Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi
https://ojs.stmik-banjarbaru.ac.id/index.php/jutisi/index
Jl. Ahmad Yani, K.M. 33,5 - Kampus STMIK Banjarbaru
Loktabat – Banjarbaru (Tlp. 0511 4782881), e-mail: puslit.stmikbjb@gmail.com
e-ISSN: 2685-0893

Classification of Chili Plant Pests Using the ConvNeXt Architecture

DOI: http://dx.doi.org/10.35889/jutisi.v15i1.3461

Jennifer Jocelyn[1*], Siska Devella[2]
Informatics, Universitas Multi Data Palembang, Palembang, Indonesia
*e-mail *Corresponding Author*. Jenniferjocelyn_2226250011@mhs.mdp.ac.id

Abstract

*Chili (*Capsicum annuum L.*) is a high-value horticultural commodity in Indonesia; however, its productivity often declines due to pest attacks that cause significant economic losses. This study aims to compare the performance of several ConvNeXt variants (V1 and V2) for chili pest classification using the Red Chili Pepper Pest dataset, which consists of four pest classes annotated with bounding boxes. The data were divided into training and testing sets, and a cropping process was applied to the object regions to ensure that the model focuses on pest images. The preprocessing stages included resizing, normalization, and data augmentation to improve model robustness against variations in image conditions. Model training was conducted using the* timm *library with uniform hyperparameter settings across all variants to ensure a fair comparison. Performance evaluation was carried out using accuracy, precision, recall, F1-score, and Area Under the Curve (AUC). In addition, computational complexity was analyzed in terms of the number of parameters, FLOPs, and inference latency. The results indicate that ConvNeXt V2 variants, particularly Nano and Tiny, achieve very high classification performance (macro-AUC > 0.99) with fewer parameters and lower latency compared to larger models. Robustness evaluation under various image degradations shows that Gaussian noise has the most significant negative impact on performance. Overall, ConvNeXt V2-Nano and V2-Tiny are recommended as the most efficient and stable models for implementing chili pest detection systems on resource-constrained devices within precision agriculture applications.*

Keywords: *Chili Pest Classification; ConvNeXt; Deep Learning; Image Processing; Smart Agriculture.*

Abstrak

Cabai (*Capsicum annuum L.*) merupakan komoditas hortikultura bernilai tinggi di Indonesia, namun produktivitasnya sering menurun akibat serangan hama yang menyebabkan kerugian ekonomi. Penelitian ini bertujuan membandingkan kinerja varian ConvNeXt (V1 dan V2) dalam klasifikasi hama cabai menggunakan dataset *Red Chili Pepper Pest* yang terdiri atas empat kelas hama dengan anotasi *bounding box*. Data dibagi menjadi data pelatihan dan pengujian, kemudian dilakukan proses cropping pada objek untuk memastikan model berfokus pada citra hama. Tahapan prapemrosesan meliputi resizing, normalisasi, dan augmentasi untuk meningkatkan ketahanan model terhadap variasi citra. Pelatihan model dilakukan menggunakan pustaka timm dengan pengaturan hiperparameter pada seluruh varian untuk menjamin perbandingan adil. Evaluasi dilakukan menggunakan akurasi, presisi, recall, F1-score, dan AUC, serta analisis kompleksitas melalui jumlah parameter, FLOPs, dan latensi inferensi. Hasil penelitian menunjukkan ConvNeXt V2, khususnya Nano dan Tiny, mencapai performa tinggi (macro-AUC > 0,99) dengan kompleksitas komputasi lebih rendah. Uji robustness menunjukkan Gaussian noise memberikan penurunan performa paling signifikan.

Kata kunci: *Klasifikasi Hama Cabai; ConvNeXt; Pembelajaran Mendalam; Pemrosesan Citra; Pertanian Cerdas.*

1. Introduction

The rapid development of digital technology in the era of globalization has significantly contributed to addressing various challenges across multiple sectors, including agriculture. The utilization of digital technologies in agriculture not only aims to improve production efficiency but also supports more accurate and data-driven decision-making processes [1]. In this context, the integration of intelligent technologies such as computer vision and deep learning has become increasingly important, particularly for complex agricultural challenges such as pest identification and management. Early and accurate pest classification is crucial for reducing crop losses, optimizing pest control strategies, and minimizing excessive pesticide use, thereby supporting sustainable agricultural practices and strengthening food security.

In the agricultural sector, horticultural crops play a vital role in supporting national food security and the community's economy, with chili pepper (*Capsicum annuum* L.) being one of the most widely cultivated high-value commodities in Indonesia [2], [3]. Chili has strategic importance as a primary raw material for various food products and shows relatively stable demand that continues to increase annually [4]. However, chili cultivation faces serious challenges due to pest infestations, which significantly reduce productivity and crop quality. Various pests such as aphids, thrips, fruit flies, and fruit caterpillars cause substantial damage to chili plants [5], with more than 60 insect pest species reported to attack chili crops [6], [7]. Identifying these pests accurately in real field conditions is challenging due to high morphological similarity, small object size, and varying lighting and background conditions. Conventional pest identification methods rely heavily on manual observation and expert knowledge, making them time-consuming, subjective, and prone to human error, thus highlighting the need for automated and reliable identification approaches.

Rapid and accurate pest identification is therefore essential to support effective and efficient pest control strategies. Digital image analysis using artificial neural networks, particularly convolutional neural networks (CNNs), has been widely applied for plant image classification tasks [8]. Previous studies have demonstrated the effectiveness of CNN architectures such as VGG, DenseNet, AlexNet, and SqueezeNet for chili leaf disease classification [9]–[13]. However, most existing studies focus on plant diseases rather than pest classification, while conventional CNNs still face limitations in handling high visual variability, subtle morphological differences, and image disturbances such as noise and illumination changes. To address these limitations, this study proposes an image-based chili pest classification approach using the ConvNeXt architecture, a modernized CNN that adopts Vision Transformer design principles while maintaining computational efficiency [14]. ConvNeXt has demonstrated superior performance across various visually complex domains, including medical image analysis [15]–[17], and its enhanced version, ConvNeXt V2, further improves generalization capability and robustness through co-design with masked autoencoders [18]. These characteristics provide strong justification for selecting ConvNeXt as an effective and efficient solution for chili pest classification under complex field conditions [19].

Despite the potential of ConvNeXt, no prior study has systematically compared multiple ConvNeXt V1 and V2 variants for chili pest classification. Therefore, this research evaluates various ConvNeXt models to identify architectures that are accurate, robust, and computationally efficient. The findings are expected to support faster and more reliable pest identification, improve pest management strategies, reduce crop losses, and contribute academically by addressing the research gap and providing empirical insights into the application of modern CNN architectures for smart agriculture systems, particularly on resource-constrained devices [13].

2. Literature Review

Various previous studies have examined the application of deep learning in agriculture, particularly to support image-based identification of plant pests and diseases. This approach is widely adopted due to its ability to automatically extract visual features and improve object recognition accuracy under diverse environmental conditions.

One of the key reference studies in the domain of chili pest analysis was conducted by [13]. The study developed a pest detection system for Indonesian red chili plants using a fine-tuned YOLOv5 model. The dataset was collected from chili plantations in Bengkulu Province and consisted of four major pest classes. The results showed that the model achieved a mean Average Precision (mAP) of 82.6% on the validation set and 81.3% on the test set, with inference speed suitable for real-time applications. Despite its strong performance, the study focused on object detection tasks and did not evaluate the model's capability for image-level pest classification on single images.

Other studies in the chili plant domain have predominantly focused on disease classification rather than pest identification. [11] developed a chili plant disease classification system using the DenseNet201 architecture. The dataset comprised five disease classes, including healthy conditions. The model achieved 90% accuracy on the training data and 84% accuracy on the test data. Although DenseNet demonstrated strong feature extraction capabilities, the observed performance drop on the test set indicates limitations of conventional CNNs in handling image variability and environmental noise.

A similar approach was adopted by [10] who compared AlexNet and SqueezeNet architectures for classifying chili pepper leaf diseases. Using 1,000 images and training for 32 epochs, the study showed that AlexNet achieved higher accuracy than SqueezeNet. Nevertheless, the results suggest that classification performance is highly dependent on architectural choice, and classical CNN models still face challenges in addressing the visual complexity of plant imagery.

The limitations of conventional CNNs have driven the development of modern CNN architectures. [14] introduced ConvNeXt as a modernization of ResNet-based CNNs by adopting design principles from Vision Transformers. ConvNeXt enhances global feature extraction through large kernel convolutions, Layer Normalization, and GELU activation functions. Experimental results demonstrated that ConvNeXt can match or even surpass Vision Transformer performance across various computer vision tasks, including image classification, while maintaining computational efficiency.

The development of ConvNeXt continued with the introduction of ConvNeXt V2 by [18]. This architecture integrates masked autoencoders and Global Response Normalization (GRN) to improve training stability and model generalization. ConvNeXt V2 has been shown to outperform the first-generation ConvNeXt, particularly on complex images and large-scale datasets.

The advantages of ConvNeXt have also been demonstrated in other visually challenging domains. [16] compared ConvNeXt Small and ResNet50 for medical CT image classification and reported superior accuracy achieved by ConvNeXt Small. In addition, [20] evaluated various backbone architectures for plant disease and pest classification under few-shot learning scenarios and found that ConvNeXt outperformed other CNN architectures.

In the context of chili plants, [21] developed a ConvNeXt-based model enhanced with multi-scale feature fusion and channel–spatial attention mechanisms for chili leaf disease classification. The proposed model achieved approximately 98% accuracy and outperformed standard ConvNeXt and other benchmark architectures. This study highlights the strong potential of ConvNeXt for agricultural applications, particularly for images containing small objects and complex visual details.

Based on previous studies, most research has focused on chili disease classification or pest detection using object detection frameworks, while image-based chili pest classification with modern CNN architectures remains limited. Moreover, no study has systematically compared multiple ConvNeXt variants from both v1 and v2 generations in terms of classification accuracy and computational efficiency. Therefore, this study aims to address this gap through a comprehensive comparison of ConvNeXt variants for chili pest classification.

3. Methodology

The methodology of this study consists of five sequential stages designed to systematically develop, train, and evaluate a robust chili pest classification system based on ConvNeXt architectures. The overall research workflow is illustrated in Figure 1, which outlines the progression from literature review to comprehensive performance and efficiency evaluation. This staged approach ensures that each phase of the research is conducted in a structured and reproducible manner, allowing objective comparison among different ConvNeXt variants.

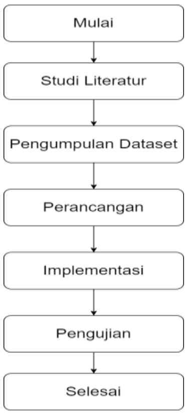


Fig. 1. Stages of Research Methodology

The first stage involved a comprehensive literature review to establish the theoretical foundation of the study. Previous works on pest and plant disease classification, convolutional neural network (CNN)-based models, and modernized convolutional architectures were examined. Particular attention was given to ConvNeXt v1, which modernizes traditional CNNs through transformer-inspired design principles [14], and ConvNeXt v2, which further enhances feature learning and generalization using masked autoencoders and Global Response Normalization [18]. In addition, studies focusing on chili pest analysis, especially object detection–based approaches such as [13], were reviewed. This review revealed a research gap in image-level chili pest classification, particularly regarding systematic comparisons between ConvNeXt v1 and v2 variants.

The second stage focused on dataset acquisition and preparation. The Red Chili Pepper Pest Dataset, consisting of 4,994 images across four pest classes—*Green peach aphid*, *Silverleaf whitefly*, *Thrips*, and *Caterpillar*—was obtained from a publicly released dataset by [13]. The images were collected under real-field conditions, making them representative of practical agricultural environments. Since the dataset was originally annotated in YOLO object detection format, it was not directly suitable for image classification. Therefore, all images were cropped using bounding box coordinates following the YOLO annotation scheme described by [22]. This process isolated individual pest objects and converted the dataset into a classification-ready format. The processed dataset was then split into training (70%), validation (20%), and testing (10%) subsets.

In the third stage, data preprocessing and system design were conducted to enhance model robustness and generalization. Each cropped image was resized to 224×224 pixels, normalized, and augmented using random cropping, horizontal flipping, color jitter (brightness and contrast adjustment), Gaussian blur, and random erasing. These augmentations were applied to increase data diversity and mitigate overfitting caused by variations in pest size, texture, and illumination. To address class imbalance, a class-weighted cross-entropy loss was applied to prevent bias toward majority classes. Multiple ConvNeXt variants were implemented, including v1 (Tiny, Small, Base) and v2 (Atto, Femto, Pico, Nano, Tiny, Base), enabling systematic comparison across different model capacities.

In the fourth stage, model training was performed using the PyTorch framework with the timm library. All ConvNeXt variants were initialized with ImageNet pretrained weights and fine-tuned on the chili pest dataset. To ensure fair and objective comparisons, identical hyperparameter settings were applied across all models. Training was conducted using the AdamW optimizer with a learning rate of 5×10^{-4}, batch size 32, and a maximum of 50 epochs. An early stopping strategy with a patience of 8 epochs was applied to prevent overfitting and ensure stable convergence. All experiments were executed in a Kaggle Notebook environment utilizing an NVIDIA Tesla P100 GPU, providing sufficient computational resources and reproducibility.

The final stage focused on comprehensive model evaluation. Classification performance was assessed using accuracy, precision, recall, macro F1-score, and confusion matrix, which together provide a detailed understanding of both overall and class-wise prediction performance. The use of macro-averaged metrics was particularly important to ensure fair evaluation across pest classes with imbalanced sample distributions. This evaluation strategy allows a more reliable assessment of the model's generalization ability under realistic agricultural conditions.

Beyond predictive performance, computational efficiency was evaluated using FLOPs and inference latency, as well as model size and number of parameters. The Area Under the ROC Curve (AUC-ROC) was also computed to assess classification effectiveness under class imbalance conditions. This multi-dimensional evaluation enabled the identification of ConvNeXt variants that achieve an optimal balance between accuracy, robustness, and computational efficiency for practical deployment in smart agriculture systems.

4. Results and Discussion

4.1 Sample of Classified Pest Images

Figure 2 presents representative samples of the cropped pest images used in this study. Each image contains a single pest object extracted using YOLO-based bounding box annotations. The dataset consists of four classes: Green Peach Aphid (MP), Silverleaf Whitefly (BT), Thrips (T), and Cotton Bollworm (C).

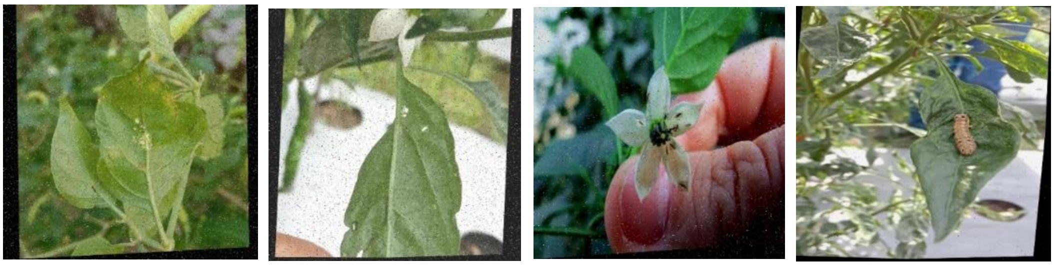


Fig. 2. Representative cropped pest images from the Red Chili Pepper Pest Dataset: (a) Green Peach Aphid (MP), (b) Silverleaf Whitefly (BT), (c) Thrips (T), and (d) Cotton Bollworm (C).

The Green Peach Aphid and Silverleaf Whitefly classes exhibit high visual similarity in terms of body size and texture, making them challenging to distinguish. Thrips are characterized by elongated body structures, while Cotton Bollworm larvae present distinct segmented body shapes. These inter-class similarities and intra-class variations increase classification complexity and justify the need for robust deep learning architectures.

4.2 Cropped Dataset Results

The initial stage of dataset preprocessing was conducted through a cropping process using bounding box annotations in the YOLO format. In the original dataset, pest objects were embedded within images containing substantial background areas, which could reduce the effectiveness of feature extraction by the model. By applying cropping, the images become more focused on the target objects, enabling the model to learn visual features more optimally, in line with the principles of visual representation employed in modern architectures such as ConvNeXt [14].

YOLO annotations provide the center coordinates (x_center, y_center) along with the bounding box dimensions (width and height) in a standardized format. These values were converted into pixel units based on the original image resolution prior to the cropping operation. Bounding boxes with extremely small sizes were filtered out to ensure sufficient image quality for model training.

The increase in the number of images after the cropping process occurred because a single original image often contained more than one pest object. Each detected object could be separated into an individual image, thereby increasing dataset diversity. Table 1 presents the distribution of the number of images after the cropping process.

Table 1. Distribution of Chili Pest Dataset After the Cropping Process

Hama	Class	Train	Validation	Test	Total Images
Green peach aphid (Myzus persicae Sulz.)	MP	1682	1407	405	3494
Silverleaf whitefly (Bemisia tabaci)	BT	1261	291	172	1724
Thrips	T	2053	540	215	2808
Cotton bollworm (Helicoverpa armigera)	C	1784	475	242	2501
Total Images		6780	2713	1034	10527

By removing irrelevant background regions, the cropped images become more focused on key characteristics such as the shape, texture, and color of the pests. Similar approaches have been shown to enhance feature extraction effectiveness in various computer vision studies [23]. Consequently, the cropped dataset becomes more representative and provides significant support for improving the performance of the ConvNeXt model.

4.2 Comparative Analysis of Training Performance

Training performance was evaluated using three validation metrics: validation accuracy, validation macro-F1 score, and validation loss. These metrics provide a comprehensive assessment of model behavior, particularly under class-imbalanced conditions where macro-F1 is essential for measuring balanced performance across all pest classes. All ConvNeXt variants demonstrated consistent improvement throughout training, although differences were observed in convergence speed, stability, and validation performance across model scales.

Table 2.Training Performance Evaluation of ConvNeXt Models

Model	Best Epoch	Train Loss	Lowest Val Loss	Val Accuracy	Macro-F1
ConvNeXt V1-Tiny	26	0.1203	0.0577	0.9797	0.9755
ConvNeXt V1-Small	20	0.1608	0.0810	0.9735	0.9642
ConvNeXt V1-Base	7	0.1332	0.0521	0.9805	0.9742
ConvNeXt V2-Atto	22	0.1176	0.0757	0.9698	0.9587
ConvNeXt V2-Femto	9	0.1759	0.0436	0.9886	0.9838
ConvNeXt V2-Pico	24	0.1750	0.1203	0.9628	0.9562
ConvNeXt V2-Nano	21	0.0939	0.0240	0.9926	0.9898
ConvNeXt V2-Tiny	17	0.1216	0.0300	0.9919	0.9899
ConvNeXt V2-Base	16	0.1939	0.0943	0.9635	0.9573

As shown in Table 2, ConvNeXt V2 variants generally outperformed their ConvNeXt V1 counterparts in terms of validation accuracy and macro-F1 score, indicating that the architectural refinements in ConvNeXt V2 enhance feature representation and learning efficiency. Notably, several smaller V2 variants achieved superior performance compared to larger models, suggesting that increased model capacity does not necessarily translate into improved validation performance for this dataset.

Among all evaluated models, ConvNeXt V2-Nano and ConvNeXt V2-Tiny achieved the strongest training results. ConvNeXt V2-Nano obtained the highest validation accuracy (0.9926), while ConvNeXt V2-Tiny achieved a slightly higher macro-F1 score (0.9899). The performance difference between these two models is negligible, indicating that both variants provide highly balanced and stable classification performance. In contrast, larger models such as ConvNeXt V2-Base did not demonstrate proportional performance gains, further supporting the suitability of compact architectures for the chili pest dataset.

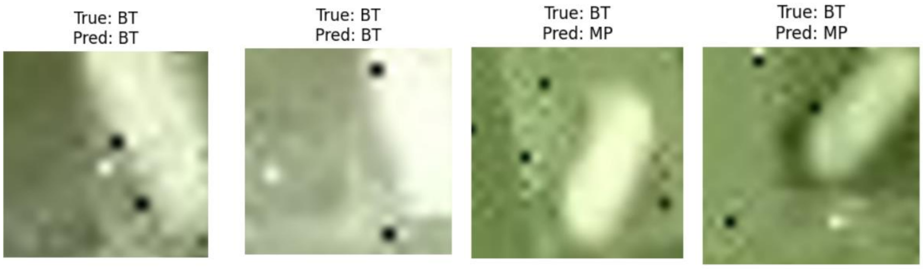


Fig. 3. Sample Predictions of ConvNeXt V2 on the Training Dataset

Figure 3 presents representative prediction samples generated by the ConvNeXt V2-Tiny model during the training phase. The figure includes both correctly classified instances (True: BT, Pred: BT) and misclassified samples (True: BT, Pred: MP). The correctly classified images indicate that ConvNeXt V2-Tiny has effectively learned discriminative visual features of the BT class. Meanwhile, the misclassified samples reveal remaining ambiguities in feature representation that may arise from visual similarities between pest categories. These qualitative observations support the quantitative results in Table 2 and illustrate the learning characteristics of the selected model.

In summary, small to medium-sized ConvNeXt V2 variants, particularly ConvNeXt V2-Nano and ConvNeXt V2-Tiny, exhibit the most stable convergence behavior and balanced validation performance. These models are therefore selected as the most reliable candidates for subsequent evaluation on the test dataset.

4.3 Comparison of Trained Model Performance on the Test Dataset

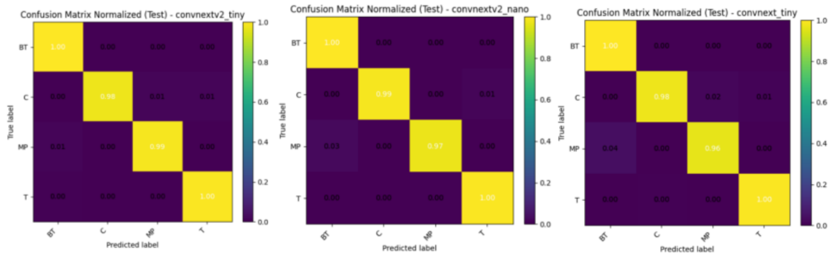
Evaluation on the test dataset was conducted to assess the generalization capability of each ConvNeXt model variant after training. Five primary metrics were employed: test loss, accuracy, macro-precision, macro-recall, and macro-F1 score. Macro-averaged metrics were selected to ensure balanced performance evaluation across all pest classes, particularly given the imbalanced class distribution. The complete results are summarized in Table 3.

Table 3.Performance Evaluation of ConvNeXt Models on the Test Dataset

Model	Test Loss	Accuracy	Macro-Precision	Macro-Recall	Macro-F1
ConvNeXt V1 – Tiny	0.0779	0.9787	0.9740	0.9834	0.9782
ConvNeXt V1 – Small	0.0860	0.9691	0.9602	0.9748	0.9666
ConvNeXt V1 – Base	0.0631	0.9758	0.9682	0.9837	0.9748
ConvNeXt V2 – Atto	0.0572	0.9816	0.9758	0.9855	0.9803
ConvNeXt V2 – Femto	0.1173	0.9671	0.9585	0.9693	0.9628
ConvNeXt V2 – Pico	0.0951	0.9594	0.9584	0.9588	0.9581
ConvNeXt V2 – Nano	0.0422	0.9845	0.9798	0.9883	0.9838
ConvNeXt V2 – Tiny	0.0435	0.9884	0.9863	0.9896	0.9878
ConvNeXt V2 – Base	0.1253	0.9584	0.9492	0.9682	0.9568

The results demonstrate that all ConvNeXt variants achieve strong generalization performance, with accuracy and macro-F1 scores consistently exceeding 95%. Consistent with the training phase, ConvNeXt V2 architectures generally outperform their ConvNeXt V1 counterparts, particularly in terms of macro-F1 score, indicating more effective feature extraction and improved class discrimination on unseen pest images.

Among all evaluated models, ConvNeXt V2-Tiny achieved the highest macro-F1 score (0.9878) and accuracy (0.9884), followed closely by ConvNeXt V2-Nano with a macro-F1 of 0.9838. The marginal performance gap between these two models indicates stable and balanced classification capability. These findings reinforce the observation that increasing model capacity does not necessarily guarantee superior generalization performance for this dataset. Larger variants such as ConvNeXt V2-Base exhibit slightly lower macro-F1 scores, suggesting that excessive model complexity may introduce reduced efficiency without significant performance gains.



(a) (b) (c)

Fig. 4. Normalized confusion matrices for three representative models: (a) ConvNeXt V1-Tiny, (b) ConvNeXt V2-Nano, and (c) ConvNeXt V2-Tiny

Figure 4 illustrates the normalized confusion matrices of three representative models. ConvNeXt V2-Nano and ConvNeXt V2-Tiny demonstrate strong diagonal dominance, indicating consistently high prediction accuracy across all pest classes. Misclassifications occur only in minor proportions, primarily between visually similar classes such as C and MP. In contrast, ConvNeXt V1-Tiny exhibits relatively higher off-diagonal values for these classes, contributing to its comparatively lower macro-F1 score. These observations confirm the superior generalization capability of ConvNeXt V2, particularly the Nano and Tiny variants.

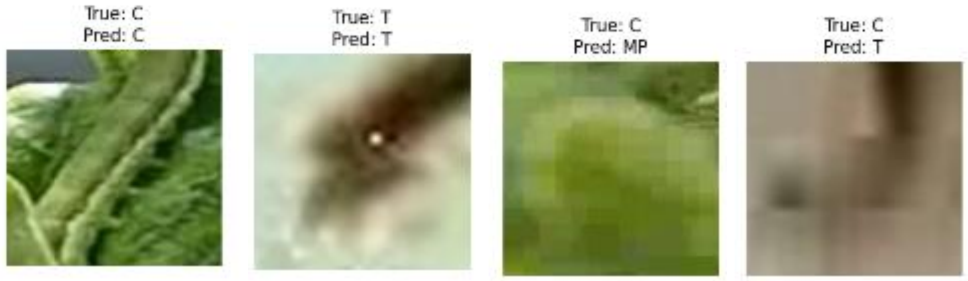


Fig. 5. Sample Predictions of ConvNeXt V2-Tiny on the Test Dataset

Figure 5 presents representative test samples classified by the ConvNeXt V2-Tiny model. The examples include both correctly classified instances (e.g., True: C, Pred: C; True: T, Pred: T) and misclassified samples (e.g., True: C, Pred: MP; True: C, Pred: T). The correct predictions demonstrate the model's ability to generalize effectively to unseen pest images. Meanwhile, the misclassified cases primarily occur between visually similar classes, suggesting subtle inter-class feature overlaps. These qualitative observations are consistent with the high macro-F1 score reported in Table 3 and further confirm the robustness of ConvNeXt V2-Tiny in practical classification scenarios.

4.4 Computational Complexity and Model Efficiency Analysis

A computational complexity analysis was conducted to evaluate the efficiency of each ConvNeXt variant for real-world deployment scenarios with constrained computational resources. Four primary aspects were examined: the number of parameters, model size, FLOPs, and inference latency. These efficiency indicators were analyzed alongside predictive performance using macro-, micro-, and weighted-AUC metrics. The results are summarized in Table 4.

Table 4. Computational Complexity and Efficiency Analysis of ConvNeXt Models

Model	Params (M)	Model Size (MB)	FLOPs (GMac)	Latency (ms)	AUC Macro	AUC Micro	AUC Weighted
V1-Tiny	27.82	106.21	4.48	6.49	0.9987	0.9986	0.9986
V1-Small	49.46	188.80	8.72	12.56	0.9989	0.9986	0.9989
V1-Base	87.57	334.19	15.41	12.62	0.9999	0.9993	0.9999
V2-Atto	3.39	12.98	0.55	6.72	0.9996	0.9995	0.9996

Model	Params (M)	Model Size (MB)	FLOPs (GMac)	Latency (ms)	AUC Macro	AUC Micro	AUC Weighted
V2-Femto	4.85	18.55	0.79	6.55	0.9980	0.9965	0.9983
V2-Pico	8.56	32.69	1.38	7.07	0.9984	0.9985	0.9982
V2-Nano	14.99	57.22	2.46	7.69	0.9996	0.9996	0.9996
V2-Tiny	27.87	106.39	4.48	9.63	0.9997	0.9995	0.9997
V2-Base	87.70	334.68	15.40	18.25	0.9982	0.9974	0.9981

Overall, smaller ConvNeXt V2 variants demonstrate superior computational efficiency while maintaining consistently high AUC values across all evaluation metrics. For instance, ConvNeXt V2-Atto achieves competitive performance with only 3.39 million parameters and 0.55 GMac FLOPs, illustrating the effectiveness of architectural refinements introduced in ConvNeXt V2.

ConvNeXt V2-Nano and ConvNeXt V2-Tiny provide the most balanced trade-off between predictive performance and computational cost. Both models achieve macro-AUC values close to 1.0 with inference latency below 10 ms per image, making them suitable for real-time implementation on edge devices. In contrast, larger models such as ConvNeXt V1-Base and ConvNeXt V2-Base exceed 80 million parameters and exhibit higher latency (12–18 ms), resulting in reduced computational efficiency despite comparable AUC performance.

These results indicate that higher parameter counts do not proportionally improve predictive performance when computational cost is considered. The ConvNeXt V2 architecture demonstrates improved capacity utilization, enabling compact models to achieve near-equivalent classification performance at substantially lower resource requirements. Based on this analysis, ConvNeXt V2-Nano and ConvNeXt V2-Tiny are recommended as the most efficient and practically deployable models for chili pest classification.

4.5 Robustness Evaluation under Image Variations and Test-Time Augmentation

Robustness evaluation was conducted to assess the resilience of ConvNeXt models under common image degradations encountered in field conditions. Five perturbation types were applied at varying severity levels: Gaussian noise, Gaussian blur, JPEG compression, brightness variation, and rotation. Additionally, Test-Time Augmentation (TTA) was evaluated to analyze its impact on prediction stability. The results are summarized in Table 5.

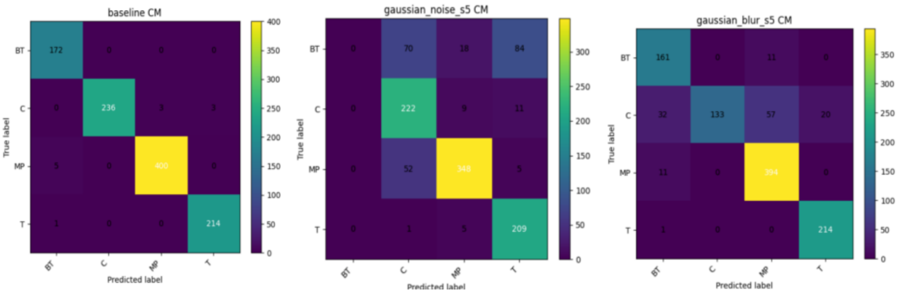
Table 5. Robustness Performance Comparison of ConvNeXt Variants Under Image Corruptions and Test-Time Augmentation

Model	Baseline	TTA	Noise s5	Blur s5	JPEG s5	Bright s5	Rot s5
ConvNeXt V1 Tiny	0.9722	0.9729	0.3356	0.9510	0.9475	0.9630	0.9134
ConvNeXt V1 Small	0.9718	0.9730	0.2516	0.9585	0.9339	0.9631	0.9609
ConvNeXt V1 Base	0.9822	0.9854	0.4870	0.8573	0.9669	0.9701	0.9365
ConvNeXt V2 Atto	0.9449	0.9467	0.2469	0.9210	0.9018	0.8771	0.9045
ConvNeXt V2 Femto	0.9733	0.9743	0.6615	0.8217	0.9518	0.9585	0.9205
ConvNeXt V2 Pico	0.9581	0.9627	0.3680	0.9265	0.9215	0.9535	0.9434
ConvNeXt V2 Nano	0.9814	0.9804	0.2986	0.7877	0.9516	0.9732	0.8393
ConvNeXt V2 Tiny	0.9878	0.9878	0.6102	0.8564	0.9557	0.9488	0.9639
ConvNeXt V2 Base	0.9568	0.9558	0.2964	0.9456	0.9283	0.9094	0.9189

Overall, performance decreases as corruption severity increases, with Gaussian noise emerging as the most detrimental perturbation across all models. While several ConvNeXt V2 variants demonstrate competitive robustness compared to their V1 counterparts, robustness varies depending on the corruption type and model scale.

Among the evaluated models, ConvNeXt V2-Tiny exhibits the most stable performance under severe corruption conditions, particularly under Gaussian noise and rotation. Compared to V2-Nano, V2-Tiny maintains higher macro-F1 under high-severity noise, suggesting that slightly increased representational capacity improves resilience to texture-level degradation. In contrast, larger variants such as ConvNeXt V2-Base do not consistently demonstrate improved robustness, indicating that higher model complexity does not guarantee better resistance to input perturbations.

The application of Test-Time Augmentation provides modest but consistent improvements across most variants. Although the gains are relatively small, TTA enhances prediction stability without additional training overhead.



(a) (b) (c)

Fig 5. Confusion matrices of ConvNeXt V2-Tiny under different evaluation conditions: (a) baseline, (b) Gaussian noise at the highest severity level, and (c) Gaussian blur at the highest severity level

To further clarify the patterns of misclassification, a confusion matrix analysis was performed on the best-performing model, namely ConvNeXt V2-Tiny. Figure 5 presents the confusion matrices under three conditions: (a) baseline, (b) Gaussian noise at the highest severity level, and (c) Gaussian blur at the highest severity level.

Under baseline conditions, the confusion matrix is dominated by the main diagonal, indicating a very high classification accuracy. Under the highest severity of Gaussian noise, an increase in misclassifications is observed, particularly between the MP and BT classes, which tend to be confused with each other. This suggests that both classes rely heavily on fine-grained texture details. In contrast, under Gaussian blur corruption, the main diagonal pattern remains relatively dominant, indicating that the model is still able to exploit global information and object shape structures. Overall, these results confirm that Gaussian noise is the most critical type of corruption affecting model performance compared to other image degradations.

4.6 Discussion

The results indicate that ConvNeXt V2 variants, particularly the Nano and Tiny configurations, achieve high classification performance while maintaining computational efficiency and model robustness. These findings reinforce current research trends showing that modern convolutional architectures remain competitive for agricultural image classification tasks.

Previous studies on chili disease and pest classification have commonly employed CNN architectures such as AlexNet, SqueezeNet, DenseNet, and transfer learning approaches [9]–[12], as well as YOLO-based detection methods [13], [22]. Although these approaches achieved strong accuracy, most studies did not explicitly evaluate computational efficiency or robustness under image degradation conditions. Furthermore, fine-grained classification studies in agricultural pests highlight the importance of detailed feature representations for distinguishing visually similar categories [18].

Since the introduction of ConvNeXt [14] and its further development into ConvNeXt V2 [19], redesigned CNN architectures have demonstrated high performance with improved efficiency. The findings of this study show that compact ConvNeXt V2 variants can achieve near-optimal macro-F1 and AUC scores with significantly fewer parameters and lower FLOPs compared to larger models. This is consistent with other studies applying ConvNeXt in medical imaging and complex image classification domains, which also report a favorable balance between accuracy and efficiency [15], [16].

The robustness analysis provides an additional contribution beyond prior chili classification studies [9]–[12], as evaluation was conducted under various image degradation scenarios that better represent real-world field conditions. The finding that Gaussian noise has the most significant negative impact suggests that fine-grained texture features play a crucial role in pest classification, consistent with multi-scale feature extraction approaches for chili leaf disease classification [21].

Interestingly, moderate-scale architectures such as ConvNeXt V2-Tiny provide the best trade-off between representational capacity, generalization, and robustness. This indicates that increasing parameter size does not necessarily lead to improved robustness, and that architectural design and pretraining strategies play a more critical role [19].

From an implementation perspective, the low inference latency and stable performance make ConvNeXt V2-Nano and V2-Tiny promising candidates for real-time chili pest detection systems deployed on resource-constrained devices, supporting digital transformation in the agricultural sector [1].

5. Conclusion

The YOLO-based bounding box cropping process effectively focuses images on pest objects, supporting improved feature extraction, in line with previous studies that integrated object detection to enhance agricultural image analysis performance. Experimental results show that all ConvNeXt variants achieve high classification performance, with accuracy and macro-F1 scores exceeding 95%, confirming the capability of modern CNN architectures to handle complex visual patterns in plant-related imagery. The superior performance and robustness of ConvNeXt V2 compared to ConvNeXt V1 are consistent with prior findings that architectural refinements in ConvNeXt V2 improve generalization under degraded image conditions. Furthermore, the strong performance of lightweight models such as ConvNeXt V2-Nano and V2-Tiny aligns with studies in other visually complex domains, including medical imaging, which report an optimal balance between accuracy and computational efficiency for compact ConvNeXt variants. These findings extend existing research by demonstrating that lightweight ConvNeXt V2 models are well suited for chili pest classification, thereby integrating modern CNN advances into practical smart agriculture applications.

References:

[1] I. Sahputra, I. Yurni, C. Agusniar, F. Nisa, and T. S. A. Sukiman, “Pemanfaatan Teknologi Informasi Digital Untuk Meningkatkan Produktivitas Petani,” J. Malikussaleh Mengabdi, vol. 3, no. 2, pp. 2829–6141, 2024, doi: 10.29103/jmm.

[2] M. Asir, R. Rahmi, A. I. Asir, N. F. Yuliani, and S. Rachman, “Implementasi pemasaran dalam meningkatkan penjualan komoditas tanaman cabai di Kabupaten Sinjai,” JPPI (Jurnal Penelit. Pendidik. Indones., vol. 8, no. 4, p. 964, 2022, doi: 10.29210/020222351.

[3] S. P. Mellinia, S. Lestari, I. Widhiono, and B. Dharmawan, “Systematic Literature Review : Rantai Pasok Dan Rantai Nilai Cabai Systematic Literature Review: Supply Chain And Value Chain Of Chili Peppers,” vol. 8, no. 4, pp. 1562–1570, 2024, doi: 10.21776/ub.jepa.2024.008.04.27.

[4] A. S. Wibowo, A. D. Irjayanti, and D. A. Khairunnisa, “Statistik Hortikultura 2023,” Badan Pus. Stat., 2024.

[5] Y. Trisnawati and E. Kustanti, Kementerian Pertanian Republik Indonesia Pusat Perpustakaan dan Penyebaran Teknologi Pertanian. Pusat Perpustakaan dan Penyebaran Teknologi Pertanian, 2021. [Online]. Available: https://repository.pertanian.go.id/items/d4ba4923-9a0d-499a-ad3c-7cf4754feac4?

[6] B. Habriantono, R. Masnilah, and F. K. Alfarisy, “Pengelolaan Serangan Kutu Kebul (Bemisi tabaci Genn.) pada Tanaman Cabai (Capsicum annuumL.) di Rumah Kaca,” J. Ilm. Inov., vol. 24, no. 2, pp. 131–139, 2024, doi: 10.25047/jii.v24i2.4650.

[7] T. M. Shivalingaswamy, A. Udayakumar, and M. Mani, “Pests and Their Management in Chillies and Bell Pepper,” in Trends in Horticultural Entomology, Springer Nature, 2022, pp. 971–982. doi: 10.1007/978-981-19-0343-4_39.

[8] D. Arniati, N. Goo, and H. R. D. D. Amannupunyo, “Serangan Hama dan Penyakit Pada Tanaman Cabai di Desa Waimital dan Waihatu Kabupaten Seram Bagian Barat,” J. Pertan. Kepul., vol. 6, no. 2, pp. 83–90, Oct. 2022, doi: 10.30598/jpk.2022.6.2.83.

[9] A. A. Anggara, A. Ridho, and C. Mutia, “Analisa Penyakit Pada Tanaman Cabai Merah (Capsicum Annuum L) dengan Membandingkan Tingkat Akurasi Menggunakan Metode Convolutional Neural Network (CNN) dan K-Nearest Neighbor (KNN),” J. Teknol. Inf., vol. 4, no. 1, p. 52, 2025, [Online]. Available: https://doi.org/10.35308/jti.v4i1.11816

[10] F. A. Danendra et al., “Klasifikasi Citra Penyakit Daun Cabai Rawit Dengan Menggunakan CNN Arsitektur AlexNet dan SqueezeNet,” Syntax J. Inform. Vol., vol. 12, no. 01, pp. 50–61, 2023, [Online]. Available: https://journal.unsika.ac.id/index.php/syntax/article/view/7947

[11] N. U. Khasanah and M. Fachrie, “Klasifikasi Jenis Penyakit Tanaman CabaiMenggunakan ArsitekturDenseNet201,” SINTECH J., vol. 7, no. 3, pp. 148 – 158, 2024, doi: doi.org/10.31598.

[12] R. A. Setyadi, S. Rahman, D. Manurung, M. Hasanah, and A. Indrawati, “Implementasi Transfer Learning Untuk Klasifikasi Penyakit Pada Daun Cabai Menggunakan CNN,” J. Teknol. Inf., vol. 5, no. 2, 2024, doi: 10.46576/djtechno.

[13] I. Agustian, R. Faurina, S. I. Ishak, F. P. Utama, K. Dinata, and N. Daratha, “Deep learning pest detection on Indonesian red chili pepper plant based on fine-tuned YOLOv5,” Int. J. Adv. Intell. Informatics, vol. 9, no. 3, pp. 383–401, 2023, doi: 10.26555/ijain.v9i3.864.

[14] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, “A ConvNet for the 2020s,” 2020. [Online]. Available: https://github.com/facebookresearch/ConvNeXt

[15] L. Abdel-Hamid, “ConvNeXt for Breast Cancer HER2 Scoring Using Different Types of Histopathological Stained Images,” J. Adv. Inf. Technol., vol. 16, no. 7, pp. 966–972, 2025, doi: 10.12720/jait.16.7.966-972.

[16] S. F. Toprak et al., “Automated Mucormycosis Diagnosis from Paranasal CT Using ResNet50 and ConvNeXt Small,” Bioengineering, vol. 12, no. 8, 2025, doi: 10.3390/bioengineering12080854.

[17] D. E. Wuri, “Penerapan Teknik Pengolahan Citra Dalam Pengenalan Pola Untuk Deteksi Penyakit Pada Citra Medis,” vol. 1, no. 1, 2024, [Online]. Available: https://coursework.uma.ac.id/index.php/informatika/article/view/705

[18] S. Woo et al., “ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders,” arXiv, 2023, doi: 10.1109/CVPR52729.2023.01548.

[19] Y. Han, C. Zhang, X. Zhan, Q. Huang, and Z. Wang, “Crossing multiple life stages : fine - grained classification of agricultural pests,” Plant Methods, vol. 20, no. 1, p. 191, Dec. 2024, doi: 10.1186/s13007-024-01317-w.

[20] A. Uzhinskiy, “Evaluation of Different Few-Shot Learning Methods in the Plant Disease Classification Domain,” MDPI, vol. 14, no. 1, p. 99, 2025, doi: 10.3390/biology14010099.

[21] D. Li, C. Zhang, J. Li, M. Li, M. Huang, and Y. Tang, “MCCM: multi-scale feature extraction network for disease classification and recognition of chili leaves,” Front. Plant Sci., vol. 15, p. 1367738, 2024, doi: 10.3389/fpls.2024.1367738.

[22] M. Yaseen, “What is YOLOv8: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector,” 2024, [Online]. Available: http://arxiv.org/abs/2408.15857

[23] S. Chen, Y. Ogawa, C. Zhao, and Y. Sekimoto, “Large-scale individual building extraction from open-source satellite imagery via super-resolution-based instance segmentation approach,” ISPRS J. Photogramm. Remote Sens., vol. 195, no. April 2022, pp. 129–152, 2023, doi: 10.1016/j.isprsjprs.2022.11.006.