Indonesian Journal of Electrical Engineering and Informatics (IJEEI)
Vol. 13, No. 2, June 2025, pp. 396~409
ISSN: 2089-3272, DOI: 10.52549/ijeei.v13i2.6419

Advanced Classification of Agricultural Plant Insects Using Deep Learning and Explainability

Hoang-Tu Vo[1], Nhon Nguyen Thien[2] , Kheo Chau Mui[3], Phuc Pham Tien[4], Huan Lam Le[5], Vuong Nguyen Phuc[6], Hieu Nguyen Trung[7], Phuong Lam Tan[8]
[1,2,3,4,5,6,7]Information Technology Department, FPT University, Can Tho 94000, Vietnam
[8]Information Technology Department, Tay Do University, Can Tho 94000, Vietnam

Article Info

Article history:

Received Mar 3, 2025
Revised May 18, 2025
Accepted Jun 1, 2025

Keywords:

Agricultural Insect
Explainable AI
XAI
Grad-CAM
Transfer Learning
Pre-trained model

ABSTRACT

This paper investigates the effectiveness of six pre-trained deep learning models to classify images of agricultural plant insects. We utilized the BAU-Insectv2 dataset, which includes images from nine classes. Aphids, Armyworm, Beetle, Bollworm, Grasshopper, Mites, Mosquito, Sawfly, and Stem borer. The models, namely Xception, MobileNetV2, ResNet50, EfficientNetV2B3, ResNet101, and DenseNet121, are fine-tuned by transfer learning from ImageNet. This approach significantly reduces training time while improving classification accuracy. Our experiments reveal that each model reliably distinguishes between insect species even when faced with varying lighting conditions and diverse viewpoints. To further clarify how these models make predictions, we employ Gradient-weighted Class Activation Mapping (Grad-CAM) to highlight critical regions in the images. The results demonstrate that each model focuses on unique biological features and offers clear explanations for its decisions. The research results contribute to demonstrating the potential of pre-trained deep learning architectures for agricultural monitoring and pest management, paving the way for promising future applications.

Corresponding Author:

Hoang-Tu Vo
Information Technology Department, FPT University,
Can Tho 94000, Vietnam.
Email: tuvh6@fe.edu.vn

1. INTRODUCTION

Agricultural productivity is often threatened by insect infestations, which can lead to significant economic losses and environmental challenges [1-2]. Rapid and accurate identification of these pests is essential for effective crop management and timely intervention. Traditional methods of insect identification, which are based on manual inspection and expert knowledge, are time-consuming and subject to human error.

Moreover, the increasing scale of modern agriculture further exacerbates the challenge of pest monitoring. Large farmlands require extensive surveillance, making manual inspection impractical and inefficient. Additionally, the diversity of insect species, seasonal variations, and environmental factors contribute to the complexity of accurate pest identification. Misidentification or delayed detection can lead to improper pest control measures, resulting in crop damage, reduced yields, and increased reliance on chemical pesticides. Therefore, improving the efficiency and reliability of pest monitoring is a critical aspect of sustainable agricultural practices. As a result, there is a growing need for automated approaches that can assist farmers and agricultural professionals in monitoring and managing pest populations.

In recent years, deep learning has revolutionized the field of image classification, offering state-of-the-art performance in various applications. Pre-trained deep learning models, in particular, have shown great promise in adapting to new tasks through transfer learning. By leveraging knowledge from large-scale datasets like ImageNet, these models can be fine-tuned to perform specific classification tasks with reduced training time and improved accuracy. This paper investigates the effectiveness of six such pre-trained models:

Xception, MobileNetV2, ResNet50, EfficientNetV2B3, ResNet101, and DenseNet121 for the classification of agricultural plant insects.

We employ the BAU-Insectv2 dataset, which consists of images spanning nine insect classes: Aphids, Armyworm, Beetle, Bollworm, Grasshopper, Mites, Mosquito, Sawfly, and Stem borer. This dataset presents various challenges, such as different lighting conditions and viewpoints, which simulate real-world scenarios in agricultural settings. Our approach involves fine-tuning each model using transfer learning from ImageNet, thereby harnessing their pre-learned features to better capture the unique characteristics of the insect species in the dataset.

Understanding the decision-making process of these deep learning models is crucial, particularly in applications where transparency is as important as accuracy. To address this, we incorporate the Gradient-weighted Class Activation Mapping (Grad-CAM) technique. Grad-CAM provides visual explanations by highlighting the critical regions in the images that influence the models' predictions. This interpretability not only builds trust in the automated system, but also aids in understanding the biological features that distinguish different insect species.

The contributions of this paper are presented in three aspects:

- First, we provide a comprehensive evaluation of six pre-trained deep learning models on a challenging agricultural insect classification task.
- Second, we demonstrate the practical advantages of transfer learning in reducing training time while maintaining high accuracy.
- Third, we use Grad-CAM to clearly show which specific biological features each model focuses on when making its predictions.

In general, the research results contribute to demonstrating the potential of pre-trained deep learning architectures for agricultural monitoring and pest management, paving the way for promising future applications. The remainder of this paper is organized as follows. Section 2 reviews related work on agriculture pest monitoring and deep learning. Section 3 describes the BAU-Insectv2 dataset. Section 4 is the methodology used in this study. Section 5 presents our experimental setup and results. Section 6 concludes the paper with suggestions for future research.

2. RELATED WORK

2.1 Transfer learning

In the field of deep learning, collecting fully labeled data for every task often poses significant challenges, especially in specialized applications such as agricultural image recognition. To overcome these limitations, transfer learning has emerged as an effective solution, enabling the transfer of features learned from a large dataset (e.g. ImageNet) to target tasks with limited data [3-9]. Numerous studies have shown that pre-trained models such as ResNet, MobileNet, and Xception can be fine-tuned to considerably improve classification performance on new tasks [10-12].

Transfer learning is also widely applied in many other fields, such as healthcare and autonomous vehicles. In healthcare, pre-trained models, initial trained on large datasets such as ImageNet, have been fine-tuned to perform complex tasks such as medical image analysis [13-17], cancer detection [18-21], and pathology segmentation on X-Ray [22-23], MRI [24-26], or CT images [27-29]. This approach helps reduce the need for extensively labeled data and supports early and accurate diagnoses, thereby improving treatment outcomes. Similarly, in the field of autonomous vehicles, transfer learning improves systems for object recognition, traffic sign detection, and identification of surrounding objects, thus improving the environmental awareness and traffic safety of a vehicle [30-33]. Specifically, in the domain of recognizing agricultural pest insects, these models not only help reduce training time but also enhance accuracy in distinguishing insect species under various conditions [34-37].

These applications demonstrate that transfer learning not only saves training time, but also enables AI systems to generalize better in changing environments, opening up many new avenues for research and practical applications.

2.2 Explaining deep learning models with Grad-CAM

Although deep learning models have achieved significant success in many fields, their complex architecture often leads to difficulties in explaining and understanding the decision-making process. This poses a challenge in the deployment of these models in applications that require high reliability.

To address this issue, recent research has proposed methods like SHAP, LIME, and Grad-CAM to clarify the inner workings of deep learning models [38-43]. These methods contribute to the field of

Explainable Artificial Intelligence (XAI) [44], with the aim of making deep learning models clearer and more transparent.

Related research on Grad-CAM has demonstrated that this method is an effective tool to explain the decisions made by deep learning models. Selvaraju et al. (2017) [45] introduced Grad-CAM as a simple technique that uses the gradients of the final layer to produce visual maps, thereby highlighting the regions in an image that greatly influence the model's decision. This method has been used successfully in many areas. For example, it is applied in medical image analysis and object recognition in self-driving cars, allowing users to better understand how the model makes its decisions.

For example, in healthcare, studies have utilized Grad-CAM to identify abnormal regions in X-ray images, thereby assisting physicians in making early and accurate diagnoses [46-50]. In agriculture, Grad-CAM is used to help researchers understand how deep learning models predict leaf diseases. By creating heatmaps over leaf images, Grad-CAM highlights the regions that have the most influence on the model's decision. This visual explanation enables experts to verify that the model is focusing on the correct symptoms of the disease, which supports a more accurate diagnosis and effective crop management.

Recent studies have applied machine learning and deep learning techniques to insect classification [55-58], often using modern machine learning techniques, traditional CNNs, transfer learning. However, these studies often lack interpretability, making it difficult to understand how the models arrive at their predictions. In contrast, our work evaluates six high-performing pre-trained architectures on the BAU-Insectv2 dataset, which presents significant visual complexity. Additionally, while Grad-CAM has been used in general image classification, its use in insect classification for agricultural purposes remains underexplored. By integrating Grad-CAM with model evaluation, we provide both performance comparison and interpretability insights, an area that has not been thoroughly addressed in prior work.

3. DATASET

In this study, we used the BAU-Insectv2 dataset [51], a comprehensive collection of high-resolution images tailored for deep learning and biomedical image analysis in agricultural settings. The dataset comprises images of nine distinct insect genera: Aphids, Armyworm, Beetle, Bollworm, Grasshopper, Mites, Mosquito, Sawfly, and Stem borer. This dataset offers a robust foundation for developing and evaluating machine learning models aimed at insect detection, classification, and analysis, thereby advancing research in precision agriculture and pest management.

The dataset is divided into three sets: the training set comprises 2092 images (80%), the validation set includes 262 images (10%), and the testing set contains 262 images (10%). Before training and evaluation, each image is resized to 224x224 pixels. Some images of the BAU-Insectv2 dataset can be seen in Figure 1.

Figure 1. Some images of the BAU-Insectv2 dataset.

4. METHODOLOGY

4.1 Proposed Model

Figure 2 presents the overall workflow of our proposed insect classification system, which combines multiple state-of-the-art Convolutional Neural Network (CNN) models with an interpretability mechanism.

First, the insect image dataset is gathered and prepared through a preprocessing phase that may include resizing images to a uniform resolution, normalizing pixel values, and applying data augmentation techniques (e.g., rotations, flips, and random crops) to enhance model robustness against variations in angle, lighting, and background.

Once preprocessing is complete, the dataset is partitioned into three subsets: a training set to fit the model parameters, a validation set for hyperparameter tuning and overfitting control, and a testing set to assess final model performance. In this study, six CNN architectures: Xception [11], MobileNetV2 [52], ResNet50 [12], EfficientNetV2B3 [53], ResNet101 [12], and DenseNet121 [54] are trained on the same dataset. Each model leverages transfer learning from large-scale pre-trained weights (e.g., ImageNet), allowing for faster convergence and improved accuracy.

Xception is an extension of the Inception architecture that replaces standard convolutions with depthwise separable convolutions, leading to improved efficiency and accuracy. MobileNetV2 is a lightweight model optimized for mobile and embedded applications, utilizing inverted residuals and linear bottlenecks to reduce computational cost. ResNet50 and ResNet101 are part of the Residual Network family, which employs skip connections to facilitate the training of very deep networks. EfficientNetV2B3 belongs to the EfficientNetV2 family, which improves both training speed and parameter efficiency through compound model scaling. Finally, DenseNet121 connects each layer to every other layer in a feed-forward fashion, enhancing feature reuse and mitigating the vanishing gradient problem. The models learn hierarchical representations of insect features, capturing both global shapes and fine-grained details critical to distinguishing among insect species.

Following model training, performance metrics such as accuracy, precision, recall, and F1-score are computed to provide a comprehensive evaluation of classification effectiveness. Additionally, the framework incorporates Grad-CAM (Gradient-weighted Class Activation Mapping) to generate heatmaps illustrating the regions in each image that most influence the model's predictions. By highlighting class-discriminative features (e.g., wings, body segments, or distinctive coloration patterns), Grad-CAM offers valuable insight into the model's decision-making process. This interpretability component is essential in agricultural settings, where understanding the basis of a model's classification decisions can guide more effective and trustworthy pest management strategies.

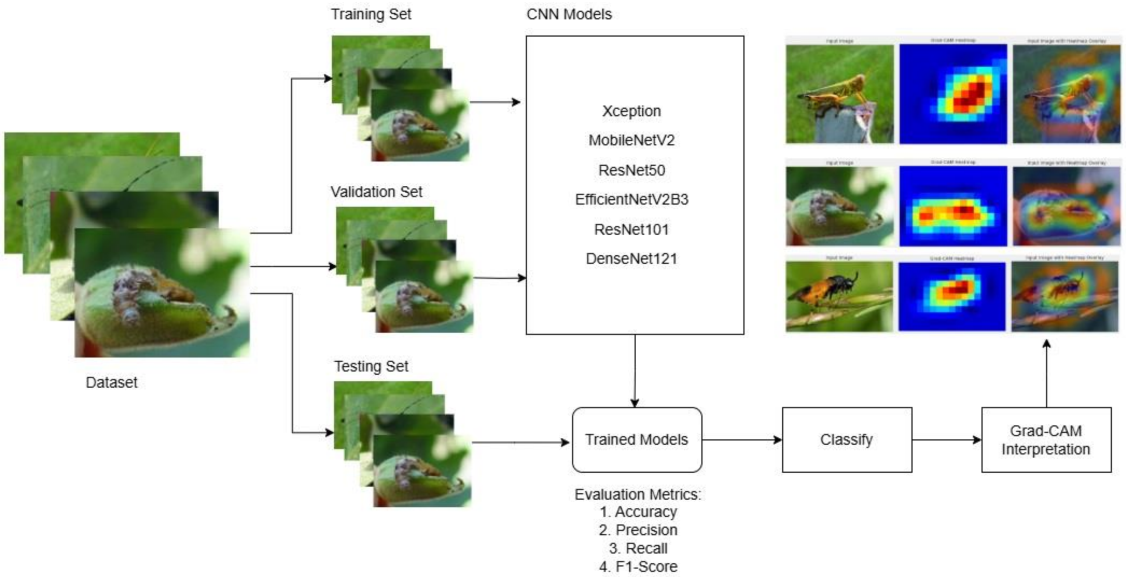

Figure 2. A diagram depicting the suggested deep learning methodology for classification and visualization.

4.2 Model Architecture

Table 1. Detailed Layer Structure of the Proposed Model

Layer	Description
Input (224×224×3)	Accepts RGB images of size 224×224×3 as the input.
Pre-trained Model	Pre-trained architectures such as Xception, MobileNetV2, ResNet50, EfficientNetV2B3, ResNet101, and DenseNet121 are leveraged to extract features.
BatchNormalization	Normalizes activations from the base model, accelerating convergence and improving training stability.

Layer	Description
Dense (256, ReLU)	A fully connected layer with 256 neurons, employing ReLU activation to capture high-level feature representations.
Dropout (0.35)	Randomly drops 35% of the neurons during training to reduce overfitting and improve generalization.
BatchNormalization	Further normalizes intermediate activations, aiding in stable and efficient training
Dense (128, ReLU)	A fully connected layer with 128 neurons, again using ReLU activation to refine learned features.
Dropout (0.3)	Drops 30% of the neurons, providing an additional regularization mechanism.
Dense (64, ReLU)	A fully connected layer with 64 neurons, using ReLU activation for further feature processing.
Dense (class_count, Softmax)	The output layer, containing class_count neurons with a Softmax activation function, producing a probability distribution over the target classes

Table 1 presents the detailed architecture of the proposed model. The model uses an RGB input image with a size of 224×224×3. Features are extracted using pre-trained models such as Xception, MobileNetV2, or ResNet50. BatchNormalization layers are added to stabilize the training, followed by Dense layers with 256 and 128 neurons, combined with Dropout to reduce overfitting. A Dense layer with 64 neurons further refines the features. Finally, the output Dense layer, using the Softmax function, generates a probability distribution for each classification class.

4.3 Setup configuration

In the proposed model training configuration, the input size is set to 224x224 pixels, consistent with the standard practices for this architecture. We utilized the Adam optimizer with a learning rate of 0.0001 to update the model weights. The training process is scheduled for 50 epochs, providing a balance that allows the model ample opportunity to learn from the data while reducing the risk of overfitting. The batch size is set to 16, a common choice that effectively balances computational efficiency and model performance.

5. RESULTS

In this section, we present and analyze the experimental results derived from six pre-trained deep learning models: Xception, MobileNetV2, ResNet50, EfficientNetV2B3, ResNet101, and DenseNet121. These models were evaluated using the BAU-Insectv2 dataset, which consists of images across nine insect classes: Aphids, Armyworm, Beetle, Bollworm, Grasshopper, Mites, Mosquito, Sawfly, and Stem borer. The performance of these models was measured through key evaluation metrics, including accuracy, precision, recall, F1 score, and support on testing phases.

5.1 Xception

The classification report for the Xception model, presented in Table 2, summarizes its performance on 262 samples across nine insect types. The overall accuracy is 0.9847, with macro-average metrics of 0.9860 for precision, 0.9839 for recall, and 0.9848 for F1-score, closely matched by the weighted averages (0.9850, 0.9847, and 0.9847, respectively). Notably, the Mites, Mosquito, and Stem Borer classes achieved perfect scores across all metrics, while the Aphids, Armyworm, Beetle, Bollworm, and Grasshopper classes also demonstrated strong performance with F1-scores ranging from 0.9508 to 0.9855. These results highlight the robustness and effectiveness of the Xception model in accurately classifying various insect types.

Table 2. Classification Report of Xception model

Insect Type	Precision	Recall	F1-Score	Support
Aphids	0.9667	0.9355	0.9508	31
Armyworm	1.0000	0.9615	0.9804	26
Beetle	0.9714	1.0000	0.9855	34
Bollworm	0.9655	1.0000	0.9825	28
Grasshopper	0.9706	1.0000	0.9851	33
Mites	1.0000	1.0000	1.0000	30
Mosquito	1.0000	1.0000	1.0000	34

Insect Type	Precision	Recall	F1-Score	Support
Sawfly	1.0000	0.9583	0.9787	24
Stem Borer	1.0000	1.0000	1.0000	22
Accuracy	-	-	0.9847	262
Macro Avg	0.9860	0.9839	0.9848	262
Weighted Avg	0.9850	0.9847	0.9847	262

5.2 MobileNetV2

The classification report for the MobileNetV2 model, as shown in Table 3, summarizes its performance on a dataset of 262 samples across nine insect types. The overall accuracy achieved is 0.9885. The macro-average metrics indicate a precision of 0.9887, a recall of 0.9875, and an F1-score of 0.9880, while the weighted averages are similarly high (0.9888 for precision, 0.9885 for recall, and 0.9885 for F1-score). Notably, several insect categories, such as Beetle, Grasshopper, Mites, Mosquito, and Stem Borer, attained perfect scores across all metrics. The strong performance across the board demonstrates that MobileNetV2 is highly effective and robust in accurately classifying various insect types.

Table 3. Classification Report of MobileNetV2 model.

Insect Type	Precision	Recall	F1-Score	Support
Aphids	1.0000	0.9677	0.9836	31
Armyworm	0.9615	0.9615	0.9615	26
Beetle	0.9714	1.0000	0.9855	34
Bollworm	0.9655	1.0000	0.9825	28
Grasshopper	1.0000	1.0000	1.0000	33
Mites	1.0000	1.0000	1.0000	30
Mosquito	1.0000	1.0000	1.0000	34
Sawfly	1.0000	0.9583	0.9787	24
Stem Borer	1.0000	1.0000	1.0000	22
Accuracy	-	-	0.9885	262
Macro Avg	0.9887	0.9875	0.9880	262
Weighted Avg	0.9888	0.9885	0.9885	262

5.3 ResNet50

The classification report for the ResNet50 model, shown in Table 4, outlines its performance on a dataset of 262 samples across nine insect types. The overall accuracy of the model is 0.9924. The macro-average metrics are exceptionally high, with a precision of 0.9923, recall of 0.9918, and F1-score of 0.9919, which are very close to the weighted averages. Notably, several insect categories, including Beetle, Bollworm, Grasshopper, and Stem Borer, achieved perfect scores (1.0000) on all metrics. These results demonstrate the robustness and precision of the ResNet50 model in accurately classifying various insect types.

Table 4. Classification Report of ResNet50 model.

Insect Type	Precision	Recall	F1-Score	Support
Aphids	1.0000	0.9677	0.9836	31
Armyworm	0.9630	1.0000	0.9811	26
Beetle	1.0000	1.0000	1.0000	34
Bollworm	1.0000	1.0000	1.0000	28
Grasshopper	1.0000	1.0000	1.0000	33
Mites	0.9677	1.0000	0.9836	30
Mosquito	1.0000	1.0000	1.0000	34

Insect Type	Precision	Recall	F1-Score	Support
Sawfly	1.0000	0.9583	0.9787	24
Stem Borer	1.0000	1.0000	1.0000	22
Accuracy	-	-	0.9924	262
Macro Avg	0.9923	0.9918	0.9919	262
Weighted Avg	0.9926	0.9924	0.9924	262

5.4 EfficientNetV2B3

The classification report for the EfficientNetV2B3 model, presented in Table 5, summarizes its performance on a dataset of 262 samples across nine insect types. The model demonstrates exceptional performance, achieving perfect scores (1.0000) for Aphids, Beetle, Grasshopper, Mites, Mosquito, Sawfly, and Stem Borer. Armyworm records an F1-Score of 0.9804 and Bollworm an F1-Score of 0.9825. Overall, the model attains an accuracy of 0.9962, with macro-average precision, recall, and F1-scores of 0.9962, 0.9957, and 0.9959 respectively, and weighted averages that are similarly outstanding. These results underscore the high reliability and effectiveness of EfficientNetV2B3 in classifying various insect types.

Table 5. Classification Report of EfficientNetV2B3 model.

Insect Type	Precision	Recall	F1-Score	Support
Aphids	1.0000	1.0000	1.0000	31
Armyworm	1.0000	0.9615	0.9804	26
Beetle	1.0000	1.0000	1.0000	34
Bollworm	0.9655	1.0000	0.9825	28
Grasshopper	1.0000	1.0000	1.0000	33
Mites	1.0000	1.0000	1.0000	30
Mosquito	1.0000	1.0000	1.0000	34
Sawfly	1.0000	1.0000	1.0000	24
Stem Borer	1.0000	1.0000	1.0000	22
Accuracy	-	-	0.9962	262
Macro Avg	0.9962	0.9957	0.9959	262
Weighted Avg	0.9963	0.9962	0.9962	262

5.5 ResNet101

Table 6 presents the classification report for the ResNet101 model, evaluated on a dataset of 262 samples spanning nine insect types. The model achieved an overall accuracy of 0.9885, with macro-average metrics of 0.9887 for precision, 0.9875 for recall, and 0.9880 for F1-score, while the weighted averages are similarly high. Notably, most insect classes, including Aphids, Beetle, Bollworm, Grasshopper, Mites, Mosquito, Sawfly, and Stem Borer, demonstrate robust performance, with several categories achieving near-perfect scores. These results underscore the effectiveness and reliability of the ResNet101 model in accurately classifying diverse insect types.

Table 6. Classification Report of ResNet101 model.

Insect Type	Precision	Recall	F1-Score	Support
Aphids	1.0000	0.9677	0.9836	31
Armyworm	0.9615	0.9615	0.9615	26
Beetle	0.9714	1.0000	0.9855	34
Bollworm	0.9655	1.0000	0.9825	28
Grasshopper	1.0000	1.0000	1.0000	33
Mites	1.0000	1.0000	1.0000	30
Mosquito	1.0000	1.0000	1.0000	34

Insect Type	Precision	Recall	F1-Score	Support
Sawfly	1.0000	0.9583	0.9787	24
Stem Borer	1.0000	1.0000	1.0000	22
Accuracy	-	-	0.9885	262
Macro Avg	0.9887	0.9875	0.9880	262
Weighted Avg	0.9888	0.9885	0.9885	262

5.6 DenseNet121

The classification report summarizes the performance of the DenseNet121 model on a dataset of 262 samples across nine insect classes (Table 7). The overall accuracy is 0.9809, with both macro and weighted averages for precision, recall, and F1-score around 0.981. Notably, the Beetle, Bollworm, Mosquito, and Stem Borer classes achieved perfect scores (1.0000) on all metrics, while the remaining classes also performed strongly with F1-scores above 0.943. These results highlight the high effectiveness and stability of DenseNet121 in accurately classifying various insect types.

Table 7. Classification Report of ResNet101 model.

Insect Type	Precision	Recall	F1-Score	Support
Aphids	0.9375	0.9677	0.9524	31
Armyworm	0.9259	0.9615	0.9434	26
Beetle	1.0000	1.0000	1.0000	34
Bollworm	1.0000	1.0000	1.0000	28
Grasshopper	1.0000	0.9394	0.9688	33
Mites	0.9677	1.0000	0.9836	30
Mosquito	1.0000	1.0000	1.0000	34
Sawfly	1.0000	0.9583	0.9787	24
Stem Borer	1.0000	1.0000	1.0000	22
Accuracy	-	-	0.9809	262
Macro Avg	0.9812	0.9808	0.9808	262
Weighted Avg	0.9816	0.9809	0.9810	262

5.7 Evaluating the Models' Overall Performance

Table 8 summarizes the testing accuracy of six deep learning models on the insect classification task. EfficientNetV2B3 achieved the highest accuracy (99.62%), followed by ResNet50 (99.24%). Xception, MobileNetV2, ResNet101, and DenseNet121 also demonstrated strong performance, with accuracies ranging from 98.09% to 98.85%. These results underscore the robust performance of these models, particularly highlighting the superior effectiveness of EfficientNetV2B3 in classifying images of agricultural plant insects.

Table 8. Testing Accuracy of Various Models.

Model	Testing Accuracy
Xception	0.9847
MobileNetV2	0.9885
ResNet50	0.9924
EfficientNetV2B3	0.9962
ResNet101	0.9885
DenseNet121	0.9809

Figure 3 illustrates the confusion matrices for six pre-trained models used in classifying images of agricultural plant insects, offering a detailed view of each model's performance in correctly identifying and occasionally misclassifying the various insect categories.

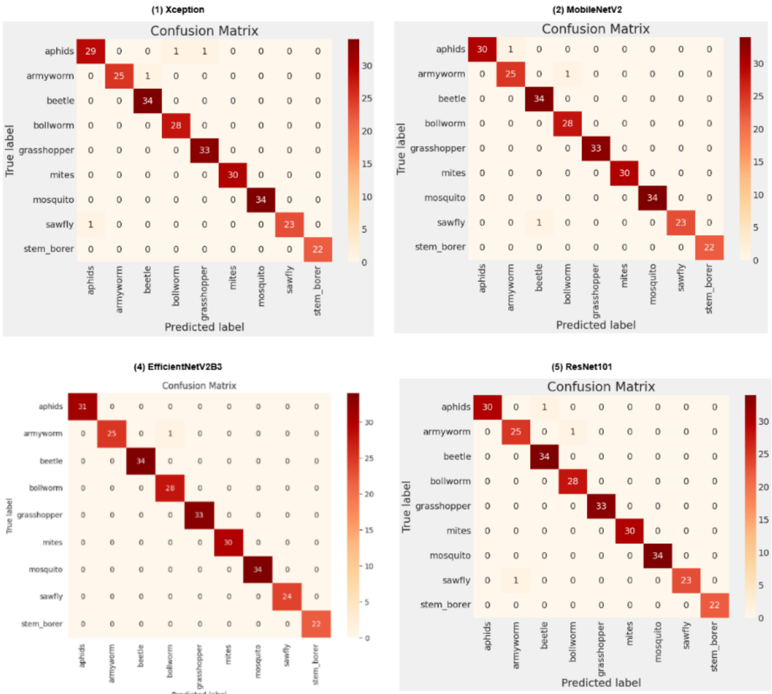

Figure. 3. Confusion matrices for six pre-trained models used to classify images of agricultural plant insects.

5.8 Grad-CAM is used to evaluate and explain the results predicted by the model

In our study, Grad-CAM is used to assess and interpret the predictions generated by our deep learning models on images of agricultural plant insects. Grad-CAM (Gradient-weighted Class Activation Mapping) helps to visualize the discriminative regions in an input image that have the most influence on the model's decision. It operates by computing the gradient of the target class score with respect to the feature maps in the final convolutional layer. These gradients are then globally averaged to obtain importance weights, which are used to produce a class-specific localization map. This map highlights the salient regions the model focuses on during classification.

By creating class-specific heatmaps, Grad-CAM highlights the key regions within each image that contribute most significantly to the model's decision-making process. This visualization technique not only aids in understanding the internal mechanisms of the deep learning models but also verifies that the model is focusing on biologically relevant features.

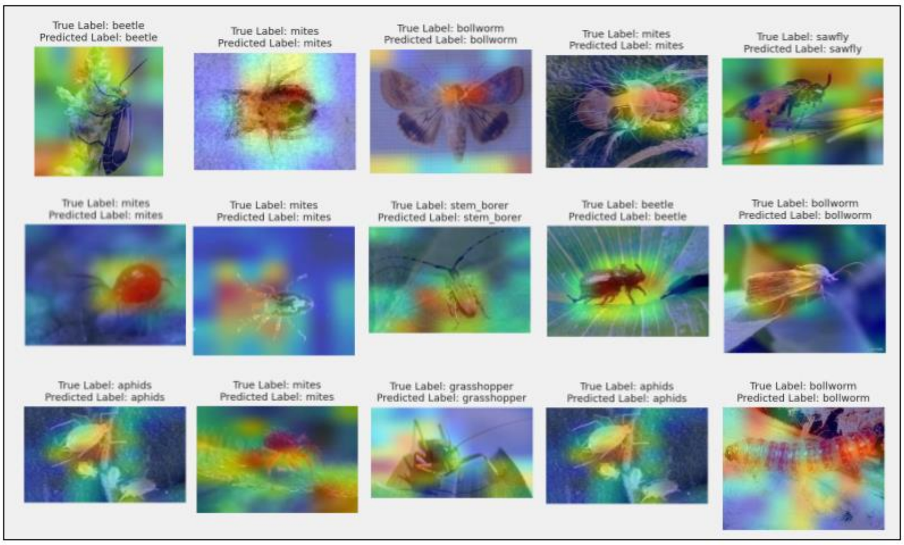

Figure 4. Heatmap-based Insect Classification Results

Figure 4 shows Grad-CAM visualizations for multiple insect images, each annotated with both the true label and the model's predicted label. The heatmaps highlight the specific regions within each image that the deep learning model relies on most heavily for classification. Notably, when the predicted label matches the true label, the heatmap often focuses on distinctive morphological features (e.g., body shape, wing patterns) that are characteristic of that insect class.

Figure 5 presents Grad-CAM visualizations generated by six deep learning models: Xception, MobileNetV2, ResNet50, EfficientNetV2B3, ResNet101, and DenseNet121 applied to an agricultural plant insect image. The first column shows the original input image, the second column displays the corresponding Grad-CAM heatmap, and the third column overlays the heatmap on the input image. Notably, each model's heatmap emphasizes the insect's body, particularly around the head, indicating that these morphological features are most influential in the classification decision.

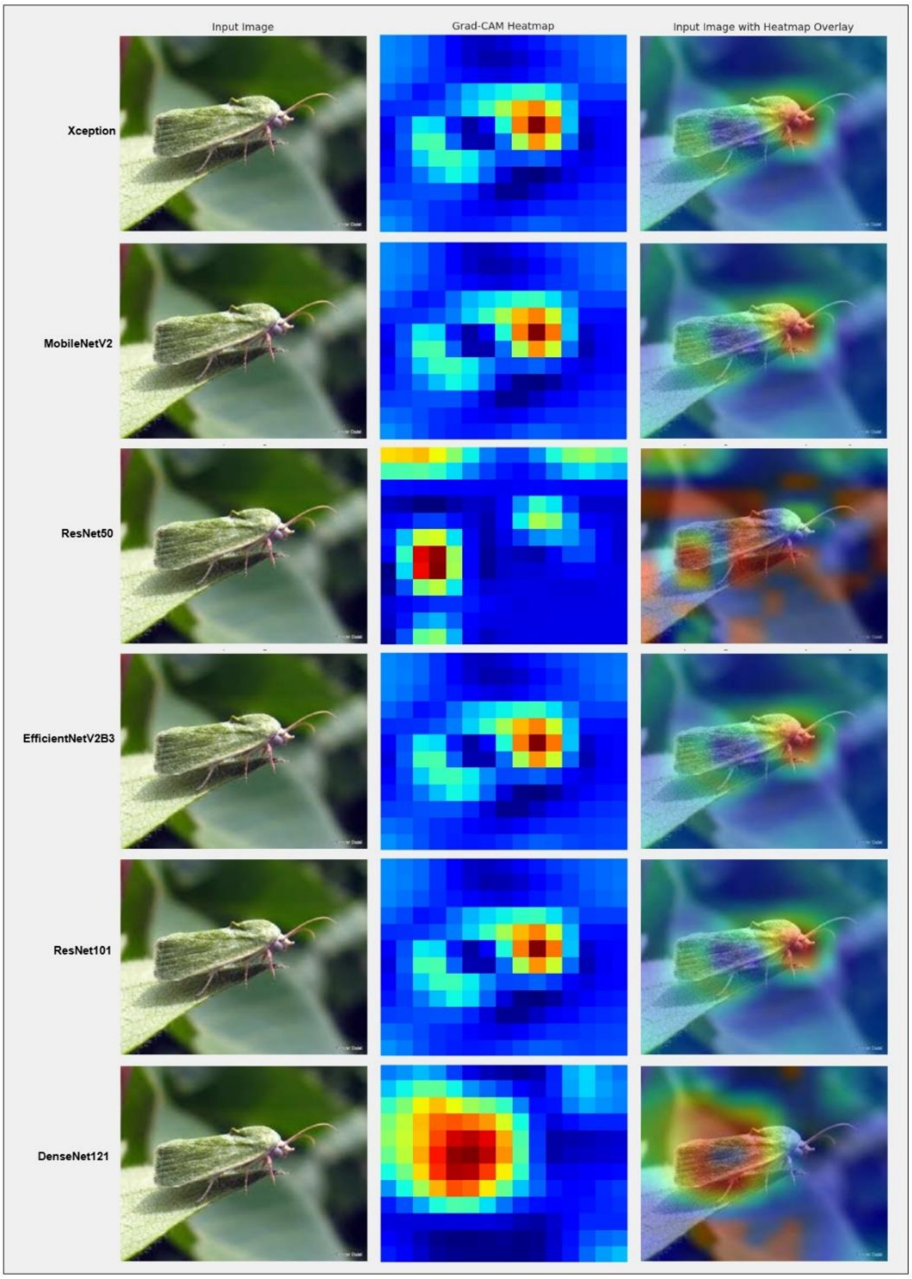

Figure 5. Grad-CAM Visualization Across Six Deep Learning Models for Agricultural Insect Classification

5. CONCLUSION AND FUTURE WORK

In this study, we evaluated the performance of six pre-trained deep learning architectures: Xception, MobileNetV2, ResNet50, EfficientNetV2B3, ResNet101, and DenseNet121 on the BAU-Insectv2 dataset comprising nine classes of agricultural plant insects. Using transfer learning, we significantly reduced training time while maintaining high classification accuracy across diverse lighting conditions and viewpoints. The Grad-CAM technique provided valuable insight into the model decision-making processes, illustrating that each network focuses on distinct biological features when identifying insect species. These findings highlight the potential of pre-trained deep learning models not only for accurate pest identification but also for improving transparency in model predictions.

While Grad-CAM has provided valuable visual explanations for the predictions of our model, future work should explore additional interpretability techniques to further elucidate the underlying decision-making process. Methods such as layer-wise relevance propagation, SHAP values, or integrated gradients could provide more detail of how different features contribute to the final classification.

REFERENCES

[1] S. Savary, A. Ficke, J.-N. Aubertot, and C. Hollier, "Crop losses due to diseases and their implications for global food production losses and food security," *Food security,* vol. 4, no. 4, pp.519–537, 2012

[2] C. A. Deutsch, J. J. Tewksbury, M. Tigchelaar, D. S. Battisti, S. C. Merrill, R. B. Huey, and R. L. Naylor, "Increase in crop losses to insect pests in a warming climate," *Science*, vol. 361, no. 6405, pp. 916–919, 2018.

[3] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[4] R. Ribani and M. Marengoni, "A survey of transfer learning for convolutional neural networks," *in 2019 32nd SIBGRAPI conference on graphics, patterns and images tutorials (SIBGRAPI-T)*. IEEE, 2019, pp. 47–57.

[5] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," *in Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27*. Springer, 2018, pp. 270–279.

[6] H. Liang, W. Fu, and F. Yi, "A survey of recent advances in transfer learning," *in 2019 IEEE 19th international conference on communication technology (ICCT)*. IEEE, 2019, pp. 1516–1523.

[7] A. W. Salehi, S. Khan, G. Gupta, B. I. Alabduallah, A. Almjally, H. Alsolai, T. Siddiqui, and A. Mellit, "A study of cnn and transfer learning in medical imaging: Advantages, challenges, future scope," *Sustainability*, vol. 15, no. 7, p. 5930, 2023.

[8] S. Niu, Y. Liu, J. Wang, and H. Song, "A decade survey of transfer learning (2010–2020)," *IEEE Transactions on Artificial Intelligence*, vol. 1, no. 2, pp. 151–166, 2020..

[9] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" *Advances in neural information processing systems*, vol. 27, 2014.

[10] A. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications. sl, sn," *arXiv preprint arXiv:1704.04861*, 2017.

[11] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *in Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *in Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[13] F. Ritter, T. Boskamp, A. Homeyer, H. Laue, M. Schwier, F. Link, and H.-O. Peitgen, "Medical image analysis," *IEEE pulse*, vol. 2, no. 6, pp. 60–70, 2011.

[14] A. P. Dhawan, *Medical image analysis*. John Wiley & Sons, 2011.

[15] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual review of biomedical engineering*, vol. 19, no. 1, pp. 221– 248, 2017.

[16] H.-P. Chan, R. K. Samala, L. M. Hadjiiski, and C. Zhou, "Deep learning in medical image analysis," *Deep learning in medical image analysis: challenges and applications*, pp. 3–21, 2020.

[17] H.-T. Vo, N. N. Thien, and K. C. Mui, "Tomato disease recognition: Advancing accuracy through xception and bilinear pooling fusion," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 8, 2023.

[18] K. Munir, H. Elahi, A. Ayub, F. Frezza, and A. Rizzi, "Cancer diagnosis using deep learning: a bibliographic review," *Cancers*, vol. 11, no. 9, p. 1235, 2019.

[19] S. Khan, N. Islam, Z. Jan, I. U. Din, and J. J. C. Rodrigues, "A novel deep learning based framework for the detection and classification of breast cancer using transfer learning," *Pattern Recognition Letters*, vol. 125, pp. 1–6, 2019.

[20] S. Guan and M. Loew, "Breast cancer detection using transfer learning in convolutional neural networks," *in 2017 IEEE applied imagery pattern recognition workshop (AIPR)*. IEEE, 2017, pp.1–8.

[21] R. Fakoor, F. Ladhak, A. Nazi, and M. Huber, "Using deep learning to enhance cancer diagnosis and classification," *in Proceedings of the international conference on machine learning*, vol. 28. ACM New York, NY, USA, 2013, pp. 3937–3949.

[22] P. L. Vidal, J. de Moura, J. Novo, and M. Ortega, "Multi-stage transfer learning for lung segmentation using portable x-ray devices for patients with covid-19," *Expert Systems with Applications*, vol. 173, p. 114677, 2021.

[23] A. Z. Abidin, B. Deng, A. M. DSouza, M. B. Nagarajan, P. Coan, and A. Wismuller, "Deep transfer learning for characterizing chondrocyte patterns in phase contrast x-ray computed tomography images of the human patellar cartilage," *Computers in biology and medicine*, vol. 95, pp. 24–33, 2018.

[24] S. U. H. Dar, M. Ozbey, A. B. ¨ C¸ atlı, and T. C¸ ukur, "A transfer-learning approach for accelerated mri using deep neural networks," *Magnetic resonance in medicine*, vol. 84, no. 2, pp. 663–685, 2020.

[25] A. Ebrahimi-Ghahnavieh, S. Luo, and R. Chiong, "Transfer learning for alzheimer's disease detection on mri images," *in 2019 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*. IEEE, 2019, pp. 133–138

[26] C. Srinivas, N. P. KS, M. Zakariah, Y. A. Alothaibi, K. Shaukat, B. Partibane, and H. Awal, "Deep transfer learning approaches in performance analysis of brain tumor classification using mri mages," *Journal of Healthcare Engineering*, vol. 2022, no. 1, p. 3264367, 2022

[27] V. Perumal, V. Narayanan, and S. J. S. Rajasekar, "Detection of covid-19 using cxr and ct images using transfer learning and haralick features," *Applied Intelligence*, vol. 51, no. 1, pp. 341– 58, 2021.

[28] A. Halder and B. Datta, "Covid-19 detection from lung ct-scan images using transfer learning approach," *Machine Learning: Science and Technology*, vol. 2, no. 4, p. 045013, 2021.

[29] T. K. Sajja, R. M. Devarapalli, and H. K. Kalluri, "Lung cancer detection based on ct scan images by using deep transfer learning," *Traitement du Signal*, vol. 36, no. 4, pp. 339–344, 2019.

[30] A. Mounsey, A. Khan, and S. Sharma, "Deep and transfer learning approaches for pedestrian identification and classification in autonomous vehicles," *Electronics*, vol. 10, no. 24, p. 3159, 2021

[31] S. Akhauri, L. Y. Zheng, and M. C. Lin, "Enhanced transfer learning for autonomous driving with systematic accident simulation," *in 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 5986–5993.

[32] S. Sharma, J. E. Ball, B. Tang, D. W. Carruth, M. Doude, and M. A. Islam, "Semantic segmentation with transfer learning for off-road autonomous driving," *Sensors*, vol. 19, no. 11, p. 2577, 2019.

[33] X. Liu, J. Li, J. Ma, H. Sun, Z. Xu, T. Zhang, and H. Yu, "Deep transfer learning for intelligent vehicle perception: A survey," *Green Energy and Intelligent Transportation*, p. 100125, 2023.

[34] K. Thenmozhi and U. S. Reddy, "Crop pest classification based on deep convolutional neural network and transfer learning," *Computers and Electronics in Agriculture*, vol. 164, p. 104906, 2019.

[35] W. Dawei, D. Limiao, N. Jiangong, G. Jiyue, Z. Hongfei, and H. Zhongzhi, "Recognition pest by image-based transfer learning," *Journal of the Science of Food and Agriculture*, vol. 99, no. 10, pp. 4524–4531, 2019.

[36] N. E. M. Khalifa, M. Loey, and M. H. N. Taha, "Insect pests recognition based on deep transfer learning models," *J. Theor. Appl. Inf. Technol*, vol. 98, no. 1, pp. 60–68, 2020.

[37] G. Pattnaik, V. K. Shrivastava, and K. Parvathi, "Transfer learning-based framework for classification of pest in tomato plants," *Applied Artificial Intelligence*, vol. 34, no. 13, pp. 981–993, 2020.

[38] S. Ali, T. Abuhmed, S. El-Sappagh, K. Muhammad, J. M. Alonso-Moral, R. Confalonieri, R. Guidotti, J. Del Ser, N. D´ıaz-Rodr´ıguez, and F. Herrera, "Explainable artificial intelligence (xai): What we know and what is left to attain trustworthy artificial intelligence," *Information fusion*, vol. 99, p. 101805, 2023.

[39] G. Feretzakis, A. Sakagianni, A. Anastasiou, I. Kapogianni, E. Bazakidou, P. Koufopoulos, Y. Koumpouros, C. Koufopoulou, V. Kaldis, and V. S. Verykios, "Integrating shapley values into machine learning techniques for enhanced predictions of hospital admissions," *Applied Sciences*, vol. 14, no. 13, p. 5925, 2024.

[40] J. Sun, C.-K. Sun, Y.-X. Tang, T.-C. Liu, and C.-J. Lu, "Application of shap for explainable machine learning on age-based subgrouping mammography questionnaire data for positive mammography prediction and risk factor identification," *in Healthcare*, vol. 11, no. 14. MDPI, 2023, p. 2000.

[41] E. Stenwig, G. Salvi, P. S. Rossi, and N. K. Skjærvold, "Comparative analysis of explainable machine learning prediction models for hospital mortality," *BMC Medical Research Methodology*, vol. 22, no. 1, p. 53, 2022.

[42] S. Rao, S. Mehta, S. Kulkarni, H. Dalvi, N. Katre, and M. Narvekar, "A study of lime and shap model explainers for autonomous disease predictions," *in 2022 IEEE Bombay Section Signature Conference (IBSSC)*. IEEE, 2022, pp. 1–6.

[43] L.-D. Quach, K. N. Quoc, A. N. Quynh, N. ThaiNghe, and T. G. Nguyen, "Explainable deep learning models with gradient-weighted class activation mapping for smart agriculture," *IEEE Access*, vol. 11, pp. 83 752–83 762, 2023.

[44] A. B. Arrieta, N. D´ıaz-Rodr´ıguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garc´ıa, S. Gil-Lopez, D. Molina, R. Benjamins ´ et al., "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," *Information fusion*, vol. 58, pp. 82–115, 2020.

[45] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradientbased localization," *in Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[46] M. Xiao, L. Zhang, W. Shi, J. Liu, W. He, and Z. Jiang, "A visualization method based on the grad-cam for medical image segmentation model," *in 2021 International Conference on Electronic Information Engineering and Computer Science (EIECS)*. IEEE, 2021, pp. 242–247.

[47] K. Lamba and S. Rani, "A novel approach of brain-computer interfacing (bci) and grad-cam based explainable artificial intelligence: Use case scenario for smart healthcare," *Journal of Neuroscience Methods*, vol. 408, p. 110159, 2024.

[48] H. Zhang and K. Ogasawara, "Grad-cam-based explainable artificial intelligence related to medical text processing," *Bioengineering*, vol. 10, no. 9, p. 1070, 2023.

[49] A. Hamza, M. Attique Khan, S.-H. Wang, M. Alhaisoni, M. Alharbi, H. S. Hussein, H. Alshazly, Y. J. Kim, and J. Cha, "Covid-19 classification using chest x-ray images based on fusion-assisted deep bayesian optimization and grad-cam visualization," *Frontiers in Public Health*, vol. 10, p. 1046296, 2022.

[50] G. M. Idroes, T. R. Noviandy, T. B. Emran, and R. Idroes, "Explainable deep learning approach for mpox skin lesion detection with grad-cam," *Heca Journal of Applied Sciences*, vol. 2, no. 2, pp. 54–63, 2024.

[51] I. Salehin, M. R. Khan, U. Habiba, N. H. Badhon, and N. N. Moon, "Bau-insectv2: An agricultural plant insect dataset for deep learning and biomedical image analysis," *Data in Brief*, vol. 53, p. 110083, 2024.

[52] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," *in Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[53] M. Tan and Q. Le, "Efficientnetv2: Smaller models and faster training," *in International conference on machine learning*. PMLR, 2021, pp. 10 096–10 106.

[54] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *in Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[55] 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.

[56] A. C. Teixeira, J. Ribeiro, R. Morais, J. J. Sousa, and A. Cunha, "A systematic review on automatic insect detection using deep learning," *Agriculture*, vol. 13, no. 3, p. 713, 2023.

[57] D. Ozdemir and M. S. Kunduraci, "Comparison of deep learning techniques for classification of the insects in order level with mobile software application," *IEEE Access*, vol. 10, pp. 35 675 35 684, 2022.

[58] X. Cao, Z. Wei, Y. Gao, and Y. Huo, "Recognition of common insect in field based on deep learning," in *Journal of Physics: Conference Series*, vol. 1634, no. 1. IOP Publishing, 2020, p.012034

BIOGRAPHY OF AUTHORS

HoangTu-Vo holds a Bachelor of Information Systems from Can Tho University, Vietnam, in 2011. In 2013, He graduated with a master's degree in Information Systems from Can Tho University in Vietnam. Currently, He is working as a lecturer in Information Technology Department at FPT University, Can Tho campus in Vietnam. His research interests include Machine learning, Deep learning, Image processing, and Computer vision. He can be contacted at email: tuvh6@fe.edu.vn

Nhon Nguyen Thien holds a Bachelor of Information Systems from Can Tho University, Vietnam, in 2013. In 2017, He graduated with a master's degree in Information Systems from Can Tho University in Vietnam. Currently, He is working as a lecturer in Information Technology Department at FPT University, Can Tho campus in Vietnam. His research areas of interest include Data science, Machine learning, Deep learning, and Web application development. He can be contacted at email: nhonnt9@fe.edu.vn

Kheo Chau Mui holds a Bachelor of Information Systems from Can Tho University, Vietnam, in 2010. In 2020, She graduated with a master's degree in Computer Science from Can Tho University in Vietnam. Currently, She is working as a lecturer in Information Technology Department at FPT University, Can Tho campus in Vietnam. Her research areas of interest include Image processing, Image classification, Data science, Object detection, Deep learning, and Machine learning. She can be contacted at email: kheocm@fe.edu.vn

Phuc Pham Tien earned his Bachelor's degree in Information Systems from Can Tho University, Vietnam, in 2002. In 2018, he completed a Master's degree in Computer Science at Ho Chi Minh City University of Technology, Vietnam National University, Ho Chi Minh City, Vietnam. Currently, he is a lecturer in Information Technology Department at FPT University, Can Tho campus, Vietnam. His research interests include machine learning and deep learning. He can be reached at phucpt10@fe.edu.vn

Huan Lam Le received his Bachelor's degree in Information Systems from Can Tho University, Vietnam, in 1996, and his Master's degree in the Information Technology from the same institution in 2010. He is currently a lecturer in the Department of Information Technology at FPT University, Can Tho campus, Vietnam. His research interests include machine learning and deep learning. He can be contacted by email at: huanll@fe.edu.vn

Vuong Nguyen Phuc holds a Bachelor of Information Systems from Hong Bang International University, Vietnam, in 2017. In 2023, He graduated with a master's degree in Computer Science from Can Tho University in Vietnam. Currently, He is working as a lecturer in Information Technology Department at FPT University, Can Tho campus in Vietnam. His research areas of interest include Data science, Machine learning, Deep learning, and Web application development. He can be contacted at email: vuongnp@fe.edu.vn

Hieu Nguyen Trung received his Bachelor's degree in Computer Science from Can Tho University, Vietnam, in 2015. In 2020, he completed his Master's degree in Computer Science, also from Can Tho University. He is currently serving as a lecturer in the Information Technology Department at FPT University, Can Tho campus, Vietnam. His research interests include Data Science, Machine Learning, Deep Learning, and Web Application Development. He can be contacted via email: hieunt239@fe.edu.vn

Phuong Lam Tan holds a Bachelor of Information Systems from Can Tho University, Vietnam, in 2008, In 2016, He graduated with a master's degree in Information Systems from Can Tho University in Vietnam. Since 2009, he has been a lecturer teaching Information Technology at Tay Do University in Can Tho city, Vietnam. His research areas include artificial intelligence, machine learning, web application development, and mobile application development. He can be contacted at email: ltphuong@tdu.edu.vn