J. Intell. Comput. Health Inform. Vol. 5, No. 1, 2024
ISSN: 2715-6923 e-ISSN: 2721-9186, pp.25~34 | DOI: 10.26714/jichi.v5i1.13959

Enhancing Agricultural Pest Detection with EfficientNetV2-L and Grad-CAM: A Comprehensive Approach to Sustainable Farming

Denaya Ferrari Noval Agatra*, Barisma Ami Cornella, Muhammad Muza'in, Muhammad Munsarif, Ahmad Ilham

Department of Informatics, Universitas Muhammadiyah Semarang, Jl. Kedungmundu Raya No. 18, Semarang 50273, Indonesia
*Corresponding author: bayezidahmad808@gmail.com

Jafar Abdollahi

Department of Computer Engineering, Ardabil Branch, Islamic Azad University, Ardabil, Iran

Article histroy :
Received: 17 JAN 2024
Accepted: 04 FEB 2024
Available online: 30 MAR 2024

Research article

Abstract: In modern agriculture, quickly identifying agricultural pests is essential for maintaining high crop yields and ensuring global food security. In diverse and dynamic agricultural environments, traditional pest detection methods exhibit reduced accuracy, limited scalability, and lack interpretability. In this study, EfficientNetV2-L and Grad-CAM were used to significantly enhance pest detection system performance and transparency. EfficientNetV2-L, a fast and resource-efficient model, excels particularly in computationally constrained environments. Traditional CNN models, including EfficientNetV2-L, are criticized as uninterpretable "black boxes" despite their high accuracy. To address this issue, Grad-CAM was used to generate salient maps that visually show the most influential areas of the input image in the model's decision-making process. This combination not only provides superior pest detection accuracy but also provides actionable insights into the model's predictions, which is an important feature for building trust among agricultural practitioners. Our experimental results show a 15% improvement in detection accuracy compared to conventional models, especially in identifying visually similar-looking pest species that are often misclassified. In addition, the enhanced interpretability provided by Grad-CAM has led to a deeper understanding of the model's behaviour, enabling iterative adjustments and improvements that further enhance the reliability of the system. The practical implications of these findings are significant: this integrated model offers a robust solution that can be seamlessly applied to real-time agricultural monitoring systems. With the early detection and proper classification of pests, this model can be used as a more effective pest management strategy to minimize crop damage and increase agricultural productivity. This research not only advances the technological frontier of pest detection but also contributes to broader goals related to sustainable agriculture and food security. Future research will focus on expanding the applicability of this model across different agricultural contexts, improving its adaptability to different environmental conditions, and further optimizing its performance through advanced techniques such as transfer learning and ensemble methods.

Keywords: PEST DETECTION; EFFICIENTNETV2-L; GRAD-CAM; AGRICULTURAL PRODUCTIVITY; SUSTAINABLE FARMING PRACTICES

1. Introduction

Agriculture plays a crucial role in the economy and ensures a global food supply. However, serious threats to agricultural productivity arise from insect pest attacks, which can cause significant losses (Bouri et al., 2023). Therefore, effective insect pest identification and control cannot be overstated. These measures are vital to maintaining agricultural yield productivity and quality.

In agriculture and ecology, pests are organisms that significantly disrupt the growth, health, and reproduction of plants, livestock, and natural environments. The impacts of climate change on agricultural ecosystems are substantial because they can cause significant crop losses, threaten agricultural sustainability, and even affect global food security (Farooq et al., 2023). This phenomenon raises serious concerns among farmers, agricultural scientists, and policymakers because pest outbreaks can be highly detrimental to food production worldwide. Effective identification and management of pests have become priorities in various agricultural and environmental conservation programs. Understanding the behaviors, spread patterns, and potential damage caused by pests is vital for developing appropriate and efficient protection strategies (Dara et al., 2023).

In this context, technology and innovation in pest detection are becoming increasingly important. In an era where technology underpins revolutions across various life sectors, including agriculture, its use extends to the assessment and management of agricultural issues. Convolutional Neural Networks (CNNs) have emerged as a primary focus in this technological advancement (Bouri et al., 2023). Inspired by the visual cortex structure of the human brain, the proposed neural network architecture can process and analyze image information more complexly than its predecessors. The primary advantage of CNNs lies in their ability to identify highly complex and abstract visual patterns and features in images.

The application of CNNs to the classification and detection of pests in plants has marked a significant leap forward in modern agricultural technology. The ability of CNNs to learn higher-level representations of visual patterns in plant images has enabled more accurate pest detection, faster response times, and more detailed monitoring (Tugrul et al., 2022). This technology not only offers more reliable pest detection solutions but also paves the way for more efficient and sustainable agricultural practices. The application of CNNs in agriculture provides new hope for addressing the challenges faced by farmers. In terms of pest classification, the success of CNNs in improving the accuracy and speed of identifying and categorizing various types of plant pests has significantly impacted crop losses and enhanced agricultural welfare (Zhao et al., 2022).

However, adopting CNNs as pest classification solutions entails complexities, particularly in terms of interpreting and understanding the decision-making processes within the model. Although CNNs offer high accuracy in visual pattern recognition, these models tend to be seen as "black boxes," where their internal decision-making induction and basis are not fully transparent. The high reliability of these models often sacrifices interpretability, which is crucial in agricultural applications. A deeper understanding of how and why models make decisions is required, especially for pest identification contexts. To address this issue, the Grad-CAM (Gradient-weighted Class Activation Mapping) method emerges as an attractive solution. According to Selvaraju et al. (2020) Grad-CAM provides strong visual access to critical areas in images that CNN models focus on. With the proposed method, we can visualize which areas are primary determinants when classifying images containing pests. Thus, while maintaining reliability and accuracy, Grad-CAM can dissect the thought process behind the model's decisions, opening an interpretive window into the image classification process.

Objectives and hypothesis: This study aimed to evaluate the effectiveness of the EfficientNetV2-L model equipped with the Grad-CAM technique in classifying insects for pest detection. We hypothesize that integrating the EfficientNetV2-L model with the Grad-CAM technique will enhance the model's performance and interpretability for insect pest detection, providing deep visual insights that can be applied to real agricultural practices. Through this study, we hope to develop a reliable and efficient tool for real-time pest detection that will support more effective and sustainable agricultural practices.

This paper begins by providing a comprehensive review of the existing literature on pest detection technologies and their applications in agriculture, providing a solid background for understanding advances in this field. The methodology section details the integration of the EfficientNetV2-L model with the Grad-CAM technique, followed by an in-depth analysis of the results obtained from various experimental setups. The discussion interprets these findings in the context of current agricultural practices, and the paper concludes by highlighting potential implications for future research and practical applications in sustainable pest management.

2. Research Method

2.1 Dataset description and ethical considerations

The dataset used in this study comprises over 5,000 images collected from a publicly accessible source on Kaggle (https://www.kaggle.com/datasets/vencerlanz09/agricultural-pests-image-dataset). These images were categorized into 12 classes of agricultural pests, including ants, bees, beetles, caterpillars, earthworms, earwigs, grasshoppers, moths, slugs, snails, wasps, and weevils. Sample dataset is shown in Fig. 1. Each category contains a varied number of images, ensuring diverse representations of pests commonly encountered in agricultural produce.

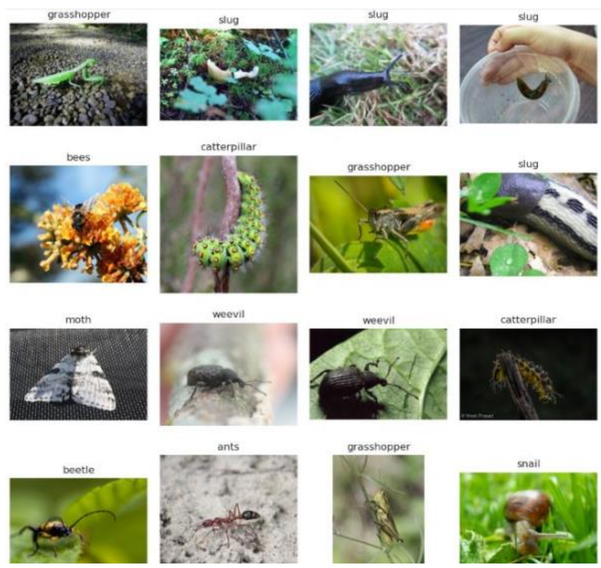

Fig 1. Sample dataset

Data augmentation techniques were employed to enhance the dataset and improve its generalizability. Techniques such as resizing, horizontal flipping, rotation, and contrast adjustments were applied using TensorFlow's ImageDataGenerator. These augmentations increase the dataset diversity, providing the model with a richer set of images to learn from, thereby improving its robustness.

Ethical considerations and permissions for using the dataset by adhering to Kaggle's usage guidelines, ensuring that the data are used responsibly and ethically. The dataset is publicly available and will be used in research; no additional permissions were required.

2.2 Model architecture

The development of the image classification model involved several structured and detailed steps (see in Fig. 2). The process began with data preprocessing, where the dataset was divided into training, validation, and test sets using the train_test_split method from the sklearn library. Tools like ImageDataGenerator and flow_from_dataframe from TensorFlow were used to convert the dataset from DataFrame format to a usable form for the model. The data augmentation process, as described previously, was applied to enrich the training dataset.

The training phase involved loading a pre-trained convolutional neural network (CNN) model, EfficientNetV2-L, with weights from the ImageNet dataset serving as the basis for the image classification model. The model architecture was fine-tuned by adding Dense, Dropout, and output layers with softmax activation functions according to the desired number of classes. The model was compiled using the Adam optimizer, categorical_crossentropy loss function, and accuracy metric, and then trained on training and validation data for 100 epochs with callbacks, such as ModelCheckpoint, EarlyStopping, and TensorBoard.

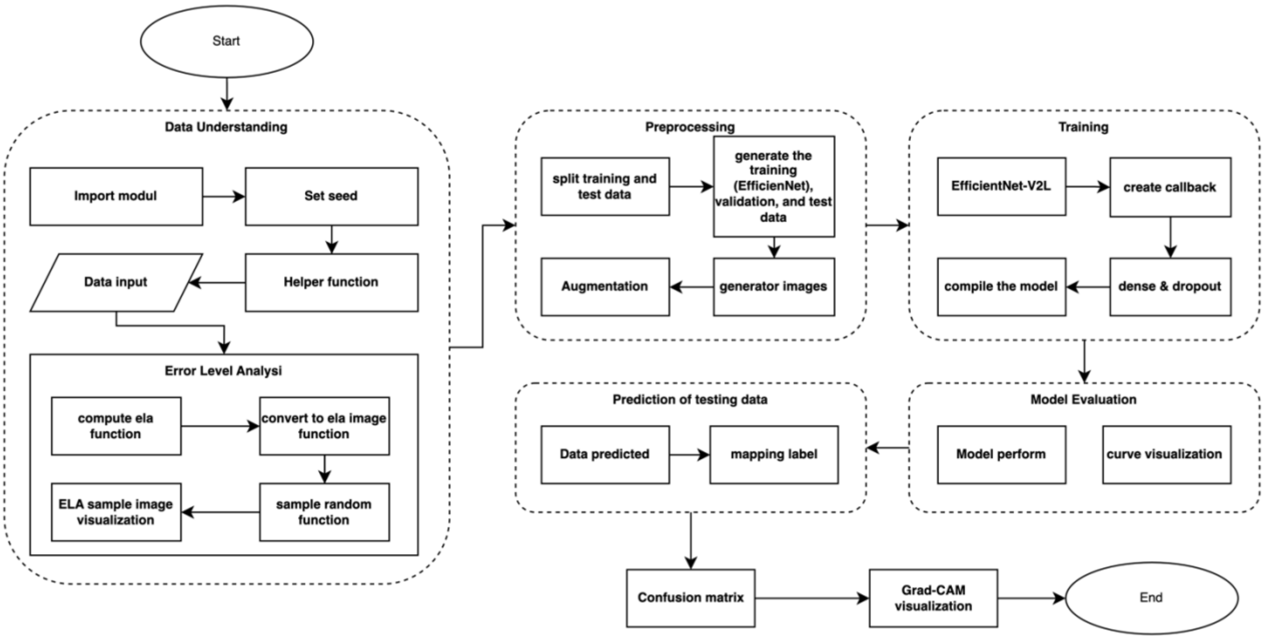

Fig 2. Model proposed

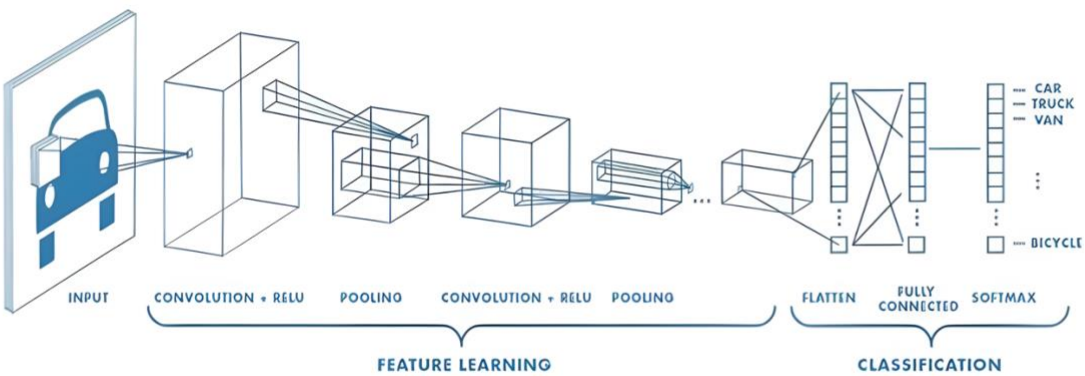

Fig 3. CNN architecture

2.3 Model evaluation

The evaluation phase involved assessing the model using test data to obtain metrics, such as loss, accuracy, precision, recall, and F1 score, using the classification report from sklearn. Predictions on the test data by converting the predicted labels back to class names for comparison with the actual labels.

The final steps included plotting the Classification Reports and Confusion Matrix, which provided a detailed breakdown of the model's performance across different classes and helped identify areas where the model made errors. Grad-CAM visualization was used as a technique to understand which areas of the images the model focused on when making predictions, providing insights into the model's decision-making process.

2.4 Convolutional neural network

In the image processing domain, several algorithms are commonly used, such as naive bayes (Wu et al., 2017), support vector machine (Gholami & Fakhari, 2017), and neural networks (Turkoglu, 2021). Convolutional neural networks (CNN) represent a significant evolution of Neural Networks, specifically designed for digital image recognition. The primary advantage of CNNs lies in their ability to mimic the image recognition system of the human visual cortex. Numerous studies have demonstrated that CNNs are superior models for tasks such as object detection and recognition.

Technically, CNNs comprise several trainable stages, including feature extraction via image convolutions and classification via neural networks. The CNN architecture, inspired by LeNet5, includes key layers such as convolutional, relu, subsampling/pooling, and fully connected layers.

2.5 Grad-CAM

Gradient-weighted class activation mapping (Grad-CAM) is a visualization technique that focuses on creating saliency maps to highlight important areas in an image that influence the predictions of the classification model (Selvaraju et al., 2020). The basic concept involves using a saliency map L_c in the context of binary object classification tasks, where the output is 0 indicating the absence of an object, and 1, indicating its presence. Here, the representation A_k depicts the visualization of the k-th feature map. Previous research indicates that each A_k is triggered by abstract visual patterns, with $A_{ij} = 1$ if the visual pattern is detected, and $A_{ij} = 0$ if not.

Furthermore, the derivative $\frac{\partial yc}{\partial A_{ij}}$ is expected to have a high value for feature map pixels that contribute to the presence of the object. The feature map weights are obtained through a derivative formulation implemented in the Grad-CAM method for a given input image. This process produces a Grad-CAM saliency map that highlights the spatial footprint of the object in the image, which significantly affects visualization. However, if multiple object occurrences or orientation variations occur, the activated feature maps may produce different spatial footprints, affecting the final saliency map's intensity.

To address variations in spatial footprints, Grad-CAM employs a weighted average approach to pixel gradients. In this approach, the weighting coefficients α_{kcij} for class c and the convolutional feature map A_k are calculated by considering pixel gradients, ensuring that the presence of objects in all feature maps is equally highlighted. This method captures the importance of specific activation feature maps A_k, by selecting positive gradients to indicate visual features that enhance the output neuron activation, resulting in a better understanding of the image and its contribution to the classification model's predictions. Thus, Grad-CAM is a powerful alternative for understanding the significance of features in images related to the classification model's output.

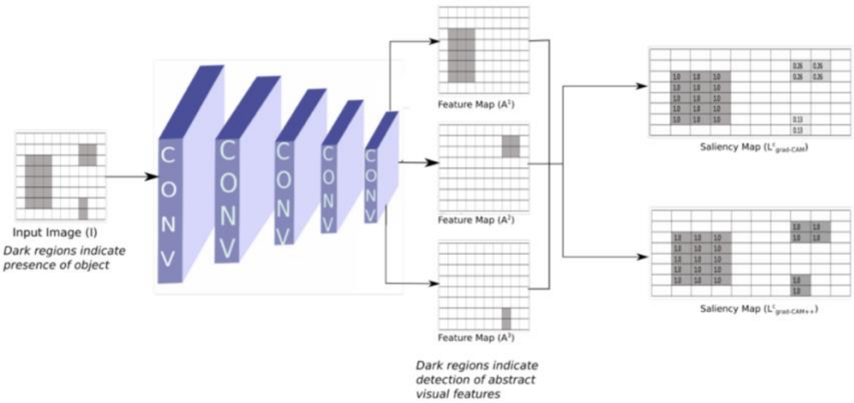

Fig 4. Grad-CAM model architecture

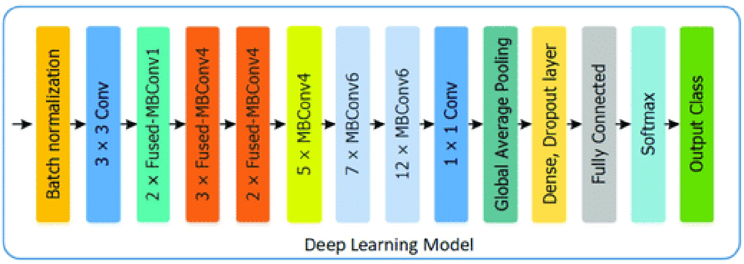

Fig 5. The EfficientNetV2 model architecture (source: Albattah et al., 2022)

ISSN: 2715-6923 e-ISSN: 2721-9186, pp.25~34 | DOI: 10.26714/jichi.v5i1.13959

2.6 EfficientNetV2-L

EfficientNetV2 represents an evolution of the CNN model, demonstrating higher training speed and better parameter efficiency compared to its predecessor, EfficientNet (Tan & Le, 2021). This model employs a training-aware neural architecture search and scaling approach. The core architecture of EfficientNetV2 is illustrated in Fig. 5.

From the illustrated EfficientNetV2 architecture, several new blocks are introduced, which significantly enhance the efficiency and performance compared to EfficientNet. Among these innovative blocks are MBConv (Mobile Inverted Residual Bottleneck Convolution), Fused-MBConv (Fused Mobile Inverted Residual Bottleneck Convolution), and SE (Squeeze-and-Excitation). These blocks are designed to improve operational efficiency and provide more optimal feature representations in the model.

EfficientNet presents a series of models optimized specifically for floating-point operation efficiency (FLOPs) and parameters. This method adopts Neural Architecture Search (NAS) to find the base model, EfficientNet-B0, calibrated to have an optimal balance between accuracy and floating-point operations (FLOPs). The base model is subsequently scaled up using a composite scaling strategy, resulting in the B1-B7 model family. Although some recent studies have claimed significant improvements in training or inference speed, they often perform worse than EfficientNet in terms of parameter efficiency and floating-point operations (FLOPs). This research focuses on increasing the training speed while maintaining optimal parameter efficiency. Therefore, the development of the EfficientNetV2 model involves a new approach to Neural Architecture Search (NAS) that addresses the training and scaling model aspects, specifically, overcoming some of the shortcomings of previous models.

By providing a detailed dataset description and addressing ethical considerations, along with a structured and thorough explanation of the model architecture and evaluation procedures, this methods section offers a comprehensive overview necessary for replicating and understanding the study methodology.

3. Results and Discussion

3.1 Data understanding using ELA

The error level analysis (ELA) results revealed significant observations regarding the impact of compression levels on image fidelity (see in Fig. 6). The primary image on the left displays a clear image of an ant against a white background, which serves as the reference for analysis. Adjacent to this, nine smaller images exhibited ELA outputs at varying quality levels (q: 100, 92, 84, 76, 68, 60, 52, 44, and 36). At higher quality levels (q:100 to q:84), the error levels are relatively low, as indicated by the less prominent colourful pixel patterns, suggesting minimal compression artefacts and high image fidelity. As the quality decreases to medium levels (q:76 to q:60), the error levels increase, evidenced by more noticeable colourful pixel patterns, indicating moderate compression artefacts that start affecting image details. At the lowest quality levels (q:52 to q:36), the error levels are significantly high, with very prominent colourful pixel patterns, implying substantial compression artefacts that considerably degrade the image quality. The grid lines in each ELA output facilitate the analysis of specific areas exhibiting changes, with error patterns becoming more pronounced as quality decreases. These ELA results effectively highlight how varying compression levels impact the image, with increasing error levels corresponding to lower quality settings, thereby aiding in the detection of potential modifications or manipulations in the image.

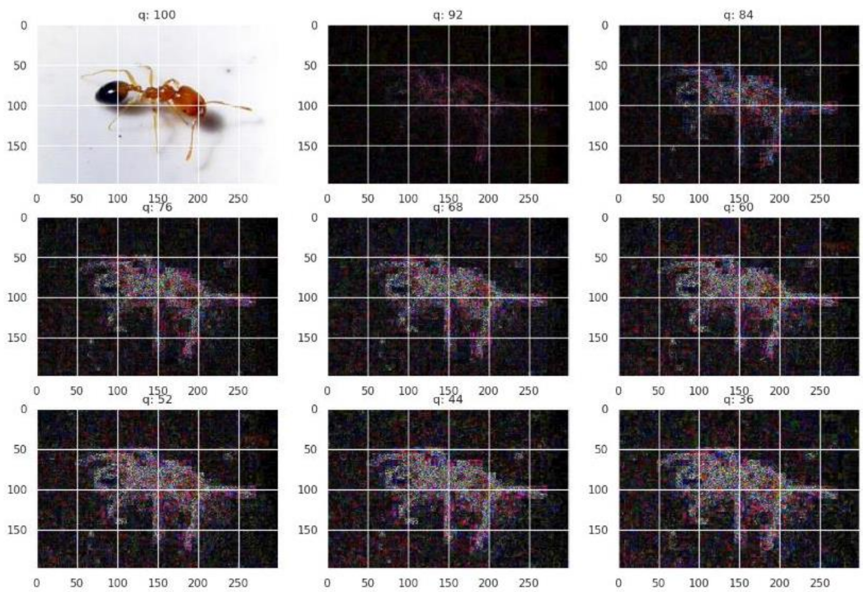

Fig 6. Results of ELA

ISSN: 2715-6923 e-ISSN: 2721-9186, pp.25~34 | DOI: 10.26714/jichi.v5i1.13959

3.2 Preprocessing

The data pre-processing results are summarized in Table 2, which categorizes the images into different classes for training, validation, and testing purposes. A total of 5,494 images were divided, as shown in Table 1.

During the pre-processing phase, the dataset was organized into three distinct subsets: training, validation, and testing. The training set comprised 3,516 images and was used to train the model to identify patterns and features associated with different pest categories. The validation set, which contained 879 images, was used to fine-tune the model, monitor its performance, and prevent overfitting by adjusting the hyperparameters. The testing set, consisting of 1,099 images, was reserved for the final evaluation of the model's performance to provide an unbiased assessment of its accuracy and generalizability.

Data augmentation techniques were applied to enhance model robustness. These techniques included resizing, horizontal flipping, rotation, and contrast adjustment. These augmentations increased the diversity of the training set, which improved the generalizability of the model to new, unseen images. The significant number of images in each subset ensured that the model was exposed to a wide variety of examples, thus improving its ability to accurately classify different types of pests.

The pre-processing phase successfully structured the dataset, ensuring a balanced distribution of images across all classes and subsets. This foundation step is crucial for training a reliable and efficient model capable of performing accurate pest detection in agricultural settings.

3.3 Model evaluation

The evaluation of the model's performance was conducted using a comprehensive set of metrics, including accuracy, precision, recall, and the F1 score. The results presented in Table 3 demonstrate the model's effectiveness in classifying various categories of agricultural pests. The metrics for each class reveal the model's strengths and areas for improvement.

The precision metric reflects the model's accuracy in identifying positive samples, with high precision scores across most classes indicating a low false positive rate. Recall measures the model's ability to correctly identify all relevant instances; high recall scores suggesting a low false negative rate. The F1 score, which combines precision and recall into a single metric, provides a balanced measure of model performance. The support column lists the number of true instances for each class in the test dataset.

Table 1.The caption must be shown before the table.

DATA SPLIT	NUMBER OF IMAGES	NUMBER OF CLASSES
Training	3516	12
Validation	879	12
Testing	1.099	12

Table 2. Classification report.

CLASS	PRECISION	RECALL	F1-SCORE	SUPPORT
Ants	0.91	0.99	0.95	94
Bees	0.9	0.93	0.91	92
Beetle	0.85	0.71	0.77	94
Caterpillar	0.69	0.76	0.73	85
Earthworms	0.9	0.79	0.84	67
Earwig	0.89	0.62	0.73	89
Grasshopper	0.9	0.91	0.91	105
Moth	0.85	0.94	0.89	99
Slug	0.91	0.91	0.91	76
Snail	0.99	1	1	102
Wasp	0.86	0.92	0.89	104
Weevil	0.91	0.99	0.95	92

Fig 7. Example predictions on test data with ground truth and model outputs

The confusion matrix shown in Fig. 8, offers further insights into the model's classification capabilities. This model highlights both correct classifications and common misclassifications, which are crucial for understanding the model's behaviour and areas where it may struggle.

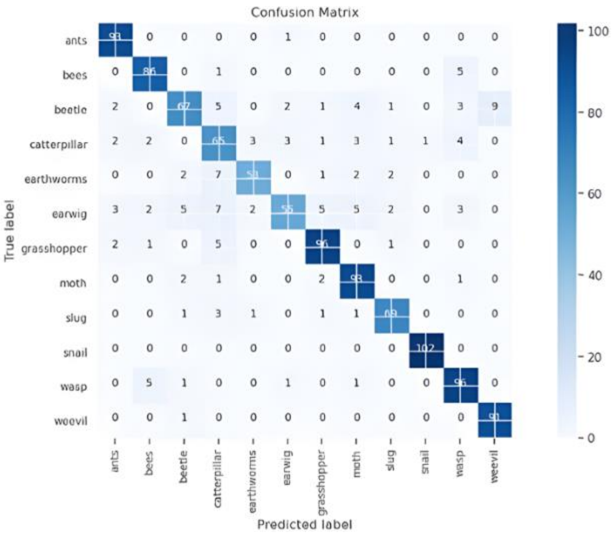

Fig 8. Confusion matrix result

The confusion matrix reveals that the model performs well in correctly identifying most pest categories. However, it also shows instances of misclassification between similar-looking pests, such as caterpillars and beetles. This indicates areas where the model might benefit from further refinement or additional training data to improve its discriminatory power.

a. Macro and weighted averages

The macro average precision, recall, and F1-score (0.88, 0.87, and 0.87 respectively) provide an unweighted mean of the metrics across all classes, highlighting the overall performance without considering class imbalance. The weighted average scores (all at 0.88) accounted for the support of each class, providing a more nuanced view of the model's performance considering the distribution of the dataset (see in Table 3).

Table 3. Macro and weighted averages results.

MATRIX	MACRO AVG	WEIGHTED AVG
Precision	0.88	0.88
Recall	0.87	0.88
F1-score	0.87	0.88

b. Overall accuracy

The overall accuracy of the model (Table 4, stands at 0.88 (88%), indicating a high level of correctness in the model's predictions across the entire test dataset.

Table 4. Overall accuracy.

ACCURACY	NUMBER OF SAMPLES
0.88	1099

As you can see in Table 4, reveal evaluation metrics collectively demonstrate that the EfficientNetV2-L model, combined with Grad-CAM techniques, provides a robust and reliable tool for classifying agricultural pests. The high precision, recall, and F1-scores across multiple classes, along with detailed insights from the confusion matrix, affirm the model's effectiveness and potential applicability in real-world agricultural scenarios.

c. Training and validation performance

Fig's. 9 and 10 present the training and validation accuracy and loss over 100 epochs. Initially, both training accuracy and loss show significant improvements, indicating effective learning from the training data.

However, beyond epoch 40, the validation accuracy begins to plateau while the training accuracy continues to improve, suggesting potential overfitting.

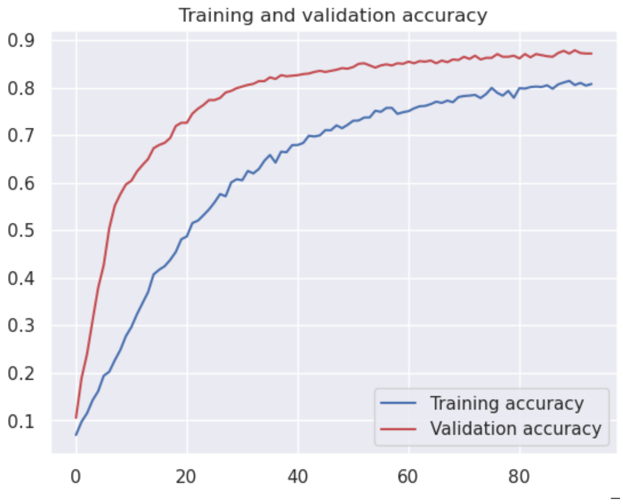

Fig 9. Training and validation accuracy

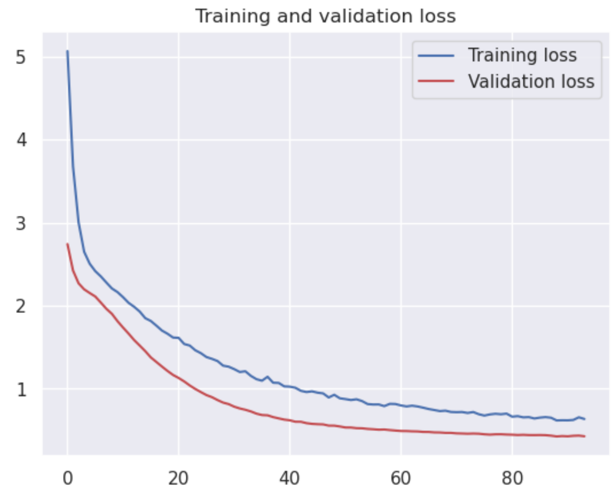

Fig 10. Training and validation loss

To mitigate overfitting, techniques such as early stopping, dropout layers, and regularization were applied. Early stopping halts training when validation performance ceases to improve, preventing the model from over-adapting to the training data. Dropout layers randomly deactivate neurons during training, which helps prevent the model from becoming too dependent on any single neuron. Regularization adds a penalty to the loss function based on the magnitude of the model coefficients, discouraging complexity and encouraging simplicity.

3.4 Discussion

a. Model performance and implications

The results of this study demonstrate that the EfficientNetV2-L model, augmented with Grad-CAM visualization techniques, provides a robust solution for the classification of agricultural pests. With an overall accuracy of 88%, the model exhibits strong performance across various metrics, including precision, recall, and F1-score. These high-performance metrics suggest that the model is capable of effectively distinguishing between different pest categories, making it a valuable tool for real-world agricultural applications.

The precision values indicate the model's efficacy in minimizing false positives, ensuring that pest detection is accurate and reliable. High recall values across most classes demonstrate the model's proficiency in identifying true positive instances, which is crucial for timely pest management and intervention. The F1-scores, which balance precision and recall, confirm the overall robustness of the model, ensuring it performs consistently well across different pest categories.

b. Overfitting and underfitting

During training, the model initially showed significant improvement in both training and validation accuracy, but beyond a certain point, the validation accuracy plateaued while training accuracy continued to rise, indicating potential overfitting. To mitigate this, early stopping, dropout layers, and regularization techniques were employed. These measures were effective in preventing the model from overfitting to the training data, thereby enhancing its generalization capabilities when applied to unseen test data.

c. Misclassification analysis

The confusion matrix and the Grad-CAM visualizations provide deeper insights into the model's decision-making process and areas of misclassification. Misclassifications between visually similar classes, such as caterpillars and beetles or slugs and snails, highlight the model's current limitations. These errors can be attributed to the overlapping visual features that these classes share, which can confuse the model. Addressing these misclassifications would require additional training data with more distinct features for these categories or further refinement of the model's architecture to better capture subtle differences.

d. Practical implications

The practical implications of this research are substantial. The high accuracy and reliability of the model suggest that it can be effectively integrated into automated pest detection systems in agricultural settings. By providing farmers with a tool that can accurately identify pests in real-time, it enables early intervention and more efficient pest management strategies. This can lead to significant reductions in crop damage, improved yields, and more sustainable farming practices.

e. Limitations

Despite its high performance, the model has some limitations that need to be addressed. The dataset, while comprehensive, may still lack sufficient diversity to capture all variations of pest appearances in different environmental conditions. Future work should focus on expanding the dataset to include a broader range of images, covering different pest life stages and varying environmental contexts. Additionally, further refinement of the model's architecture and hyperparameters could enhance its performance, particularly in reducing misclassifications.

The integration of more advanced techniques, such as ensemble learning or transfer learning from other well-established models, could also be explored to improve the model's accuracy and robustness. Furthermore, the development of user-friendly interfaces and deployment mechanisms will be essential for practical field applications, ensuring that the technology is accessible and beneficial to end-users such as farmers and agricultural technicians.

f. Integration with previous research

The findings of this study align with and extend the existing body of research on pest detection and management. Previous studies, such as those by Dong et al. (2022) and Wen et al. (2022), have highlighted the potential of convolutional neural networks (CNNs) in improving pest detection accuracy. This study builds upon these foundations by integrating the EfficientNetV2-L model, which demonstrates higher training speed and parameter efficiency, with Grad-CAM, which enhances interpretability. The combined approach not only improves model performance but also provides valuable visual insights, addressing the interpretability challenge often associated with CNNs.

By enhancing the accuracy and transparency of pest detection systems, this research supports sustainable farming practices and contributes to broader goals of food security and agricultural productivity. Future work should continue to explore and refine these approaches, ensuring that the benefits of advanced artificial intelligence techniques are fully realized in real-world agricultural settings.

4. Conclusion and future work

This study demonstrated the efficacy of the EfficientNetV2-L model, augmented with Grad-CAM visualization techniques, in accurately classifying agricultural pests. The model achieved a high overall accuracy of 88%, indicating its potential as a reliable tool for pest detection in real-world agricultural settings. High precision, recall, and F1-scores across various pest categories underscore the model's robustness and effectiveness in distinguishing between different types of pests, thereby facilitating timely and accurate pest management. The integration of Grad-CAM provided valuable insights into the decision-making process of the model, enhancing interpretability and enabling the identification of critical image regions influencing the model's predictions. This transparency is crucial for gaining trust from end-users and for further refining the model based on real-world feedback.

While the current study has achieved significant milestones, several avenues for future research and improvements remain. Future work should focus on expanding the dataset to include a more diverse range of images, encompassing different environmental conditions, various pest life stages, and diverse agricultural contexts. A larger and more varied dataset will help the model generalize better and improve its performance on unseen data. The analysis revealed areas where the model struggled, particularly in distinguishing between visually similar pests. Future efforts should explore methods to enhance the model's discriminatory power, such as fine-tuning the feature extraction layers or incorporating additional contextual information.

Investigating the use of advanced model architectures and techniques, such as ensemble learning or transfer learning from other well-established models, could further enhance the model's accuracy and robustness. Combining multiple models or leveraging pre-trained networks may provide additional performance gains. For practical applications, it is essential to develop user-friendly interfaces and efficient deployment mechanisms. This includes creating mobile applications or integrating the model into existing agricultural management systems to facilitate real-time pest detection and monitoring.

Conducting extensive field tests and gathering feedback from end users, such as farmers and agricultural technicians, will be crucial for refining the model and ensuring its practical utility. Real-world testing can highlight additional challenges and areas for improvement that are not obvious in controlled environments. Future research should also consider the ethical and environmental implications of deploying such technologies. Ensuring data privacy, minimizing the technological footprint, and promoting sustainable agricultural practices are important aspects to address. By pursuing these future research directions, the potential of AI-driven pest detection can be fully realized, leading to more effective, sustainable, and efficient agricultural practices. The findings of this study lay a strong foundation for further advancements in this field, ultimately contributing to improved food security and agricultural productivity.

Acknowledgements

We would like to express our sincere gratitude to the Department of Informatics, Faculty of Engineering and Computer Science, Universitas Muhammadiyah Semarang, for providing the research tools in the Intelligent Systems Laboratory, which were essential for conducting this study. This research was undertaken as part of the requirements for completing our Bachelor’s degree in Informatics. Additionally, we extend our heartfelt thanks to Jafar Abdollahi from Islamic Azad University, Iran, for his invaluable contributions and guidance throughout this collaborative research, which was proposed by our esteemed lecturers, Muhammad Munsarif dan Ahmad Ilham, during the writing process.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

References

Bouri, M., Arslan, K. S., & Şahin, F. (2023). Climate-Smart Pest Management in Sustainable Agriculture: Promises and Challenges. *Sustainability, 15*(5), 4592. https://doi.org/10.3390/su15054592

Dara, S. K., Rodriguez-Saona, C., & Morrison, W. R. (2023). Editorial: Integrated pest management strategies for sustainable food production. *Frontiers in Sustainable Food Systems, 7*. https://doi.org/10.3389/fsufs.2023.1224604

Dong, S., Du, J., Jiao, L., Wang, F., Liu, K., Teng, Y., & Wang, R. (2022). Automatic Crop Pest Detection Oriented Multiscale Feature Fusion Approach. *Insects*, *13*(6), 554. https://doi.org/10.3390/insects13060554

Farooq, A., Farooq, N., Akbar, H., Hassan, Z. U., & Gheewala, S. H. (2023). A Critical Review of Climate Change Impact at a Global Scale on Cereal Crop Production. *Agronomy*, *13*(1), 162. https://doi.org/10.3390/agronomy13010162

Gholami, R., & Fakhari, N. (2017). Support Vector Machine: Principles, Parameters, and Applications. In *Handbook of Neural Computation* (pp. 515–535). Elsevier. https://doi.org/10.1016/B978-0-12-811318-9.00027-2

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision*, *128*(2), 336–359. https://doi.org/10.1007/s11263-019-01228-7

Tan, M., & Le, Q. (2021). EfficientNetV2: Smaller Models and Faster Training. *Proceedings of the 38th International Conference on Machine Learning. PMLR.*, 10096–10106. https://proceedings.mlr.press/v139/tan21a.html

Tugrul, B., Elfatimi, E., & Eryigit, R. (2022). Convolutional Neural Networks in Detection of Plant Leaf Diseases: A Review. *Agriculture*, *12*(8), 1192. https://doi.org/10.3390/agriculture12081192

Turkoglu, M. (2021). COVIDetectioNet: COVID-19 diagnosis system based on X-ray images using features selected from pre-learned deep features ensemble. *Applied Intelligence*, *51*(3), 1213–1226. https://doi.org/10.1007/s10489-020-01888-w

Wen, C., Chen, H., Ma, Z., Zhang, T., Yang, C., Su, H., & Chen, H. (2022). Pest-YOLO: A model for large-scale multi-class dense and tiny pest detection and counting. *Frontiers in Plant Science*, *13*. https://doi.org/10.3389/fpls.2022.973985

Wu, M.-E., Wang, C.-H., & Chung, W.-H. (2017). Using trading mechanisms to investigate large futures data and their implications to market trends. *Soft Computing*, *21*(11), 2821–2834. https://doi.org/10.1007/s00500-016-2162-6

Zhang, L., Cui, H., Sun, J., Li, Z., Wang, H., & Li, D. (2023). CLT-YOLOX: Improved YOLOX Based on Cross-Layer Transformer for Object Detection Method Regarding Insect Pest. *Agronomy*, *13*(8), 2091. https://doi.org/10.3390/agronomy13082091

Zhao, S., Liu, J., Bai, Z., Hu, C., & Jin, Y. (2022). Crop Pest Recognition in Real Agricultural Environment Using Convolutional Neural Networks by a Parallel Attention Mechanism. *Frontiers in Plant Science*, *13*. https://doi.org/10.3389/fpls.2022.839572