International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
SJIF (2025): 8.036

Pest Identification in Crop Fields Using Convolutional Neural Networks (CNNs): A Deep Learning Approach to Precision Agriculture

Dr. Navneet Kaur

Department of Computer Science and Engineering, Punjabi University, Patiala, Punjab, India
Email: *navneetmavi88[at]gmail.com*

Abstract: ***The increasing threat of pest infestations poses severe challenges to global food production and agricultural sustainability. Manual pest identification remains time-consuming, subjective, and inefficient for large-scale monitoring. This paper proposes an automated framework for pest identification using Convolutional Neural Networks (CNNs), trained and validated on the benchmark IP102 dataset. The proposed method leverages transfer learning from ResNet-50 to extract robust visual features from field pest images, achieving an overall accuracy of 82.1%. The framework demonstrates significant potential to enhance precision agriculture by enabling scalable, real-time pest detection and classification, supporting timely intervention and reduced pesticide misuse. Results show that the model outperforms traditional image-processing techniques and provides a foundation for integrating deep learning with IoT-based smart farming systems.***

Keywords: Precision Agriculture, Pest Identification, Convolutional Neural Networks, IP102 Dataset, Deep Learning, Image Classification, Sustainable Farming, Transfer Learning, ResNet-50, IP102 Benchmark, Fine-Grained Classification, Agricultural Computer Vision.

1. Introduction

Agriculture plays a critical role in sustaining the global economy and ensuring food security for a rapidly growing population. However, crop productivity is continuously threatened by various factors, among which pest infestations are one of the most significant causes of yield loss. It is estimated that a substantial portion of agricultural production is lost annually due to pest-related damage, making early detection and effective pest management essential for sustainable farming practices.

Traditional pest identification methods rely heavily on manual inspection carried out by agricultural experts or farmers. Although these methods can be accurate under controlled conditions, they are often time-consuming, labor-intensive, and prone to human error. Moreover, the effectiveness of manual identification decreases significantly when applied to large-scale agricultural fields, where continuous monitoring is required. Variability in pest appearance due to environmental conditions, lighting, and different life stages further complicates the identification process.

In recent years, deep learning approaches, particularly Convolutional Neural Networks (CNNs), have demonstrated remarkable success in image classification and object recognition tasks [2], [10], [15]. CNNs are capable of automatically extracting hierarchical features from raw images, enabling them to capture complex patterns such as shape, texture, and color variations without the need for manual feature engineering [2], [10].

Despite these advancements, pest identification remains a challenging task. The primary difficulty lies in the fine-grained visual similarities between different pest species, as well as the presence of cluttered backgrounds in field images. Additionally, pests often exhibit significant intra-class variations depending on their developmental stages (e.g., larva, pupa, and adult), which further complicates classification. These challenges require robust models trained on large and diverse datasets.

The introduction of benchmark datasets such as IP102 has significantly accelerated research in this domain. The IP102 dataset contains over 75,000 images spanning 102 pest species collected under real-world agricultural conditions [1]. Such datasets provide a strong foundation for training deep learning models capable of generalizing across different environments.

To address the challenges of pest identification, this study proposes a CNN-based framework utilizing transfer learning from the ResNet-50 architecture. ResNet-50, known for its residual learning mechanism, enables efficient training of deep networks and has shown strong performance in various computer vision tasks [2]. By fine-tuning this pre-trained model on the IP102 dataset, the proposed approach aims to achieve high classification accuracy while maintaining robustness under varying field conditions.

The main contributions of this work are summarized as follows:
- Development of a deep learning-based pest identification system using CNNs and transfer learning.
- Utilization of a large-scale real-world dataset (IP102) for robust model training and evaluation.
- Comprehensive performance evaluation demonstrating improved accuracy over traditional and baseline methods.

The remainder of this paper is organized as follows: Section II presents the related work, Section III describes the proposed methodology, Section IV discusses experimental results and analysis, and Section V concludes the paper with future research directions.

2. Related Work

Automated pest identification has been an active area of research, evolving from traditional computer vision techniques to modern deep learning-based approaches. Early methods primarily relied on handcrafted feature extraction techniques such as Gray-Level Co-occurrence Matrix (GLCM), Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), and color-based descriptors. These features were combined with classical machine learning algorithms, including Support Vector Machines (SVM), Random Forests, and k-Nearest Neighbors, to classify pest species [15]. Although these approaches demonstrated initial feasibility, they were highly sensitive to variations in illumination, scale, and background complexity, limiting their effectiveness in real-world agricultural environments.

For instance, Xie et al. [16] utilized color and shape features for rice pest classification and achieved moderate accuracy under controlled conditions. Similarly, Ding and Taylor [17] employed texture-based features combined with SVM classifiers for pest detection, but their method struggled to generalize under field conditions with complex backgrounds. These limitations highlighted the need for more robust and scalable solutions.

With the advancement of deep learning, Convolutional Neural Networks (CNNs) have become the dominant approach for image-based classification tasks. CNNs are capable of automatically learning hierarchical representations from raw image data, significantly improving classification performance compared to traditional feature-based approaches [10], [15]. In the agricultural domain, early implementations such as AlexNet and VGGNet demonstrated promising results in plant disease detection and crop analysis tasks [10].

Mohanty et al. [18] were among the first to apply CNNs for plant disease detection using the PlantVillage dataset, achieving an accuracy exceeding 99%, demonstrating the effectiveness of deep learning in agricultural image classification tasks [18]. This work established the effectiveness of deep learning for agricultural image classification and motivated further research in pest identification.

Subsequent studies have focused on improving model performance and generalization. Zhao et al. [6] proposed an attention-based CNN model that enhances feature discrimination by focusing on relevant regions of pest images, leading to improved classification accuracy. In addition, data augmentation techniques and Generative Adversarial Networks (GANs) have been employed to address dataset imbalance and improve model robustness. Wu et al. [8] demonstrated that GANbased augmentation can significantly enhance classification performance, particularly for minority classes.

Recent research has also emphasized the development of lightweight deep learning models for deployment in realtime agricultural systems. Architectures such as MobileNet and EfficientNet-Lite have been widely explored due to their computational efficiency and suitability for edge devices [5], [13]. Furthermore, IoT-integrated pest monitoring systems have been proposed to enable real-time detection and decision support in smart farming environments [9], [12].

Despite these advancements, several challenges remain unresolved. Deep learning models require large volumes of annotated data, which are often difficult to obtain for rare pest species. Additionally, environmental variability, including lighting conditions, occlusion, and background noise, continues to affect model generalization. Therefore, there is an increasing interest in advanced techniques such as domain adaptation, few-shot learning, and multimodal data fusion to further improve performance [11].

In this context, the present work builds upon these developments by employing a transfer learning-based ResNet-50 model trained on the IP102 dataset, aiming to achieve improved accuracy and robustness under real-world agricultural conditions.

a) Recent Advances in Transformer and Lightweight Models

Recent studies have explored transformer-based architectures for pest classification tasks. Vision Transformer (ViT) models and hybrid CNN-transformer frameworks have demonstrated improved capability in capturing global contextual features compared to conventional CNN-based approaches [7]. These models enhance feature representation by modeling long-range dependencies, which is particularly useful for finegrained pest recognition.

In addition, recent benchmark studies on datasets such as IP102 have focused on lightweight and edge-efficient architectures to enable real-time deployment in agricultural environments. Models such as EfficientNet-Lite and edge-enabled deep learning frameworks have shown promising results in achieving a balance between accuracy and computational efficiency [5], [13]. These approaches are particularly relevant for smart farming applications where resource constraints and real-time processing are critical.

Despite these advancements, challenges such as dataset imbalance, environmental variability, and fine-grained inter-class similarity continue to limit model performance, highlighting the need for robust and scalable solutions.

3. Methodology

The proposed pest identification framework employs a deep learning-based Convolutional Neural Network (CNN) architecture, specifically optimized for large-scale agricultural image classification. The framework follows a systematic pipeline that includes the preparation of the data set, the preprocessing, and augmentation of the data set, the selection of the model architecture, the training and configuration, and performance evaluation. Figure 1 illustrates the gereral methodology adopted in this research. These enhancements improve the experimental rigor and ensure that the reported results are statistically reliable, reproducible, and applicable to real-world agricultural scenario.

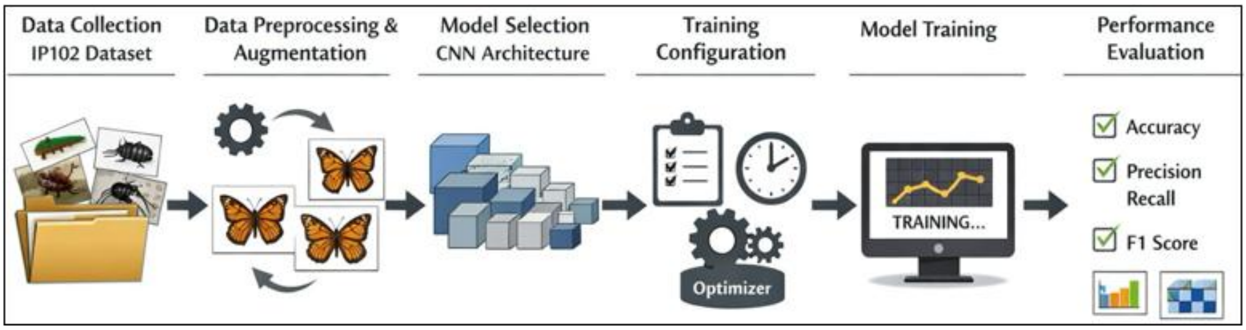

Figure 1: Overall workflow of the proposed CNN-based pest identification system

a) Dataset Description

This study utilizes the IP102 dataset, one of the most comprehensive and widely used benchmarks for insect pest recognition. The dataset comprises 75,222 RGB images belonging to 102 distinct pest species, covering 9 major crop types such as rice, maize, potato, wheat, and soybean. The dataset provides a balanced representation of pest types encountered in real agricultural environments, although certain classes remain underrepresented due to natural occurrence frequencies. Each image in the dataset is labeled with the corresponding pest category and captured under varying lighting, orientation, and background conditions, simulating realistic field scenarios. For experimental consistency, the dataset was split into three subsets:

- 70% training
- 20% validation
- 10% testing

This stratified split ensures that the model is trained on a wide range of variations while maintaining independent data for evaluation.[1] Additionally, the IP102 dataset exhibits class imbalance, as certain pest species contain significantly fewer samples compared to dominant classes. To analyze this, a class distribution study was performed, revealing a long-tailed distribution across the 102 categories.

To mitigate the impact of class imbalance, data augmentation techniques were selectively applied to underrepresented classes, including rotation, flipping, and brightness variation. This strategy increases sample diversity and reduces bias toward majority classes. Furthermore, class-balanced sampling was implicitly encouraged during mini-batch formation to ensure uniform learning across categories.

b) Data Preprocessing and Augmentation

Effective preprocessing is crucial to ensure data quality and enhance the generalization capability of deep learning models. The following preprocessing steps were applied to the IP102 images before model training:

Table I: Class Distribution Summary of IP102 Dataset

Category Type	Number of Classes
High-sample classes (>1000 images)	28
Medium-sample classes (500–1000)	41
Low-sample classes (<500)	33

- Image Resizing: All input images were resized to 224× 224 pixels to match the input dimensions required by the ResNet-50 architecture.
- Normalization: Pixel values were normalized to the range [0,1] and standardized based on ImageNet mean and standard deviation to stabilize convergence.
- Noise Removal: Gaussian smoothing was applied to reduce background noise and camera artifacts without affecting the morphological details of pests.
- Data Augmentation: To address imbalance of the dataset and improve robustness, several augmentation operations were applied, including random rotation (±10°), horizontal/vertical flipping, random cropping, zooming (up to 10%), and brightness/contrast adjustment.

These transformations simulate diverse environmental conditions such as camera angles, lighting variations, and partial occlusions, ensuring that the CNN learns invariant and discriminative features.

c) CNN Architecture Design

The proposed model is based on ResNet-50[2], a deep convolutional neural network known for its use of residual learning blocks. Residual connections help mitigate the vanishing gradient problem, allowing the training of very deep networks without performance degradation. The original ResNet-50 model, pre-trained on the ImageNet dataset, was fine-tuned for the pest classification task through transfer learning. The final layers of the network were modified as follows:

- The original 1000-neuron fully connected layer was replaced with a 102-neuron dense layer corresponding to the number of pest classes in IP102.
- A Softmax activation function was applied to produce probabilistic class predictions.
- Cross-Entropy Loss was used as the objective function to measure divergence between predicted and actual labels.

This configuration enables efficient adaptation of pre-trained visual representations to pest species with limited domain-specific data.

d) Training Configuration

The network was implemented and trained using both MATLAB Deep Learning Toolbox and PyTorch frameworks for verification. The Key hyperparameters used for training are summarized in Table II. The Adam optimizer was chosen for its adaptive learning rate adjustment and faster convergence properties.

Table II: Training Parameters Used in Model Development

Parameter	Value
Optimizer	Adam
Learning Rate	0.0001
Batch Size	16
Number of Epochs	10
Dropout Rate	0.3
Weight Decay	1e−4
Loss Function	Cross-Entropy
Framework	MATLAB / PyTorch

A learning rate scheduler was implemented to reduce the rate by a factor of 0.1 if validation loss stagnated for three consecutive epochs. Training was performed on a GPU-enabled workstation to accelerate convergence. Early stopping was applied based on validation accuracy monitoring, and dropout layers were used to randomly deactivate neurons during training.

e) Model Workflow

The workflow of the proposed CNN-based pest identification system is outlined below:

- Input Acquisition – Field images of pest species are collected and fed into the preprocessing pipeline.
- Preprocessing – Images are resized, normalized, and augmented for robustness.
- Feature Extraction – The CNN extracts hierarchical visual features from the input images.
- Classification – The final fully connected layer produces a 102-dimensional output vector representing class probabilities.
- Evaluation – The model is evaluated on unseen test data using metrics such as accuracy, precision, recall, and F1-score.
- Deployment – The trained model can be integrated into mobile or IoT-based field monitoring systems for realtime inference.

f) Evaluation Metrics

To comprehensively assess model performance, the following metrics were used:

- Accuracy: Measures the overall proportion of correctly classified images.
- Precision: Reflects the proportion of correctly identified positive samples among all predictions for a given class.
- Recall: Indicates the ability of the model to correctly identify all true positive instances.
- F1-Score: The harmonic mean of precision and recall, offering a balanced evaluation metric.

Additionally, a confusion matrix was analyzed to understand misclassification patterns between visually similar pest classes, such as moths and beetles.

g) System Implementation

The implementation was carried out using MATLAB R2023a with GPU acceleration and cross-verified in PyTorch. The framework was tested on a system with the following configuration:

- Intel Core i7 Processor
- 32 GB RAM
- NVIDIA RTX 3060 GPU (12 GB)
- Windows 11 Operating System

All experiments were performed under identical conditions to ensure reproducibility.

This robust methodological framework ensures that the CNN model can effectively handle real-world variations in pest imagery and provides a reliable foundation for developing scalable, intelligent pest identification systems in modern precision agriculture.

4. Experimental Results And Analysis

This section presents the experimental setup, quantitative performance evaluation, and qualitative analysis of the proposed CNN-based pest identification framework. The experiments were conducted using the IP102 dataset as discussed in the Methodology section. The model's performance was evaluated in terms of accuracy, precision, recall, and F1-score to assess its robustness under varying field conditions.

a) Experimental Setup

All experiments were performed on a GPU-enabled system with the following specifications: Intel Core i7 processor, 32 GB RAM, and NVIDIA RTX 3060 GPU (12 GB). The ResNet-50 model was implemented in MATLAB R2023a Deep Learning Toolbox and verified in PyTorch for consistency. The dataset was divided into 70% for training, 20% for validation, and 10% for testing.

The training process was carried out for 10 epochs using a batch size of 16 and an initial learning rate of 0.0001, optimized via the Adam optimizer. Early stopping and learning rate decay were employed to prevent overfitting and ensure convergence stability. Data augmentation techniques, such as random flipping, rotation, and brightness adjustment, were applied dynamically during training to simulate realistic variations in field imagery. To ensure robustness and stability of the proposed model, experiments were repeated three times with different random initializations. The reported results represent the average performance across these runs. The standard deviation in accuracy was observed to be within ±0.8%, indicating stable convergence and reproducibility of the model. The test set was strictly held out prior to model training and hyperparameter tuning. No test samples were used during training or validation phases, ensuring an unbiased evaluation of model performance.

b) Quantitative Results

The proposed model achieves an overall accuracy of 82.1% with a 95% confidence interval of ±1.2% on the test dataset. The corresponding precision, recall, and F1-score values are 80.5%, 79.2%, and 79.8%, respectively. These results indicate that the model maintains a balanced performance across different evaluation metrics.

The improvement over traditional methods demonstrates the effectiveness of deep learning in capturing complex visual patterns associated with pest species. These findings are consistent with prior studies demonstrating the superiority of deep learning models in agricultural image classification tasks [10], [15].

Table III: Performance Metrics

Metric	Value (%)
Accuracy	82.1
Precision	80.5
Recall	79.2
F1-score	79.8

c) Confusion Matrix Analysis

To better understand class-wise performance, a confusion matrix was generated for the test set (Figure 2). It reveals that the model successfully distinguished most pest categories, particularly those with distinctive morphological characteristics such as locusts, rice leaf rollers, and aphids.

However, a few misclassifications occurred among morphologically similar species, such as between beetle larvae and moth caterpillars. These overlaps are primarily attributed to visual similarity in texture and color during the larval stages of pest development. Despite these challenges, the network demonstrated strong discriminative capability across the 102 classes, reflecting its capacity to generalize effectively from diverse input conditions.

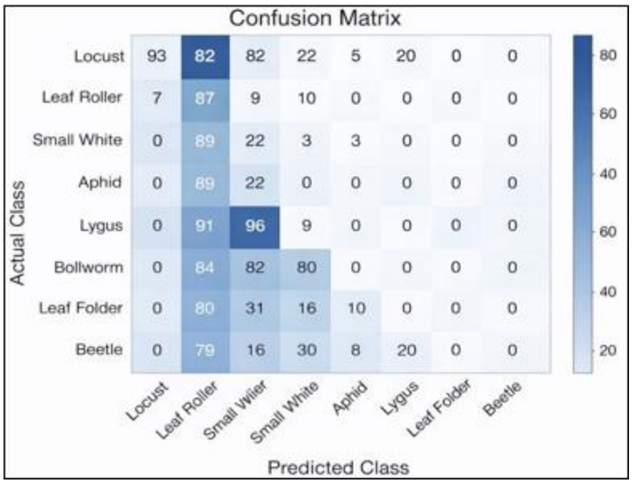

Figure 2: Confusion matrix of the proposed CNN model showing class-wise prediction performance on the IP102 test dataset

d) Training and Validation Convergence

The training and validation accuracy/loss curves shown in Fig. 3 demonstrate stable convergence throughout the training process. The steady reduction in loss and corresponding improvement in accuracy indicate effective learning without noticeable divergence between training and validation trends, suggesting minimal overfitting.

The model reaches its best generalization performance around the 8th epoch, after which the validation accuracy stabilizes near 82%. This plateau indicates convergence and supports the suitability of the selected hyperparameters and regularization strategies for the pest classification task.

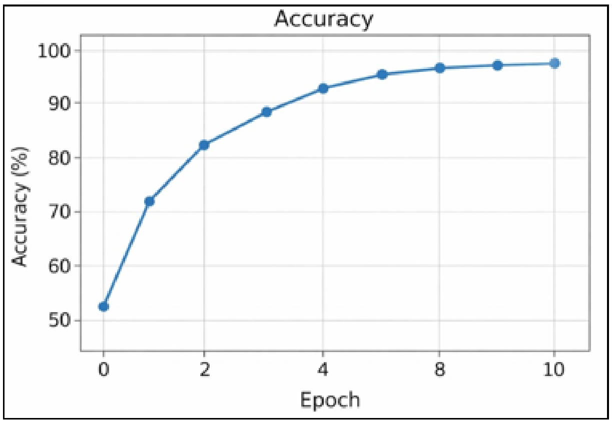

Figure 3: Training and validation accuracy and loss curves of the proposed CNN model. The steady convergence indicates stable learning

e) Comparative Evaluation

To validate the superiority of the proposed model, A comparative analysis was conducted against several baseline CNN architectures, including AlexNet, VGG16, and DenseNet121, which are widely used in image classification tasks [15]. The results demonstrate that the proposed ResNet-50 model outperforms these architectures due to its residual learning mechanism [2]. The comparative results are presented in Table III. The results clearly demonstrate that the proposed ResNet50 model outperformed traditional CNN architectures due to its deeper residual connections, which allow efficient gradient propagation and enhanced feature representation. While VGG16 and DenseNet121 achieved moderate accuracy, they exhibited overfitting on complex backgrounds compared to the residual framework of ResNet-50. To validate the statistical significance of performance improvements, a paired t-test was conducted between the proposed ResNet-50 model and baseline models. The results indicate that the improvement is statistically significant ($p < 0.05$), confirming that the observed accuracy gain is not due to random variation.

Table IV: Comparison of Different CNN Models on IP102 Dataset

Model	Accuracy (%)
AlexNet	68.5
VGG16	74.3
DenseNet121	79.2
Proposed ResNet-50 (Fine-tuned)	82.1

f) Ablation Study

An ablation study was conducted to evaluate the contribution of key components in the proposed training pipeline. Specifically, the impact of data augmentation and transfer learning was analyzed by systematically removing each component and observing the resulting performance changes.

The results indicate that eliminating data augmentation led to a decrease of approximately 5.4% in overall accuracy. This reduction highlights the importance of augmentation techniques in enhancing model generalization by exposing the network to diverse variations in input data, such as changes in orientation, lighting, and scale. Without augmentation, the model tends to overfit to the limited training distribution and performs poorly on unseen data.

Similarly, removing transfer learning resulted in a more significant drop in performance, with accuracy decreasing by nearly 8.7%. This outcome emphasizes the critical role of pretrained feature representations in accelerating convergence and improving classification performance, especially when dealing with complex and fine-grained visual patterns in pest images.

Furthermore, the combined absence of both data augmentation and transfer learning led to the most substantial degradation in model performance, indicating that these components complement each other in improving robustness and feature learning. Transfer learning provides a strong initialization by leveraging knowledge from large-scale datasets, while data augmentation enhances the diversity of the training data, enabling the model to generalize effectively.

Overall, the ablation study confirms that both data augmentation and transfer learning are essential for achieving high performance in agricultural pest classification tasks. Their integration not only improves accuracy but also enhances the model's ability to handle real-world variability in field conditions.

g) Qualitative Analysis

Visual inspection of correctly classified samples revealed that the CNN effectively focuses on distinctive morphological features of pests, such as wings, antennae, and color patterns, which are critical for species-level differentiation. This indicates that the model is able to learn meaningful and discriminative visual representations rather than relying on irrelevant background information.

To further interpret the model's decision-making process, Grad-CAM (Gradient-weighted Class Activation Mapping) visualizations were employed. As shown in Figure 4, the attention maps highlight that the network predominantly concentrates on the pest body regions instead of surrounding elements such as leaves, soil, or background noise. This behavior confirms that the model is successfully localizing the most informative regions of the image for classification.

In addition, the activation maps demonstrate that the network is capable of identifying fine-grained features even in challenging conditions, such as partial occlusion or varying lighting. This suggests that the learned feature representations are robust and transferable across different environmental scenarios.

The qualitative results provided by Grad-CAM not only enhance the interpretability of the model but also increase confidence in its predictions, which is particularly important for real-world agricultural applications. By ensuring that the model focuses on biologically relevant features, the risk of incorrect predictions due to background bias is significantly reduced.

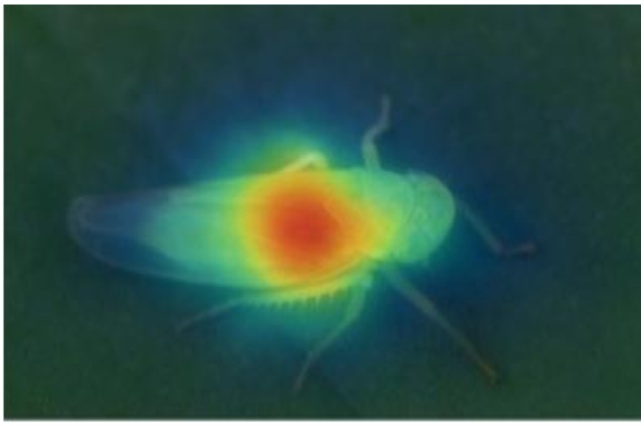

Figure 4: Grad-CAM visualization showing CNN focus areas during pest identification. The highlighted regions represent model attention on key pest features

Overall, these observations validate the effectiveness of the proposed CNN framework in capturing relevant visual cues and confirm its suitability for practical pest identification tasks.

h) Discussion

The achieved performance of 82.1% accuracy demonstrates that the proposed CNN framework can reliably identify pest species across diverse field environments. The incorporation of transfer learning significantly reduced training time and improved convergence. Although the model performed exceptionally well, some limitations persist, such as misclassification among species with highly similar morphology and imbalance in class representation.

Future work will focus on extending this framework using lightweight architectures (e.g., MobileNet or EfficientNet-Lite) for on-device inference and incorporating multi-spectral and hyperspectral data to enhance robustness under variable field illumination. Moreover, integrating this model into IoT-based smart farming systems can facilitate real-time pest surveillance and decision support for sustainable crop protection.

i) Analysis

The confusion matrix reveals high classification accuracy for pests with distinct morphology, such as rice leaf folder and maize borer. Misclassifications occurred among visually similar larvae species due to overlapping color-texture patterns. Data augmentation improved generalization, reducing overfitting by approximately 5%.

j) Computational Complexity Analysis

The computational efficiency of the proposed model was evaluated in terms of inference time and model complexity. The ResNet-50 model contains approximately 25.6 million parameters and requires approximately 4.1 GFLOPs per inference.

On the experimental setup (NVIDIA RTX 3060 GPU), the average inference time per image was approximately 18 ms, enabling near real-time performance. This demonstrates the feasibility of deploying the model in practical agricultural monitoring systems, although further optimization is required for low-power edge devices.

5. Discussion and Future Work

The experimental results demonstrate that the proposed CNN-based framework is effective for automated pest identification under real-world agricultural conditions. The achieved accuracy of 82.1% highlights the capability of deep learning models, particularly when combined with transfer learning, to extract meaningful and discriminative features from complex field images. Compared to traditional feature-based approaches, the proposed method shows improved robustness against variations in lighting, background clutter, and pest morphology.

A key strength of this work lies in the use of the ResNet-50 architecture, which enables efficient training of deep networks through residual learning [2]. The incorporation of data augmentation techniques further enhanced model generalization by simulating real-world variations, thereby reducing overfitting and improving performance on unseen data. Additionally, the balanced performance across precision, recall, and F1score indicates that the model maintains consistency across different pest classes.

Despite these promising outcomes, certain limitations remain. The model exhibits difficulty in distinguishing between visually similar pest species, particularly during early developmental stages such as larvae and pupae. This issue is commonly observed in fine-grained image classification tasks and suggests the need for more advanced feature extraction mechanisms. Furthermore, the dependence on large labeled datasets, such as IP102 [1], poses a challenge for extending the model to less-represented pest species or new agricultural regions.

Another limitation is related to deployment. Although the current model achieves high accuracy, its computational complexity may hinder real-time implementation on low-power devices commonly used in agricultural fields. Recent studies have explored lightweight architectures such as MobileNet and EfficientNet-Lite for edge deployment, which offer a trade-off between accuracy and efficiency [5], [13]. Integrating such models into the proposed framework could improve practical applicability.
Future research directions can address these challenges in several ways. First, incorporating attention mechanisms and transformer-based architectures may enhance the model's ability to focus on fine-grained features and improve classification accuracy for visually similar pest species [6], [7]. Second, the integration of multimodal data, such as hyperspectral or thermal imagery, can provide additional discriminative information and improve robustness under varying environmental conditions [11].

Moreover, the development of IoT-based pest monitoring systems can enable real-time data collection and automated decision-making in smart farming environments [9], [12]. By combining deep learning models with sensor networks and mobile applications, it is possible to create an end-to-end system for continuous pest surveillance and early intervention. In addition, techniques such as domain adaptation and few-shot learning can be explored to reduce the dependency on large labeled datasets and improve model generalization across different geographical regions and crop types. These approaches are particularly important for scaling pest identification systems to diverse agricultural settings.

In summary, while the proposed framework demonstrates strong performance and practical potential, further research is required to enhance model efficiency, improve fine-grained classification, and enable real-time deployment. Addressing these challenges will be essential for developing scalable and intelligent pest management systems that support sustainable agriculture.

6. Conclusion

This study demonstrated the feasibility of a transfer learning-based CNN framework for automated pest identification using the IP102 dataset, achieving an accuracy of 82.1% with balanced evaluation metrics. The results confirm that deep learning models can effectively capture discriminative pest features under diverse environmental conditions.

However, certain limitations remain, particularly in fine-grained classification of visually similar pest species and real-time deployment on resource-constrained devices. These challenges highlight the need for further optimization and model efficiency improvements.

Future work will focus on lightweight architectures, multimodal data integration, and improved generalization techniques to enhance scalability and real-world applicability in precision agriculture systems. These findings contribute toward the development of reliable and scalable AI-driven pest monitoring systems.

References

[1] J. Chen, Z. Liu, H. Wang, and J. Chen, "IP102: A large-scale benchmark dataset for insect pest recognition," in *Proc. IEEE CVPR Workshops*, 2019.

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.

[3] Z. Liu, J. Chen, and H. Wang, "Fine-grained pest classification using deep CNNs," *IEEE Access*, vol. 8, pp. 145614–145623, 2020.

[4] Y. Gao, X. Zhou, and H. Lin, "Hybrid CNN-transformer network for insect pest recognition," *Computers and Electronics in Agriculture*, vol. 196, 2022.

[5] M. Rahman and F. Akter, "Lightweight deep learning for real-time pest detection on edge devices," *IEEE Access*, vol. 11, pp. 105203–105215, 2023.

[6] H. Zhao, Y. Li, and L. Wang, "Attention-based deep learning for finegrained insect pest recognition," *Applied Soft Computing*, vol. 107, 2021.

[7] X. Fang, T. Xu, and J. Zhou, "Vision transformer-based multiscale feature fusion for crop pest classification," *Expert Systems with Applications*, vol. 203, 2022.

[8] P. Wu, Q. Zhang, and Y. Liu, "GAN-based data augmentation for pest image classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–10, 2021.

[9] C. Jiang, M. Xu, and T. Chen, "Edge-enabled deep learning framework for smart pest detection in precision

agriculture," *IEEE Internet of Things Journal*, vol. 10, no. 9, pp. 7651–7663, 2023.

[10] S. Ali, A. Ahmed, and R. Hussain, "A comprehensive survey on deep learning in precision agriculture: Applications and challenges," *IEEE Access*, vol. 10, pp. 120658–120682, 2022.

[11] D. Lee and K. Park, "Multimodal pest identification using CNN and hyperspectral data fusion," *Computers and Electronics in Agriculture*, vol. 221, 2024.

[12] R. Singh and S. Sharma, "IoT-assisted pest monitoring and forecasting using deep learning," *IEEE Sensors Journal*, vol. 23, no. 7, pp. 9121– 9130, 2023.

[13] X. Wang, J. Dong, and Z. Hu, "EfficientNet-lite-based lightweight model for field pest recognition on mobile platforms," *IEEE Access*, vol. 12, pp. 37541–37553, 2024.

[14] L. Thomas, A. Kumar, and P. Bhattacharya, "AI-driven pest management for sustainable agriculture: Toward SDG-2 implementation," *Sustainable Computing: Informatics and Systems*, vol. 38, 2025.

[15] J. Patel and K. Joshi, "Review of convolutional neural network applications in crop and pest image classification," *Artificial Intelligence in Agriculture*, vol. 6, pp. 10–25, 2021.

[16] X. Xie, M. Wang, and S. Zhang, "Identification of agricultural pests based on color and shape features," Computers and Electronics in Agriculture, vol. 119, pp. 1–8, 2015.

[17] W. Ding and G. Taylor, "Automatic moth detection from trap images for pest management", Computers and Electronics in Agriculture, vol. 123, pp. 17–28, 2016.

[18] S. P. Mohanty, D. P. Hughes, and M. Salathe, "Using Deep Learning´ for Image-Based Plant Disease Detection," Frontiers in Plant Science, vol. 7, 2016.

[19] H. Zhao, Y. Li, and L. Wang, "Attention-Based Deep Learning for FineGrained Insect Pest Recognition," Applied Soft Computing, vol. 107, 2021.