2023 26th International Conference on Computer and Information Technology (ICCIT)
13-15 December, Cox's Bazar, Bangladesh

Revolutionizing Pest Detection for Sustainable Agriculture: A Transfer Learning Fusion Network with Attention-Triplet and Multi-Layer Ensemble

Nazmul Haque[1], Anwar Hossain Efat [2], S. M. Mahedy Hasan[3],Nahrin Jannat [4], Mahjabin Oishe [5], Mostarina Mitu[6]
[1,2,3,4,5,6]*Department of Computer Science & Engineering*
Rajshahi University of Engineering & Technology, Rajshahi-6204, Bangladesh
Emails: nazmulruetcse18@gmail.com,anwarhossainefat@gmail.com, mahedy@cse.ruet.ac.bd,
nahrinnipun17@gmail.com, mahjabinoishe@gmail.com,mostarinabinre@gmail.com

***Abstract*—Pests, insidious and destructive, pose a significant threat to agriculture, stored goods, and the environment. Detecting and mitigating their impact is crucial for safeguarding crops, food safety, and economic stability. Automated detection systems have emerged as critical tools, offering rapid and precise identification for timely interventions. However, existing solutions grapple with limitations, such as extensive labeled datasets and struggles in adapting to new pest species. This study aims to overcome these challenges, developing a robust pest recognition system setting new standards in accuracy and adaptability. Meticulous dataset collection and preprocessing form the foundation for accurate pest recognition. Leveraging pre-trained models and fine-tuning contribute significantly. Integrating Transfer Learning Fusion (TLF) harmoniously blends deep features, and Multi-Layer Ensemble (MLE) techniques enhance performance, achieving 98% accuracy. Our approach resulted in a model with the highest accuracy and precision in pest recognition, surpassing previous limitations. Our work signifies progress in agriculture, emphasizing the importance of automated pest detection and providing a superior solution. Contributions mark a significant stride in safeguarding crops, preserving food safety, and bolstering agricultural sustainability.**

***Index Terms*—Pest Detection, Transfer Learning Fusion (TLF), Attention-Triplet (AT) and Multi-Layer Ensemble (MLE).**

I. Introduction

In agriculture, pests, harmful organisms affecting crops, stored goods, and the environment, lead to financial losses and compromise food safety by targeting various plant parts. They can be categorized based on their damage focus into three types [1].

The Food and Agriculture Organization (FAO) reports annual pest-related crop losses of 20-40%, amounting to a substantial $220 billion global economic impact [2]. Although pesticide use is common, it carries environmental and health risks, including cancer, respiratory problems, genetic disorders, and fetal harm [3]. This research aims to establish an automated pest detection system for early identification and targeted pest management.

Previous research focused on automating pest classification, encountering challenges such as extensive labeled data needs, overfitting risks, and adaptation difficulties to new pests. Additionally, hurdles included substantial computational requirements, resource-intensive training, model interpretability, and minimizing false results. This study successfully overcame these limitations.

Beyond addressing these limitations, research inquiries (RIs) are posed to enhance the study's quality:

RI1: ***Examining the feasibility and impact magnitude of integrating a pre-trained model with a customized Convolutional Neural Network (CNN) model.***

RI2: ***Identifying methodologies to emphasize crucial features, particularly within specific regions or areas.***

RI3: ***Evaluating the advantages of classifying pests using a single classifier or a consortium of multiple classifiers to enhance overall effectiveness.***

These inquiries are extensively investigated, leading to several contributions:

1) Amalgamating a pre-existing model with a tailored CNN model, resulting in an enhanced transfer learning model.
2) Harnessing multiple pre-trained models, combining their performances for superior outcomes and broader applicability.
3) Introducing a novel approach, the Attention-Triplet (AT) method, within the proposed custom transfer learning model, designed to assign greater significance to pivotal features.
4) Enhancing model robustness, precision, and adaptability across diverse datasets through the incorporation of ensemble techniques at various levels, termed the Multi-Layer Ensemble (MLE).

Our paper follows this structure: Section II reviews the literature. Section III presents the dataset overview, and Sections IV-A to IV-E provide detailed information on our research approach, data preprocessing, Architecture of Proposed Transfer Learning Fusion (TLF), Attention-Triplet (AT), Multi-Layer Ensemble Approach (MLE), and Justification of Our Procedural Architecture. Section V analyzes experimental results, while Sections V-A to V-D provide detailed experimental results cover Evaluation Metrics, Experimental Setup, Result Assessment, and Research Question Answers. Sections VI and

VII contain discussions, comparisons, and considerations of potential result validity threats. Finally, in Section VIII, we summarize findings and suggest future research directions, followed by references.

II. Literature Review

Various research endeavors have employed diverse models in their investigations. For instance, certain studies utilized Convolutional Neural Network (CNN) models, as evidenced by references [4], [5], [6], and [7]. Meanwhile, reference [4] applied the Inception-v3 model, and [8] and [9] adopted a blend of Inception-v3, ResNet, GoogleNet, and VGG. Conversely, [10] and [11] opted for YOLO. Additionally, [12] and [7] incorporated Artificial Neural Networks (ANN), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Naive Bayes (NB) in their investigations.

Souza et al. [12] achieved 97.00% accuracy using the Inception-V3* model for insect pest classification. Thenmozhi and Reddy [4] obtained accuracies of 96.75%, 97.47%, and 95.97% for NBAIR, Xie1, and Xie2 datasets. TÜRKOĞLU and HANBAY [8] examined nine deep learning networks, with GoogleNet reaching a 95.22% accuracy with SVM and ResNet50 achieving 97.86% with SVM. Other models, like ResNet101, InceptionV3, InceptionResNetV2, and SqueezeNet, also exhibited strong performance.

Deng et al. [13] achieved an 85.50% accuracy using SIFT-HMAX features for insect pest recognition. Yanfen Li et al. [5] utilized deep CNNs and GoogleNet to attain a 94.61% accuracy in crop pest identification. Jin Wang et al. [9] introduced the CPAF dataset and obtained a 92.26% recognition accuracy with CPAFNet.

Zhankui Yang et al. [6] achieved 91.64% accuracy using an enhanced SqueezeNet for insect recognition in natural settings. Min Dai et al. [10] introduced a pest detection approach with YOLOV5m, showcasing high precision, recall, F1 score, and mAP values. Kasinathan et al. [7] employed machine learning techniques, with the CNN model achieving classification rates of 91.50% and 90.00% for nine and 24 insect classes, respectively. Kuzuhara et al. [11] employed YOLOv3 and Xception for insect pest detection and re-identification, benefiting from an effective data augmentation strategy.

Collectively, these studies furnish valuable insights and methodologies for the recognition and categorization of insect pests, thus contributing significantly to advancements in pest control and agricultural practices.

III. Dataset Description

Table I provides a concise overview of our ”pest” dataset, which is readily accessible on Kaggle [14]. Figure 1 illustrates the dataset’s representation in terms of samples for each class.

TABLE I
SUMMARY OF THE DATASET

No of Images	Format	No of Classes	Source
3150	JPEG	9	kaggle.com

IV. Methodology

Methodology begins with dataset acquisition, followed by vital preprocessing. ”Transfer Learning Fusion (TLF)” involves feeding input images into pre-trained models, reshaping tensors, and fine-tuning. Two convolution blocks with varying kernel sizes, filters (128 and 256), ’BatchNormalization,’ MaxPooling2D, and Attention layer were used. ReLU activation functions applied. Final max-pooling output flattened, passing through three fully connected layers. Softmax activation predicted class probabilities. ’Multi-Layer Ensemble (MLE)’ introduced for prediction with Majority Voting, Softmax Averaging, and Weighted Averaging. The process is depicted in Figure 3.

The workflow of our research is illustrated in Figure 2.

A. Data Preprocessing

Preprocessing, vital for data and feature selection ,emphasized by [15] and [16], optimized algorithm precision. The dataset initially divided into training and test sets, with the training set further partitioned into 90% for training and 10% for validation. No Data Augmentation implemented due to well-balanced class distribution and sufficient remaining images for effective training and model assessment.

B. Architecture of Proposed Transfer Learning Fusion (TLF)

Our approach initiates by feeding input images into pre-trained models, followed by tensor reshaping for fine-tuning. Two convolution blocks (CBs) with 128 and 256 filters each are incorporated, featuring varying kernel sizes (7x7, 5x5, 3x3, and 1x1) and four ’BatchNormalization’ layers. Each CB concludes with a MaxPooling2D layer and an Attention layer for robustness, using ReLU activation functions on all convolution layers to mitigate the vanishing gradient problem.

The final max-pooling layer’s output is flattened and passed through three fully connected layers (512, 256, and 9 neurons) with ReLU activation functions. The last layer employs a softmax activation function for predicting class probabilities.

Next, we introduce ’Multi-Layer Ensemble (MLE)’ with Majority Voting (MV), Softmax Averaging (SAvg), and Weighted Averaging (WAvg) in the initial prediction phase. The best-performing ensemble progresses to subsequent layers.

Pre-trained models, including DenseNet, ResNet, and Xception, are strategically selected for training via different architecture types. The architectural overview is depicted in Figure 3.

C. Attention-Triplet (AT)

Attention modules serve the purpose of selectively directing the network’s focus toward significant input features while disregarding irrelevant ones.

1) Channel Attention (CA):: CA improves feature maps by computing channel-wise attention weights.

$$\mathbf{w_c} = \sigma\big(\mathbf{W_2}\delta\big(\mathbf{W_1 x}\big)\big) \tag{1}$$

$$\mathbf{y_c} = \mathbf{w_c} \odot \mathbf{x} \tag{2}$$

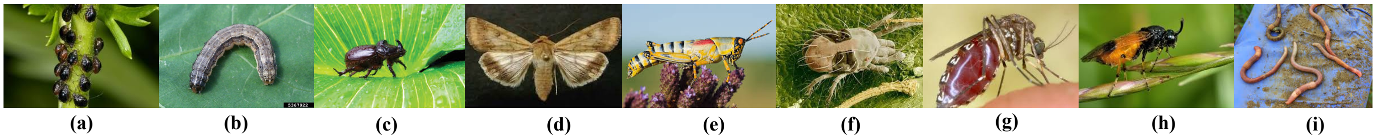

Fig. 1. Sample images for each classes.

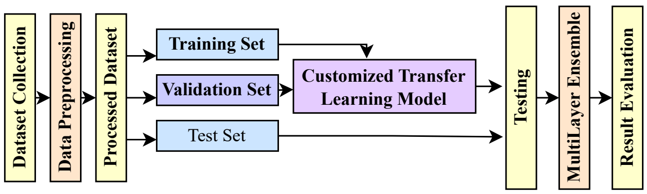

Fig. 2. Sequential workflow of the proposed methodology

where, $\mathbf{x}$ = the input feature maps ($C \times H \times W$). $\mathbf{W_1}$ and $\mathbf{W_2}$ = weight matrices. δ = ReLU. σ = Sigmoid. $\mathbf{w_c}$ = attention weights. $\odot$ = Element-wise multiplication.

2) Soft Attention (SA):: The SA module operates by concentrating on specific portions of the input, assigning weights to each element to determine their relative significance.

$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^{T} \exp(e_j)} \tag{3}$$

where α_i = Attention weight. T = Length of the input sequence. e_i = Scalar value.

3) Squeeze and Excitation Attention (SEA):: The SEA module includes two essential actions: squeezing to reduce spatial dimensions in input feature maps and exciting to learn channel-specific attention weights, focusing on vital features.

Let $\mathbf{x}$ be the input feature maps of size $C \times H \times W$.

$$\mathbf{z} = \text{GlobalAveragePooling}(\mathbf{x}) \tag{4}$$

$$\mathbf{s} = \text{ReLU}(\mathbf{W}_2 \text{sigmoid}(\mathbf{W}_1 \mathbf{z})) \tag{5}$$

$$\mathbf{y} = \mathbf{s} \odot \mathbf{x} \tag{6}$$

where $\mathbf{W}_1 \in \mathbb{R}^{C/r \times C}$ and $\mathbf{W}_2 \in \mathbb{R}^{C \times C/r}$ = learnable weight matrices.

D. Multi-Layer Ensemble (MLE)

Ensembling, a technique that combines multiple models to enhance predictive performance, boost model stability, reduce overfitting, and enhance robustness, played a pivotal role in our study. We employed methodologies such as Majority Voting (MV), Softmax Averaging (SAvg), and Weighted Averaging (WAvg) in the initial prediction layer. The best-performing ensemble approach advanced to subsequent layers, creating a cascading effect. The process for Multi-Layer Ensemble (MLE) is thoughtfully elucidated in Figure 4, providing a visual representation of our approach's intricacies.

E. Justification of Our Procedural Architecture

In CNN, applying batch normalization addresses internal covariate shift, ensuring stable input distribution to each layer during training. This integration expedites training, promotes robust convergence, and enhances overall performance.

Utilizing diverse attention modules yields independent feature sets, reinforcing various regions and extracting diverse feature types. CA heightens the relevance of pivotal features, SEA optimizes computational efficiency, and SA affords insights for model interpretation.

Multi-Layer Ensemble methodology amalgamates architectural strengths, resulting in heightened prediction accuracy.

V. Performance Analysis

A. Performance Evaluation Metrics

We have employed a range of metrics to assess the effectiveness of our models, encompassing accuracy, precision, recall (sensitivity), f1-score, specificity, and the Area Under the Receiver Operating Characteristic Curve (ROC-AUC). Below, we present the mathematical expressions that represent these measures. The mathematical representations of these metrics are provided as follows [17], [18].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{10}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{11}$$

B. Experimental setup

We performed architectural operations on Kaggle using a GPU P100 and a 2-core Intel Xeon CPU (690 ms/step). Input images were (224,224,3), with separate training and testing folders. We split the training dataset 0.85 for training and 0.15 for validation, with no holdout set. Models trained for 50 epochs (batch size 16) with Adam optimizer (lr: 0.001, loss: categorical cross-entropy). Early stopping used Reduce on Plateau (patience: 25).

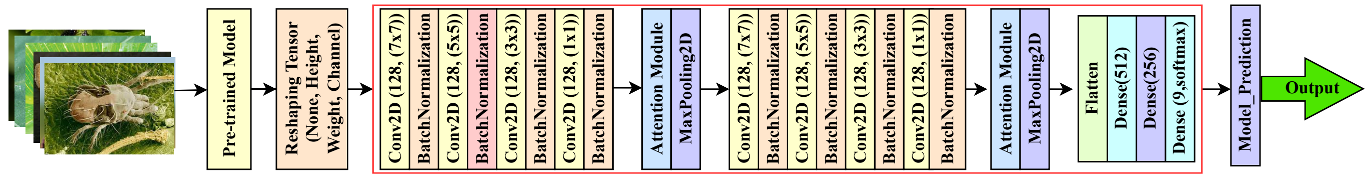

Fig. 3. Architectural view of the proposed methodology

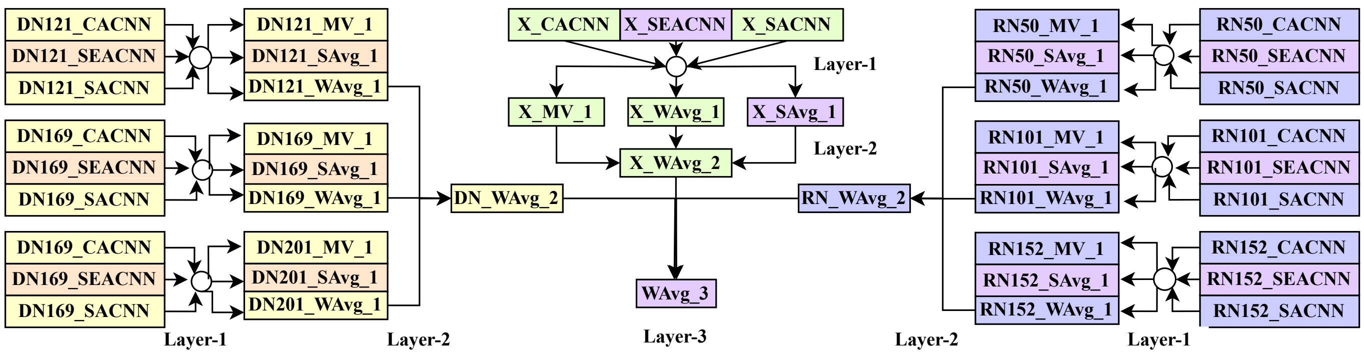

Fig. 4. Multi-Layer Ensemble Approach

C. Results Assessment

Table II presents the performance results obtained from our TLF architectures and MLE approaches. It is evident that the utilization of the Attention-triplet within the TLF model has significantly improved performance across various metrics.

Our study employed three core architectures: DenseNet, ResNet, and Xception. In the case of DenseNet, we explored three variants—DenseNet121, DenseNet169, and DenseNet201. Similarly, for ResNet, we examined three variants—ResNet50, ResNet101, and ResNet152. However, Xception, being unique, had no additional variants. Each of these architectures was seamlessly integrated with a customized CNN model, enriched with the attention triplet mechanism.

DenseNet121-based models excelled, with channel attention CNN achieving 97.78% accuracy. First-layer ensemble brought WAvg to the same accuracy. DenseNet169 saw soft attention-based CNN lead at 97.11%. After first-layer ensemble, MV, SAvg, and WAvg improved to 97.56%. DenseNet201's channel attention-based CNN reached 97.77%. WAvg maintained this accuracy, and a second-layer ensemble improved it to 97.78%.

ResNet50's soft attention-based CNN scored 97.56%, and MV, SAvg, and WAvg matched this accuracy post first-layer ensemble. ResNet101, with soft attention-based CNN, reached 97.78%. WAvg maintained this accuracy. ResNet152's Squeeze and Excitation attention triplet-based CNN led at 97.33%. WAvg reached the same accuracy after first-layer ensemble, improving to 97.88% with a second-layer ensemble.

Xception, without variants, combined with attention triplet-based and Squeeze and Excitation triplet-based CNNs for the highest accuracy at 96.89%. MV and SAvg reached 97.33% after first-layer ensemble.

Finally, after a third-layer ensemble using the three best WAvg models, the accuracy improved to 98.00%. The confusion matrix of the final model is depicted in Figure 5, which is also in agreement with the results.

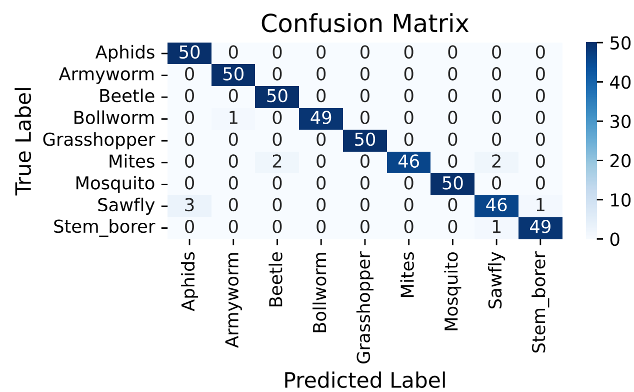

Fig. 5. Confusion Matrix of the proposed model

However, the ROC-AUC curve in Figure 6 has been observed graphically to assess the convergence of the architecture which is actually the ideal scenario based on [19], [20]. It is evident that the fluctuations in the true positive rate and false positive rate are minimal, forming nearly straight lines. This indicates that our model has achieved a stable convergence.

D. Answers to The Research Inquiries

Answer to RI1: *Our research focused on integrating a pre-trained model with a customized CNN. Experiment results unequivocally show the substantial benefits. Fine-tuning*

TABLE II
EXPERIMENTAL RESULTS ASSESSMENT

Model	Accuracy	Precision	Recall	F1-score
DN121-CACNN	97.78	97.81	97.78	97.77
DN121-SEACNN	96.67	96.77	96.67	96.64
DN121-SACNN	96.67	96.81	96.67	96.63
MV (1)	97.33	97.44	97.33	97.31
SAvg (1)	97.33	97.44	97.33	97.31
WAvg (1)	97.74	97.71	97.74	97.72
DN169-CACNN	96.44	96.56	96.44	96.43
DN169-SEACNN	96.89	96.99	96.89	96.87
DN169-SACNN	97.11	97.32	97.11	97.10
MV (1)	97.56	97.66	97.56	97.54
SAvg (1)	97.56	97.66	97.56	97.54
WAvg (1)	97.56	97.66	97.56	97.54
DN201-CACNN	97.77	97.81	97.77	97.75
DN201-SEACNN	96.22	96.49	96.22	96.20
DN201-SACNN	96.67	96.76	96.67	96.64
MV (1)	97.26	97.11	97.11	97.09
SAvg (1)	97.11	97.26	97.11	97.09
WAvg (1)	97.77	97.81	97.77	97.75
DN_WAvg (2)	97.78	97.83	97.78	97.76
RN50-CACNN	97.11	97.15	97.11	97.08
RN50-SEACNN	95.78	95.97	95.78	95.78
RN50-SACNN	97.56	97.73	97.56	97.57
MV (1)	97.78	97.87	97.78	97.77
SAvg (1)	97.78	97.87	97.78	97.77
WAvg (1)	97.78	97.80	97.78	97.77
RN101-CACNN	97.11	97.17	97.11	97.09
RN101-SEACNN	96.22	96.35	96.22	96.19
RN101-SACNN	97.78	97.82	97.78	97.77
MV (1)	97.56	97.59	97.56	97.54
SAvg (1)	97.56	97.59	97.56	97.54
WAvg (1)	97.78	97.82	97.78	97.77
RN152-CACNN	96.67	96.80	96.67	96.63
RN152-SEACNN	97.33	97.38	97.33	97.30
RN152-SACNN	96.67	96.84	96.67	96.61
MV (1)	97.56	97.64	97.56	97.52
SAvg (1)	97.56	97.64	97.56	97.52
WAvg (1)	97.78	97.83	97.78	97.75
RN_WAvg (2)	97.88	97.90	97.88	97.76
X-CACNN	96.67	96.82	96.67	96.64
X-SEACNN	96.89	96.96	96.89	96.85
X-SACNN	96.67	96.87	96.67	96.66
MV (1)	97.33	97.40	97.33	97.31
SAvg (1)	97.33	97.40	97. 33	97.31
WAvg (1)	97.11	97.17	97.11	97.07
X_WAvg (2)	97.33	97.38	97.33	97.30
WAvg (3)	98.00	98.06	98.00	97.99

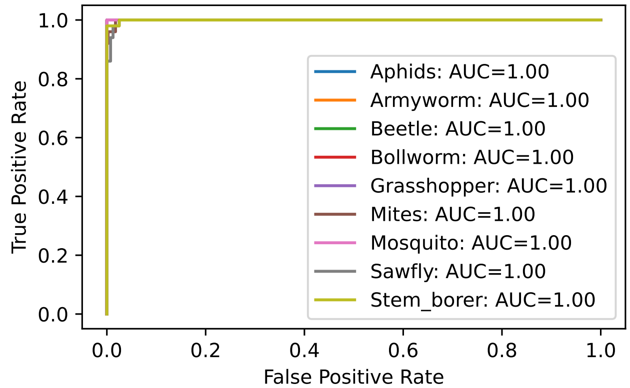

Fig. 6. ROC-AUC Curve of the proposed model

consistently yielded superior performance, enhancing overall model effectiveness and accuracy.

Answer to RI2: *To address emphasizing crucial features, especially in specific regions, we introduced the "Attention Triplet." This innovative approach incorporates three attention modules—CA, SEA, and SA—into our CNN. These modules enable prioritization of essential features, disregarding less relevant information. Meticulous experimentation has demonstrated the Attention Triplet's remarkable success in enhancing feature selection and overall model performance, providing a robust solution to emphasizing critical features within specific regions.*

Answer to RI3: *Our research highlights the advantage of using an ensemble of multiple classifiers over a single classifier for pest classification. We designed a custom CNN with three attention mechanisms—CA, SEA, and SA—allowing each to capture diverse pest features. Testing showed autonomous identification of unseen data by each model. Employing diverse ensemble strategies, our findings unequivocally demonstrate that this approach significantly enhances results compared to relying on a single classifier, consistently improving accuracy.*

VI. DISCUSSION AND EXTENDED COMPARISON

Table III can now be enhanced with a comprehensive comparison, illustrating that our final results significantly surpass previous researches. It's important to note that even though Turkoglu and Hanbay [8] achieved a higher accuracy of 97.86% in prior work, they only utilized a small subset of the dataset with complex algorithms. In contrast, we employed a dataset comprising 3150 images, affirming the superiority of our model in this context. Our less complex yet advanced and user-friendly model has proven its efficacy.

VII. THREATS TO VALIDITY

Our study has limitations in dataset generalization. The pest classification model, constructed from a single dataset, may be less effective in diverse environments. We solely analyzed pest images, neglecting accompanying metadata. Integrating this metadata in future work holds potential to boost the prediction model's accuracy and performance.

TABLE III
PERFORMANCE COMPARISON

Article	Accuracy	Precision	Recall	F1-score	Specificity
[12]	97.00	-	-	-	-
[4]	97.47	-	-	-	-
[8]	97.86	-	-	97.14	-
[13]	86.9	-	-	-	-
[5]	94.61	-	-	-	-
[9]	92.26	-	-	-	-
[6]	91.64	-	-	-	-
[10]	96.4%	95.7	93.1	94.38	-
[7]	91.5	-	-	-	-
[11]	78%	-	-	-	-
Ours	**98.00**	**98.06**	**98.00**	**97.99**	**99.72**

VIII. CONCLUSION

The term ”pest” encompasses a spectrum of agricultural threats, ranging from insects to pathogens, capable of causing extensive damage to crops, stored goods, and the ecosystem. Timely and accurate detection of these threats is crucial, directly influencing food production, economic stability, and environmental well-being.The advent of automated pest detection systems has revolutionized agriculture, offering swift and precise identification to manage pests proactively, mitigating potential damage. Despite their promise, existing solutions face significant limitations, including the need for extensive labeled data, susceptibility to overfitting, and challenges in adapting to emerging pest species and variations.Our mission was clear: transcend these limitations and engineer a pest recognition system redefining standards of accuracy and adaptability. A cornerstone in our journey involved assembling and refining a comprehensive dataset, a prerequisite for precision. By harnessing pre-trained models, we injected domain expertise, finely tuning them for nuanced pest recognition. Innovative Transfer Learning Fusion (TLF) brought a harmonious amalgamation of deep features, elevating discernment. Incorporating Multi-Layer Ensemble (MLE) techniques, including Majority Voting, Softmax Averaging, and Weighted Averaging, added finesse to performance.Our journey culminated in creating a model reaching unparalleled heights in precision and accuracy in pest recognition. Surmounting past limitations, our work signifies a leap forward in agriculture. In summary, our research underscores the pivotal role of automated pest detection in modern agriculture, providing a tangible and impactful solution surpassing existing methods. Our contributions advance crop protection and significantly contribute to the sustainability and resilience of agricultural practices in a dynamic world.

REFERENCES

[1] N. Jannat, S. M. Hasan, A. Y. Srizon, M. Oishe, A. H. Efat, M. F. Taraque, M. Mitu, and M. F. Faruk, “Efficient detection of crop leaf diseases: A lightweight convolutional neural network approach for enhanced agricultural productivity,” in *2023 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*. IEEE, 2023, pp. 99–103.

[2] “Fao - news article: new standards to curb the global spread of plant pests and diseases,” https://www.fao.org/news/story/en/item/1187738/icode/, (Accessed on 09/30/2023).

[3] S. Narenderan, S. Meyyanathan, and B. Babu, “Review of pesticide residue analysis in fruits and vegetables. pre-treatment, extraction and detection techniques,” *Food Research International*, vol. 133, p. 109141, 2020.

[4] 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.

[5] Y. Li, H. Wang, L. M. Dang, A. Sadeghi-Niaraki, and H. Moon, “Crop pest recognition in natural scenes using convolutional neural networks,” *Computers and Electronics in Agriculture*, vol. 169, p. 105174, 2020.

[6] Z. Yang, X. Yang, M. Li, and W. Li, “Automated garden-insect recognition using improved lightweight convolution network,” *Information Processing in Agriculture*, 2021.

[7] 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.

[8] M. Türkoğlu and D. Hanbay, “Plant disease and pest detection using deep learning-based features,” *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 27, no. 3, pp. 1636–1651, 2019.

[9] J. Wang, Y. Li, H. Feng, L. Ren, X. Du, and J. Wu, “Common pests image recognition based on deep convolutional neural network,” *Computers and Electronics in Agriculture*, vol. 179, p. 105834, 2020.

[10] M. Dai, M. M. H. Dorjoy, H. Miao, and S. Zhang, “A new pest detection method based on improved yolov5m,” *Insects*, vol. 14, no. 1, p. 54, 2023.

[11] H. Kuzuhara, H. Takimoto, Y. Sato, and A. Kanagawa, “Insect pest detection and identification method based on deep learning for realizing a pest control system,” in *2020 59th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*. IEEE, 2020, pp. 709–714.

[12] W. S. Souza, A. N. Alves, and D. L. Borges, “A deep learning model for recognition of pest insects in maize plantations,” in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. IEEE, 2019, pp. 2285–2290.

[13] L. Nanni, G. Maguolo, and F. Pancino, “Insect pest image detection and recognition based on bio-inspired methods,” *Ecological Informatics*, vol. 57, p. 101089, 2020.

[14] “Pest dataset,” https://www.kaggle.com/datasets/simranvolunesia/pest-dataset/data, (Accessed on 09/30/2023).

[15] S. Datta, S. M. Hasan, M. F. Faruk, A. Y. Srizon, M. R. Hossain, A. H. Efat, and M. Mitu, “Improved diabetes prediction with reduced feature sets: Evaluating feature selection techniques in machine learning,” in *2023 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*. IEEE, 2023, pp. 104–108.

[16] S. Datta, S. M. Hasan, M. Mitu, M. F. Taraque, N. Jannat, and A. H. Efat, “Hyperparameter-tuned machine learning models for complex medical datasets classification,” in *2023 International Conference on Electrical, Computer and Communication Engineering (ECCE)*. IEEE, 2023, pp. 1–6.

[17] M. Mitu, S. M. Hasan, A. H. Efat, M. F. Taraque, N. Jannat, and M. Oishe, “An explainable machine learning framework for multiple medical datasets classification,” in *2023 International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM)*. IEEE, 2023, pp. 1–6.

[18] N. Jannat, S. M. Hasan, A. H. Efat, M. F. Taraque, M. Mitu, M. Al Mamun, and M. F. Faruk, “Stacking ensemble technique for multiple medical datasets classification: A generalized prediction model,” in *2023 International Conference on Electrical, Computer and Communication Engineering (ECCE)*. IEEE, 2023, pp. 1–6.

[19] A. H. Efat, S. M. Hasant, N. Jannat, M. Mitu, M. F. Taraque, S. Ferdous, and S. Datta, “Inquisition of the support vector machine classifier in association with hyper-parameter tuning: A disease prognostication model,” in *2022 4th International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)*. IEEE, 2022, pp. 131–134.

[20] M. F. Taraque, S. M. Hasan, N. Jannat, A. H. Efat, M. Mitu, S. Datta, and S. Ferdous, “Early stage prediction of autism spectrum disorder: Analyzing different hyperparameter tuned machine learning classifier,” in *2022 4th International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)*. IEEE, 2022, pp. 1–4.