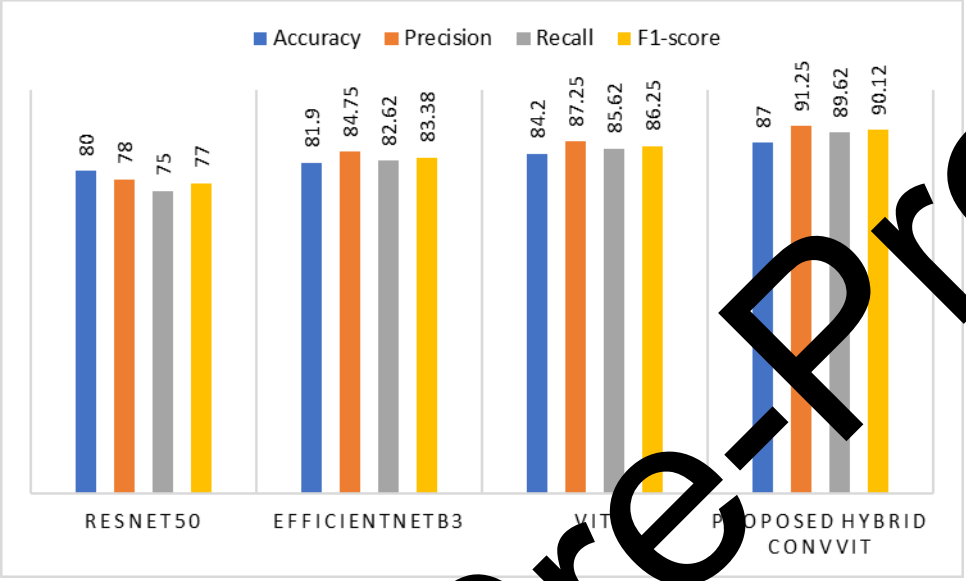
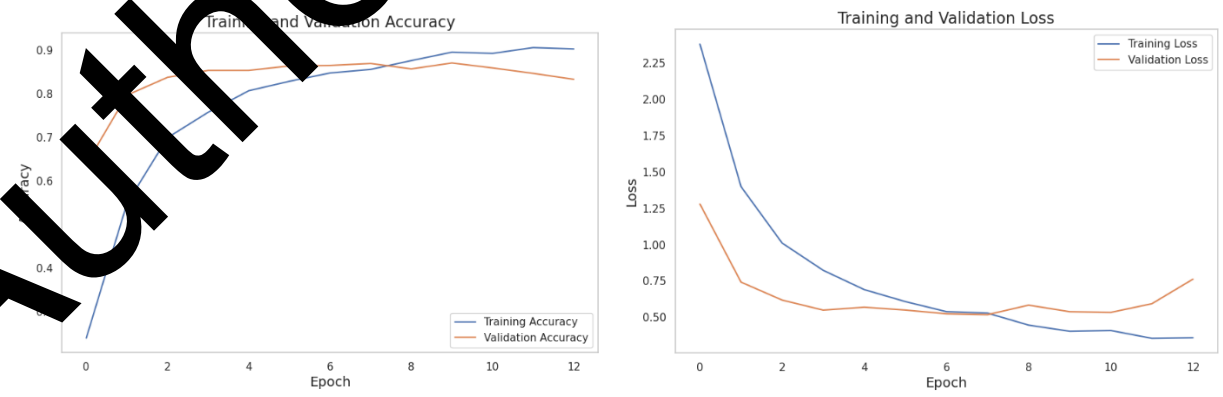


Figure 9: Comparison of classification performance f[illegible] AI-powered models in pest image classification

4.4 Accuracy and Loss Curves Analysis for Propos[illegible] Hyb[illegible] ConvViT Model

We also analyzed the training and validation curves of t[illegible] est-performing model (Hybrid ConvViT) across 12 epochs to further understand the learning dynamics and the generalizat[illegible] behavior of the best-performing model. Figure 10(a) shows the training and validation accuracy trend as well, and Figure 10(b) shows the loss curves.

From Figure 9(a), it is evident that the [illegible] accu[illegible]cy keeps increasing from about 25% to about 90%, however, the validation accuracy starts at about 62% and rises to about 88%, then decreases slightly in the last epoch. This implies that the model had learned these discrimina[illegible]e features from the training data early on and then generalized to the mid-epochs well. However, there is a sig[illegible] of [illegible]mal [illegible]erfitting in the slight dip in validation accuracy, but overall, the trend is still consistent and stable.



(a) **(b)**

Figure 10: Training and validation accuracies and losses over epochs for the proposed hybrid ConvViT model

Secondly, Figure 10(b) depicts a sharp decrease in both training and validation losses through the first few epochs of training, with training loss dropping from approximately 2.3 to below 0.2 and validation loss stabilizing at around 0.5. The training loss is decreased while the validation loss is fluctuating slightly at the end, which agrees with the observed validation accuracy drop in its corresponding curve. However, this pattern is quite common in DL models trained on moderately imbalanced datasets whose performance in some minority classes slightly fluctuates after prolonged training.

All in all, these curves verify that the proposed hybrid ConvViT model is a well-adapted stimulus to learn, having strong convergence characteristics and a high level of generalization. This also clearly demonstrates the model's ability to find meaningful representations with little overfitting.

4.5 Precision Analysis

Figure 11 shows how the precision scores vary as a function of the four DL models (ResNet50, EfficientNetB3, ViT, and the proposed hybrid ConvViT over the eight pest classes (beetle, grasshopper, earthworms, ants, earwig, snail, caterpillar, and weevil). However, precision, defined as the proportion of TPs among all positive predictions, is important to avoid false alarms in real-world pest detection systems where there are many FPs.

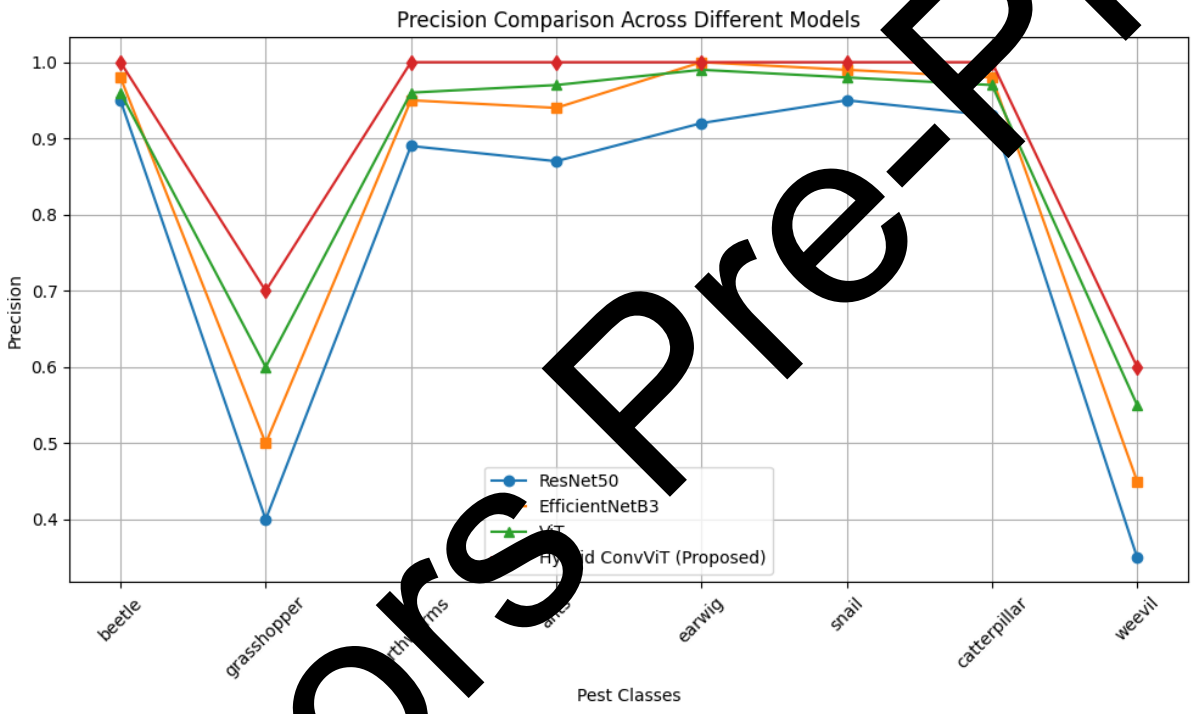


Figure 11 Precision comparison of various DL models across different pest classes

The figure shows that Hybrid ConvViT achieves excellent precision on all classes and a 1.00 score on six of the eight classes. Including beetles, earthworms, ants, earwig, snail, and caterpillar, it is obvious that the model accurately makes highly accurate positive predictions but does not incorrectly classify other classes with the label of such pests. The proposed model still reaches significantly higher precision scores (0.70 for grasshopper and 0.60 for weevil) than other architectures, even for the traditionally difficult to classify categories of grasshopper and weevil. Whereas the precision value, 0.40, 0.35, respectively, for grasshopper and weevil using the ResNet50 model is much lower, as it struggles particularly. On beetle (0.95) and snail (0.95), its performance is decent, but its inconsistent performance indicates that it may lack resilient feature extraction capabilities for small or visually ambiguous classes. With high precision values (above 0.90) in six classes, EfficientNetB3 has better performance than ResNet50 while still not beating grasshopper (0.50) and weevil (0.45). In most categories, we find that ViT, with its attention mechanism powers, defeats both ResNet50 and EfficientNetB3, but still cannot reach the consistent precisions on all classes of the hybrid ConvViT.

4.6 Recall Analysis

Figure 4 is the recall comparison across the same 8 pest classes for the 4 models. The recall metric helps us to recall whether the model can completely detect all the actual positive instances (proportion of TPs out of all the actual positives). Recall

value of a high value means less of FNs, leading to the importance of this measure when it comes to pest detection, due to it spelling possible crop damage or ecosystem imbalance in case a real pest is missed.

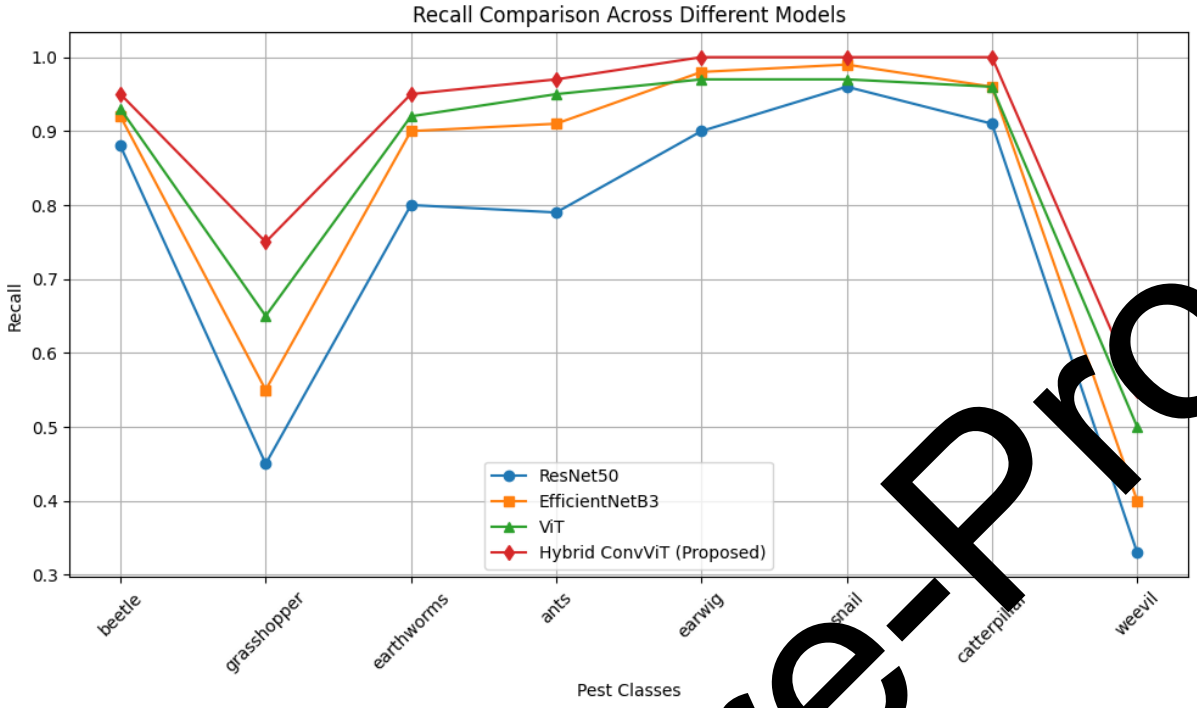


Figure 12: Recall comparison of various DL models across different pest classes

Performance is once again very good (although still not perfect) with Hybrid ConvViT performing best and attaining near-perfect recall (higher than 0.99) in all classes except earwig, snail, and catterpillar, where it had a perfect performance (1.00). Yet, the model also reads surprisingly well on recall of classes that other models crumble on, such as weevil (0.55) and grasshopper (0.75), showing a strong sign of enhanced generalization and robustness in terms of pest detection with arbitrary visual characteristics. Out of the 4, ViT comes in as the second best with high levels of recall (greater than 0.90) for most classes, such as earthworm, ant, and beetle, respectively, at 0.92, 0.95, and 0.93. It, however, does not perform as well as ConvViT in more ambiguous attributes like weevil (0.50) and grasshopper (0.65). Generally, EfficientNetB3 shows reasonable performance over most classes, but has smaller recall for weevil (0.40) and grasshopper (0.55), which indicates a limitation to the narrow class variability for certain classes. Weevil (0.33) and grasshopper (0.45) have the lowest recall for ResNet50, which may indicate shallow learning capacity or sensitivity to data imbalance for these specific classes. Clearly, this figure shows that the proposed ConvViT architecture succeeds in avoiding FN resulting in a more reliable solution in situations where missed detection can have severe repercussions. The high precision together with its good recall makes it a balanced and powerful model.

4.7 F1-score analysis

In Figure 5, the evaluation is combined in consolidating the F1-scores for each pest class. It is a harmonic mean of precision and recall and a single metric whose value balances both FPs and FNs, and is called the F1-score. If class distribution is not balanced and precision and recall are equally important, it is very useful. The hybrid ConvViT model is once again superior to all baselines here, having F1-scores of 1.00 for earwig, snail, and catterpillar, as well as above 0.95 for all other classes except weevil (1.57) and grasshopper (0.72), which still outperform the other models. The high F1 scores indicate that the model is capable of balanced and reliable classification for a wide variety of pest categories.

Commendably, ViT does well in ants (0.96), beetle (0.95), and earthworms (0.94). The slight inconsistencies in classifying visually less distinct pests are again reflected in their scores for weed (0.52) and grasshopper (0.62). Similar to EfficientNetB3, EfficientNetB4 performs well on snail, catterpillar, and earwig, but has some weaknesses in the aforementioned challenging classes. ResNet50 provides an acceptable F1-score in a few classes, but overall it has poor results in F1-score on most of the classes, especially in the grasshopper and weevil classes, with an F1-score below 0.50.

This analysis further verifies the superiority of the proposed hybrid ConvViT in maintaining a harmonious balance among the balance between precision and recall, i.e., achieving both accuracy and completeness in pest classification. In fact, the high F1 scores maintained across most classes guarantees effectiveness in real agricultural settings with these tradeoffs being essential.

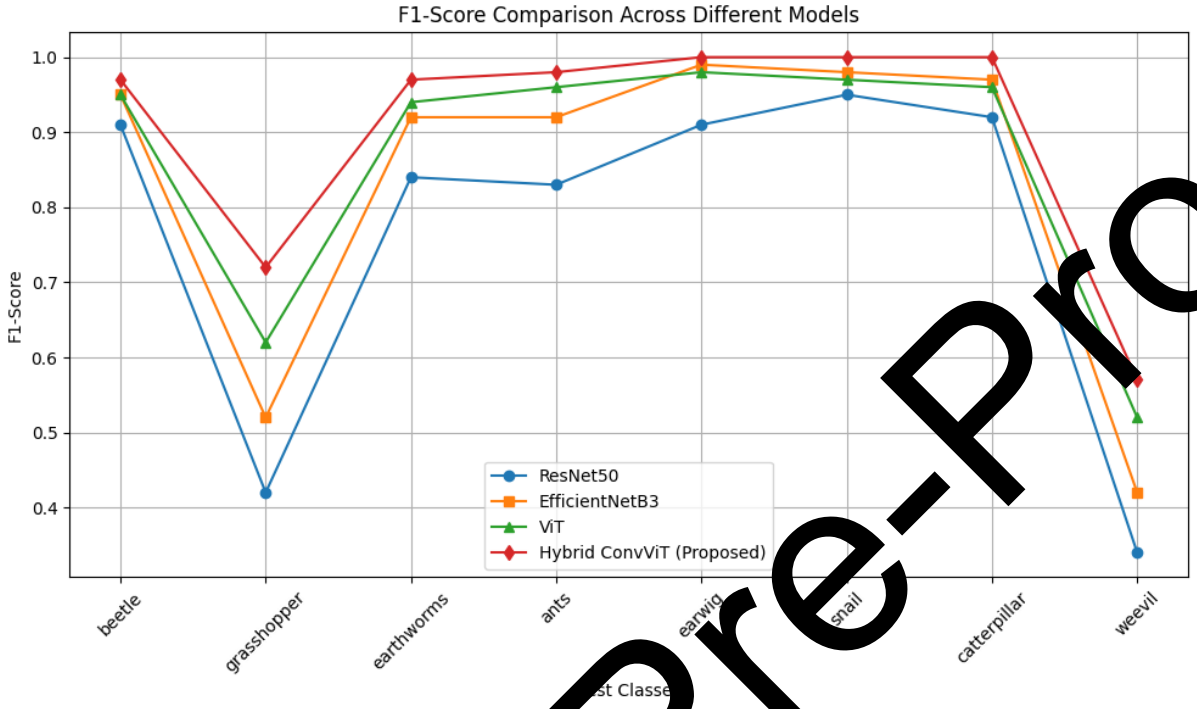


Figure 13: F1-score comparison of various DL models across different pest classes

4.8 Error Analysis for the proposed Hybrid ConvViT Model

A confusion matrix, as shown in Figure 14, was used to perform an error analysis to adequately evaluate the performance and reliability of the proposed hybrid ConvViT model in agricultural pest image classification. This analysis does a good job of exposing the strengths and weaknesses in the classification behavior of this model in TP, FP, TN, and FN values for each pest class.

This confusion matrix shows that the hybrid ConvViT model has almost perfect classification ability for most of the pest categories. In particular, the model performed perfectly for the beetle, earwig, snail, and caterpillar. For all of these classes, there were no FPs or FNs, with TP values of 1 for each class, meaning all samples from these classes were correctly classified with no misclassifications. In addition, the TN values for these classes were very high, which demonstrated the model's robustness.

In the confusion matrix for the grasshopper class, both TP and FP are 0, indicating no sample from this class was present in the test dataset and no prediction was made for this class. Despite that Corporate Control pattern was the one with the lowest classification error in practice with ants and earthworms. In the case of earthworms, the model correctly classified a sample (TP = 1), misclassified one sample as ants (FN = 1), and had a slight overlap in its feature representation for these categories. In the same way, the ants class produced one sample correctly predicted (TP = 1) and one misclassified as a weevil (FN = 1), leading us to suspect visual similarity among these pest types.

Additionally, for the weevil class, the confusion matrix of predictions shows that no prediction was made (TP = 0, FP = 0, FN = 0), in line with the fact that there are no true samples for this class in the test dataset. However, interestingly, the model does not make any FP predictions across any pest category, which is a great strength of the proposed method, meaning that the proposed approach is highly reliable in avoiding incorrect classifications of negative samples.
Overall, the results justify the usefulness and effectiveness of the transformer-based hybrid model for the efficient classification of different agricultural pests. With very few misclassification errors, excluding the earthworms and ants

classes, the model showed very good results in the prediction. However, the level of tolerance to these minor errors is allowable in the context of real-world agricultural environments in which some pest species are quite similar visually. Therefore, the design of pest detection and classification based on the proposed model is found to be very effective, and it helps enlarge the basis of precision farming and sustainable crop protection systems.

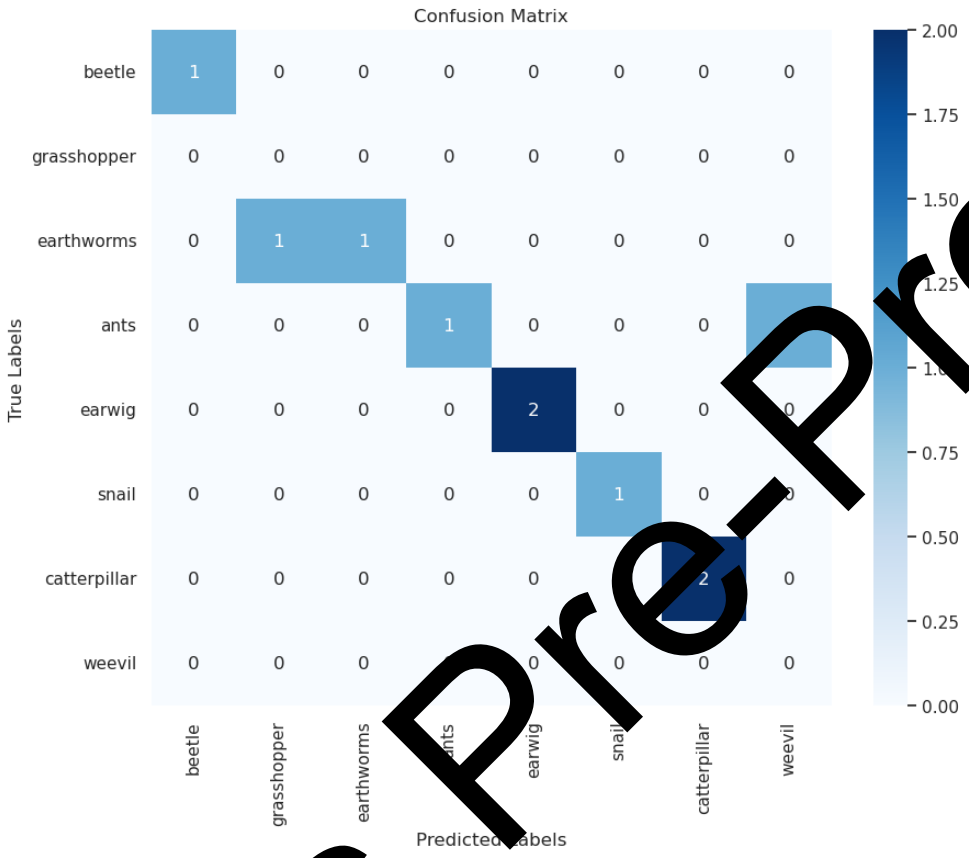


Figure 14: Confusion matrix representing the class-wise prediction performance of the proposed hybrid CovViT model [illegible] agricultural pest images.

4.9 Comparative Analysis

Table 7 highlights a comprehensive comparison of the proposed Hybrid ConvViT model to existing state-of-the-art models published in the literature for the existing pest classification tasks. The models used in the above are ProtoNet, MADN, and RCNN. Classification accuracy metric is used to make the comparison in which is based on the direct measure of the overall correctness of each of the pest models.

Table 7: Accuracy comparison between the proposed Hybrid ConvViT model and existing pest classification models

Reference	Model	Accuracy (%)
[21]	ProtoNet	86.33
[22]	MADN	75.28
[23]	RCNN	67.79
Proposed	**Hybrid ConvViT**	**87.0**

From the table, we found that our proposed Hybrid ConvViT outperforms other methods in accuracy of 87.0%. On the contrary, in comparison to the next best performing model, the ProtoNet model proposed by Gomes and Borges [21] with 86.33% accuracy can be improved in relative terms by 0.77%. The gain appears modest, but the significance of it arises from the fact that such improvements in such imbalanced and complex pest datasets are meaningful in terms of better generalization and classification stability. A final performance result of 75.28% was achieved by the model in Peng et al's research [22], where the MADN architecture was used, which implied that the proposed model improves over this baseline by approximately 11.72%. The combination of convolutional and transformer-based mechanisms proves to be fruitful as they are able to capture both local as well as global feature representations of pest images. Furthermore, the accuracy of

the RCNN model by Xu et al. [23] is the lowest, with a value of 67.79%. In particular, the absolute improvement of the proposed Hybrid ConvViT over the baseline was as high as 19.21%, which indicates a large improvement in classification capability. Overall, this comparative analysis demonstrates that by employing the Hybrid ConvViT, we obtain the state-of-the-art performance and also show an outstanding accuracy improvement compared to the standard DL approaches, as well as with recent works in this context.

V. Conclusion

The subject of classification of agricultural pest images has been shown in this study to be capable of leveraging a novel hybrid DL model combining the local structure of CNNs and global contextual features of ViTs. Compared with ResNet50, EfficientNetB3, as well as standalone ViT models, the proposed model achieved better accuracy and generalization ability. The performances of ConvViT on a carefully curated dataset of 12 pest species (8 out of which are representative are extensively tested in the form of experiments on 8 different classes (including geometrically ambiguous or minority classes) to show that it significantly boosts pest recognition. A pipeline comprising data augmentation, error level analysis, and model optimization solved the problem of image variability, class imbalance, and data quality with a health success rate of its own. The best test results (87.0) for the ConvViT hybrid model surpassed previously published ones like ProtoNet, MADN, and RCNN. This suggests that ConvViTs should be deployed in precision farming, where pests must be identified as early as possible in order to ensure sustainable crop protection. This research has paved the way for further advances in smart agriculture through promoting the development of lightweight and edge-deployable AI solutions, the growth of the datasets in this regard to include more pest species and life stages, and potentially multimodal systems incorporating IoT and sensor data for total agriculture monitoring.

References

[1] B. A. Khan *et al.*, "Pesticides: Impacts on Agriculture Productivity, Environment, and Management Strategies," pp. 109–134, 2023, doi: 10.1007/978-3-031-22269-6_5.

[2] M. D. Junaid and A. F. Gokce, "Global agricultural losses and their causes," *Bulletin of Biological and Allied Sciences Research*, vol. 2024, no. 1, p. 66, 2024.

[3] A. Awad Fahad, "Modern techniques in integrated pest management to achieve sustainable agricultural development," *International Journal of Family Studies, Food Science and Nutrition Health*, vol. 4, no. 1, pp. 1–14, Jun. 2023, doi: 10.21608/IJFSNH.2024.293410.1070.

[4] S. Ashique *et al.*, "Artificial Intelligence Integration with Nanotechnology: A New Frontier for Sustainable and Precision Agriculture," *Curr Nanosci*, vol. 20, no. 2, pp. 242–273, Jan. 2024, doi: 10.2174/0115734137275111[illegible]120607[illegible]CITE/REFWORKS.

[5] K. Sharma and S. K. Shivandu, "Integrating artificial intelligence and Internet of Things (IoT) for enhanced crop monitoring and management in precision agriculture," *Sensors International*, vol. 5, p. 100292, Jan. 2024, doi: 10.1016/J.SINTL.2024.100292.

[6] Y. Liu, H. Pu, and D. W. Sun, "Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices," *Trends Food Sci Technol*, vol. 113, pp. 193–204, Jul. 2021, doi: 10.1016/J.TIFS.2021.04.042.

[7] N. Aijaz, H. Lan, T. Raza, M. Yaqub, R. Iqbal, and M. S. Pathan, "Artificial intelligence in agriculture: Advancing crop productivity and sustainability," *J Agric Food Res*, vol. 20, p. 101762, Apr. 2025, doi: 10.1016/JJAFR.2025.101762.

[8] A. R. Onteddu, R. R. Kundavaram, A. Kamisetty, J. C. S. Gummadi, and A. Manikyala, "Enhancing Agricultural Efficiency with Robotics and AI-Powered Autonomous Farming Systems," *Malaysian Journal of Medical and Biological Research*, vol. 12, no. 1, pp. 7–22, 2025.

[9] S. Ahuja, H. R. Manjunath, I. Alam, and A. Rastogi, "The Role of AI in Modern Farming: Precise Pest Management and Optimal Water Use," *Int. J. Chem. Biochem. Sci*, vol. 25, no. 13, pp. 249–255, 2024.

[10] U. Mishra, "Harnessing AI Technologies for Sustainable Agricultural Practices: Innovations in Soil Analysis and Crop Management," *Biology, Engineering, Medicine and Science Reports*, vol. 11, no. 1, pp. 9–13, Mar. 2025, doi: 10.5530/BEMS.11.1.2.

[11] T. Hashem, J. Joudeh, and A. Ahmad Zamil, “Smart Farming (Ai-Generated) as an Approach to Better Control Pest and Disease Detection in Agriculture: POV Agricultural Institutions,” *Migration Letters*, vol. 21, no. S1, pp. 529–547, 2024.

[12] D. K. Gupta, A. Pagani, P. Zamboni, and A. K. Singh, “AI-powered revolution in plant sciences: advancements, applications, and challenges for sustainable agriculture and food security,” *Open Exploration 2019 2:5*, vol. 2, no. 5, pp. 443–459, Aug. 2024, doi: 10.37349/EFF.2024.00045.

[13] P. Spagnolo, “Advancements in Precision Agriculture: Integrating AI and IoT for Smart Crop Monitoring and Management,” *Europian Journal of Crop Science and Technology P-ISSN 3051-0139 en E-ISSN 3051-01[illegible]*, vol. 1, no. 01, pp. 21–29, 2025.

[14] D. Patil, “Artificial Intelligence Innovations In Precision Farming: Enhancing Climate-Resilient [illegible] Management,” Nov. 2024, doi: 10.2139/SSRN.5057424.

[15] N. Negi, S. K. Singh, and A. Agarwal, “RESNET-50 Based Pest Identification in Plants,” *2024 International Conference on Intelligent Systems and Advanced Applications, ICISAA 2024*, 2024, doi: 10.1109/ICISAA62385.2024.10828595.

[16] J. Sharma, “EfficientNetB3 for High-Performance Insect Identification,” *2nd International Conference on Intelligent Cyber Physical Systems and Internet of Things, ICoICI 2024 - Proceedings*, pp. 955–959, 2024, doi: 10.1109/ICOICI62503.2024.10696281.

[17] R. Zhang, Y. Zhu, Z. Ge, H. Mu, D. Qi, and H. Ni, “Transfer Learning for Leaf Small Dataset Using Improved ResNet50 Network with Mixed Activation Functions,” *Forests 2022, Vol. 13, Page 2072*, vol. 13, no. 12, p. 2072, Dec. 2022, doi: 10.3390/F13122072.

[18] K. Hu *et al.*, “Rice pest identification based on multi-scale double-branch GAN-ResNet,” *Front Plant Sci*, vol. 14, p. 1167121, 2023.

[19] E. M. Roopa Devi, R. Shanthakumari, R. Raj Devi, A. Vasuyaa, Harini, and Lokesh, “Rice Leaf Disease Diagnosis Using Dense EfficientNet Model,” *Lecture Notes in Networks and Systems*, vol. 1050 LNNS, pp. 200–210, 2024, doi: 10.1007/978-3-031-64847-2_18.

[20] Puttaswamy, B. S., and N. Thillaiarasu. "Fine DenseNet based human personality recognition using english hand writing of non-native speakers." Biomedical Signal Processing and Control 99 (2025): 106910.

[21] J. C. Gomes and D. L. Borges, “Insect Pest Image Recognition: A Few-Shot Machine Learning Approach including Maturity Stages Classification,” *Agronomy 2022, Vol. 12, Page 1733*, vol. 12, no. 8, p. 1733, Jul. 2022, doi: 10.3390/AGRONOMY12081733.

[22] H. Peng *et al.*, “Crop pest image classification based on improved densely connected convolutional network,” *Front Plant Sci*, vol. 14, p. 1133060, Apr. 2023, doi: 10.3389/FPLS.2023.1133060/BIBTEX.

[23] W. Xu, L. Han, C. Zhao, B. Liu, Z. Yang, and W. Yang, “Deep Learning-Based Image Recognition of Agricultural Pests,” *Applied Sciences 2022, Vol. 12, Page 12896*, vol. 12, no. 24, p. 12896, Dec. 2022, doi: 10.3390/APP122412896.

[24] Thillaiarasu, T., Suman Lata Tripathi, and V. Dhinakaran, eds. Artificial Intelligence for Internet of Things: Design Principles, Modernization, and Techniques. CRC Press, 2022.

[25] C. Zhang, J. Su, Y. Ju, K. M. Lam, and Q. Wang, “Efficient Inductive Vision Transformer for Oriented Object Detection in Remote Sensing Imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, 2023, doi: 10.1109/TGRS.2023.3292418.