Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 7, No. 3, June 2026, Page. 2670-2689
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5649

Optimization of ShuffleNetV2 Using Self-Knowledge Distillation for Cocoa Fruit Disease Classification

H.R Merdu Wira Jasa[*1], Anjar Wanto[2], Rizky Khairunnisa Sormin[3]

[1,2,3]Information Systems, STIKOM Tunas Bangsa, Indonesia

Email: [1]hamdanhrmwj@gmail.com

Received : Jan 28, 2026; Revised : Feb 10, 2026; Accepted : Feb 10, 2026; Published : Jun 15, 2026

Abstract

Timely cocoa fruit disease diagnosis is critical for field management, yet manual inspection is subjective and inconsistent, while many accurate deep learning models remain too computationally demanding for practical on-device use. This study aims to optimize cocoa fruit disease classification by applying self-knowledge distillation (Self-KD) to a lightweight ShuffleNetV2 architecture without increasing inference complexity. Using a three-class dataset (healthy, pod borer, and black pod rot) with preprocessing and class balancing, ShuffleNetV2 was selected as the baseline and trained with Self-KD, improving accuracy from 96.84% to 98.34% along with consistent gains in precision, recall, and F1-score. These results indicate that Self-KD provides a learning-level optimization that enhances robustness and prediction stability in lightweight CNNs, which is especially relevant for edge AI deployment in agricultural environments. Therefore, the proposed approach supports efficient, scalable, and sustainability-oriented AI (Green/Sustainable AI) for smart farming, with potential transferability to other crops that exhibit similar visual symptom patterns.

Keywords : *Cocoa Disease Classification, Image-Based Classification, Lightweight CNN, Self-Knowledge Distillation, ShuffleNetV2.*

1. INTRODUCTION

Cocoa remains a strategic commodity for Indonesia's smallholder plantation economy, yet recent national statistics indicate a persistent downturn. Between 2019 and 2022, the harvested area declined from approximately 1,560,944 ha to 1,421,009 ha, while production dropped from about 734,793 tonnes to 667,294 tonnes [1], [2]. The industry is overwhelmingly smallholder-driven (>99%), and most farmers cultivate less than 2 ha, making rapid and affordable on-site diagnosis critical for sustaining productivity [2]. Among the major yield-limiting factors, pod pests and diseases especially black pod rot caused by Phytophthora spp. are consistently reported as highly destructive and can lead to severe losses when detection and control measures are not implemented promptly [3]–[5]. These yield and quality losses translate directly into reduced farmer income and increased production costs, thereby posing a significant economic burden for smallholder-based cocoa production systems [6]. Because pod disorders often present as observable surface symptoms (e.g., color shifts, texture changes, lesion enlargement, and distinctive damage patterns), field assessment still largely depends on visual judgment and experiential knowledge. However, such manual inspection is inherently time-intensive and prone to subjectivity, which can produce inconsistent decisions under heterogeneous field conditions and when symptom boundaries across classes are subtle [6]–[10]. In this context, computer vision and deep learning have emerged as promising tools for extracting discriminative symptom cues directly from images and enabling faster, more objective disease recognition in plantation environments [10]–[15]. Even so, practical deployment typically demands models that preserve accuracy while remaining

Vol. 7, No. 3, June 2026, Page. 2670-2689
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5649

computationally efficient for mobile or edge platforms with limited resources, thereby encouraging the use and refinement of lightweight neural network architectures for cocoa disease classification [13]–[16].

Convolutional Neural Networks (CNNs) are the dominant paradigm in image classification because they learn hierarchical feature representations from data, progressing from low-level structures (such as edges and textures) to higher-level semantic patterns [17]–[19]. This capability is particularly relevant for cocoa pod disease classification, where symptom appearance can vary with cultivar, fruit maturity, and imaging conditions, and where certain categories may differ only marginally[20], [21]. By learning task-specific features automatically, CNN-based approaches can provide strong discriminative performance without relying on hand-crafted feature engineering [22]–[24].

ShuffleNetV2 was developed as a lightweight neural network architecture with an emphasis on achieving optimal performance while maintaining computational efficiency, making it particularly suitable for plant disease classification tasks in environments with limited resources [25], [26]. The ShuffleNetV2 architecture employs channel split and channel shuffle mechanisms to reduce computational complexity while ensuring effective information exchange across channels [27]. This design enables ShuffleNetV2 to operate with a relatively small number of parameters without compromising its feature representation capability. The efficiency characteristics of ShuffleNetV2 make it well suited for image classification tasks on resource-constrained devices, including deployments in field-based agricultural systems [28], [29]. Previous studies have reported that the use of lightweight architectures such as ShuffleNetV2 can deliver efficient performance in recognizing visual variations of plant diseases while supporting practical system implementation in real-world agricultural environments [30].

Knowledge distillation is not exclusively implemented through teacher–student schemes involving different network architectures. Distillation strategies can also be realized within a single network architecture through a self-distillation mechanism, in which knowledge is transferred internally across different learning depths of the network [31], [32]. In self-distillation approaches, predictive information generated by deeper network representations is utilized to guide the learning process of other parts of the network, thereby promoting consistency and stability in model predictions [33]. Several studies have demonstrated that self-distillation mechanisms can significantly enrich a model’s internal feature representations without increasing network size or computational burden, making them particularly suitable for lightweight architectures [34], [35]. Furthermore, ensemble self-distillation approaches applied to the ShuffleNetV2 architecture on other agricultural commodities have been reported to enhance prediction stability and generalization capability in image-based disease classification, indicating their relevance for adaptation to cocoa fruit disease classification tasks characterized by complex visual patterns [36], [37].

Several previous studies have demonstrated that image-based cocoa disease classification can be achieved with strong performance using deep learning approaches; however, opportunities for improvement remain in terms of efficiency and prediction stability. CNN-based studies on cocoa disease classification have reported an accuracy of approximately 91.79% in disease classification scenarios involving comparisons across multiple architectures (e.g., Custom CNN, VGG16, EfficientNetB0, and ResNet50), indicating that variations in disease symptoms and image conditions may reduce prediction consistency when models are not properly optimized [38]. For cocoa fruit disease classification involving healthy conditions, pest infestation, and black pod disease, MobileNet-based CNN approaches have been reported to achieve an accuracy of approximately 86.04%, indicating the suitability of lightweight models for field deployment while underscoring the need for further optimization strategies to enhance robustness under diverse environmental conditions [23]. Other studies comparing EfficientNetB0 and ResNet50 for cocoa fruit disease detection have reported accuracies of approximately 96.00%, emphasizing that transfer learning can yield strong performance but remains highly dependent on training configurations and the characteristics of minority classes within the dataset [22]. From a practical implementation perspective, smartphone-based applications for cocoa disease diagnosis further underline the need for models that are not only accurate but also lightweight and stable for effective use by farmers in field conditions [39]. At a broader scale within the plant disease domain, knowledge distillation has been shown to be effective in transferring predictive knowledge from more complex models to lightweight models, enabling smaller networks to retain performance while reducing computational cost, which is highly relevant as an optimization foundation for lightweight architectures [40]. In addition, self-distillation research in image classification tasks has demonstrated that distillation within a single model framework can enhance performance without requiring heterogeneous teacher–student architectures, thereby aligning methodologically with the self-distillation concept applied to ShuffleNetV2 in this study [41].

A review of previous studies indicates that deep learning approaches show considerable potential for image-based cocoa fruit disease classification, either through the use of conventional CNN architectures or lightweight models combined with transfer learning techniques. However, the primary focus of most existing studies remains limited to comparative evaluations among network architectures or the application of pretrained models, while more in-depth learning optimization strategies for lightweight architectures have received relatively little attention. In agricultural applications, knowledge distillation is typically implemented using teacher–student schemes with heterogeneous architectures, whereas self-distillation within a single architecture has rarely been examined in a comprehensive manner. Moreover, studies that specifically integrate ShuffleNetV2 with knowledge distillation mechanisms for cocoa fruit disease classification are still very limited, particularly with respect to evaluating their impact on prediction stability and the efficiency of lightweight models. These conditions highlight the need for investigations that assess the effectiveness of self-distillation applied to ShuffleNetV2 as a learning optimization strategy capable of improving model performance without increasing architectural complexity, especially for image-based cocoa fruit disease classification tasks.

This study aims to evaluate the effectiveness of applying self-knowledge distillation to the ShuffleNetV2 architecture in improving the performance of cocoa fruit disease classification. The proposed approach focuses on optimizing the learning process without increasing architectural complexity or computational burden, thereby remaining suitable for deployment on resource-constrained devices. The main contributions of this research include the application of a self-distillation scheme to ShuffleNetV2 within the cocoa fruit disease domain, an analysis of performance improvements achieved by the lightweight model compared to conventional training schemes, and an evaluation of prediction stability and model efficiency on a cocoa image dataset. The findings of this study are expected to provide a methodological reference for the development of efficient, adaptive, and practical image-based plant disease classification systems for cocoa plantation environments.

2. METHOD

This section describes the research methodology employed to achieve the objectives of the study, encompassing data processing stages, baseline model selection, and the optimization strategy applied. The research begins with the collection and preprocessing of a cocoa fruit image dataset, followed by an initial evaluation of several convolutional neural network architectures to identify the most suitable baseline model. Based on the evaluation results, ShuffleNetV2 is selected as the baseline architecture and subsequently optimized using a self-knowledge distillation approach. In addition, this section outlines the training configuration, evaluation metrics, and research framework implemented to ensure that the study is conducted in a systematic manner and that the experimental procedures are reproducible.

Vol. 7, No. 3, June 2026, Page. 2670-2689
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5649

2.1. Data Collection

The dataset used in this study consists of cocoa fruit images obtained from the Roboflow Universe platform through the Cacao dataset published in 2024 and provided by Babayo (https://universe.roboflow.com/babayo/cacao-xnsty). The dataset is released under the Creative Commons Attribution 4.0 (CC BY 4.0) license and contains a total of 4,390 labeled cocoa fruit images. All images are annotated into three classes Healthy, Pod Borer, and Black Pod Rot based on visual appearance and damage patterns observable on the fruit surface. This annotation scheme provides the basis for supervised learning in the proposed image-based cocoa fruit disease classification task.

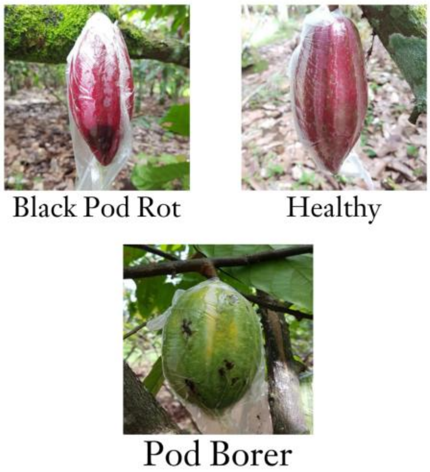

Figure 1. Sample of Image Cocoa Disease

Figure 1 presents representative sample images from the cocoa fruit disease dataset used in this study. The Healthy class represents cocoa fruits in normal conditions without visible disease symptoms, the Pod Borer class indicates damage caused by pest infestation, and the Black Pod Rot class describes cocoa fruit rot conditions commonly associated with fungal infections [20]. These examples illustrate the visual differences in color, texture, and surface damage patterns that form the basis for image-based disease classification.

2.2. Data Preprocessing

At this stage, a series of data preprocessing procedures is conducted to ensure the quality, consistency, and balance of the dataset prior to its use in the model training process. The preprocessing stage includes handling class imbalance among image categories, removing non-representative data, and adjusting the dataset to meet the requirements of deep learning models. This stage is crucial, as data quality and distribution have a significant influence on model performance as well as its ability to generalize in image-based cocoa fruit disease classification tasks.

Table 1. Distribution of the Dataset Before and After Preprocessing

Class	Before Cleaning and Augmentation	After Cleaning and Augmentation
Black Pod Rot	943	2667
Healthy	3344	2667
Pod Borer	103	2667
Total	**4390**	**8001**

Table 1 summarizes the class distribution of image samples before and after the data cleaning and augmentation stages. Initially, the dataset was imbalanced, with the Healthy class containing the highest

Vol. 7, No. 3, June 2026, Page. 2670-2689
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5649

number of samples (3,344 images), while the Pod Borer class had the fewest images (103 images). During data cleaning, the Healthy class was reduced to 2,667 samples. Subsequently, the Black Pod Rot and Pod Borer classes were increased using augmentation-based oversampling until each class reached 2,667 samples. To prevent data leakage, the dataset was split into training, validation, and testing subsets first, and augmentation was applied only to the training subset. The final dataset achieved a balanced distribution with a total of 8,001 images, making it more suitable for training and evaluating the proposed cocoa fruit disease classification model.

Table 2. Dataset Split for Training, Validation, and Testing

Class	Train	Val	Test
Black Pod Rot	1866	400	401
Healthy	1866	400	401
Pod Borer	1866	400	401

Table 2 presents the division of the dataset, which has undergone preprocessing and balancing, into training, validation, and testing subsets. The dataset is split proportionally across all classes using a composition of 70% for training data, 15% for validation data, and 15% for testing data, thereby maintaining balanced data distribution across all subsets. The training data are used to construct and train the model, the validation data are employed to monitor model performance and assist in selecting the optimal model during the training process, while the testing data are used to evaluate the model's generalization capability on unseen data. This data partitioning scheme is applied to ensure that model evaluation is conducted objectively and provides a more reliable representation of model performance.

2.3. Baseline Model Evaliation and Selection

The selection of baseline models in this study is based on the need to represent diverse characteristics of convolutional neural network architectures commonly employed in image classification, particularly under constraints of limited computational resources and field deployment requirements. ShuffleNetV2 and MobileNetV2 are selected as lightweight CNN models specifically designed for computational efficiency while maintaining competitive performance, which has led to their widespread adoption in resource-constrained applications [42], [43]. YOLOv5s Classification is included as a representative of YOLO-based architectures adapted for classification tasks, which have demonstrated strong performance in several studies involving agricultural imagery and natural object classification [44], [45]. Meanwhile, ResNet18 is utilized as a comparative model based on residual learning, offering greater depth and stability and frequently serving as a reference architecture in CNN performance evaluations [46]. By evaluating these four models, the study provides a comprehensive perspective on the trade-offs among classification accuracy, model complexity, and computational efficiency, thereby enabling an objective and comparative justification for selecting ShuffleNetV2 as the baseline architecture for subsequent optimization.

2.4. Proposed Method

ShuffleNetV2 is employed as the baseline architecture in this study because it is specifically designed to achieve a balance between computational efficiency and classification performance. The architecture is developed with consideration for hardware resource constraints, making it well suited for field deployment scenarios and systems with limited computational capacity. By leveraging channel split, channel shuffle, and inverted residual block concepts, ShuffleNetV2 effectively reduces operational complexity without compromising the feature extraction capability required for image-based cocoa fruit disease classification.

Vol. 7, No. 3, June 2026, Page. 2670-2689
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5649

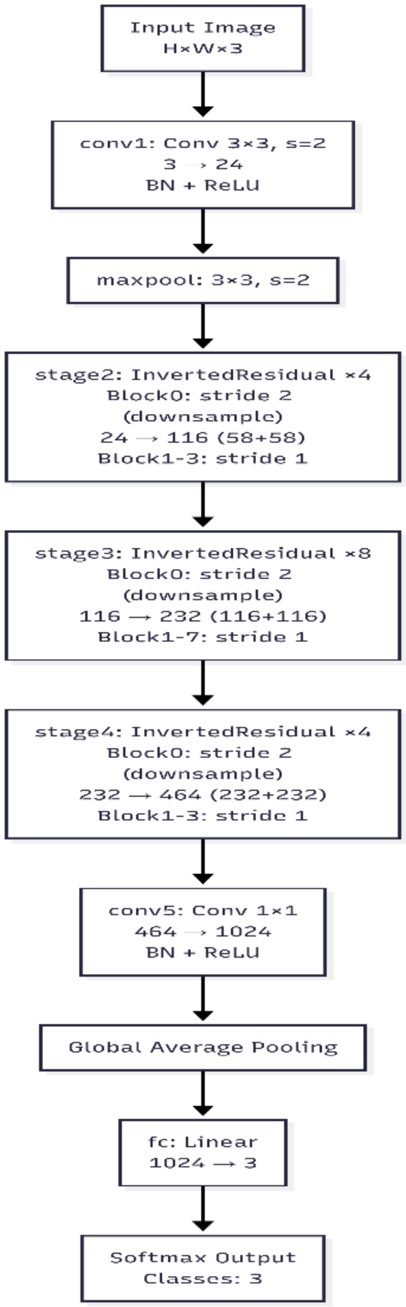

Figure 2. Architecture of the ShuffleNetV2 Model

Figure 2 illustrates the ShuffleNetV2 architecture, which represents the image processing pipeline starting from an input image with a spatial resolution of H × W and three color channels. The input image is initially processed by a 3 × 3 convolutional layer with a stride of 2 to extract low-level visual features while reducing the spatial resolution. This operation is followed by a max-pooling layer before the feature maps are propagated through a sequence of Inverted Residual blocks organized into stage2, stage3, and stage4. In each stage, the first block employs a stride of 2 for downsampling, whereas the remaining blocks preserve the feature resolution using a stride of 1. The channel split and channel shuffle mechanisms are applied to facilitate efficient cross-channel information exchange with minimal computational overhead. After the feature extraction stages, the network output is passed through a 1 × 1 convolutional layer to increase the channel dimension, followed by global average pooling and a fully connected layer with a softmax activation function to generate class probabilities. In this study, the output layer is designed to classify images into three cocoa fruit disease categories.

Vol. 7, No. 3, June 2026, Page. 2670-2689
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5649

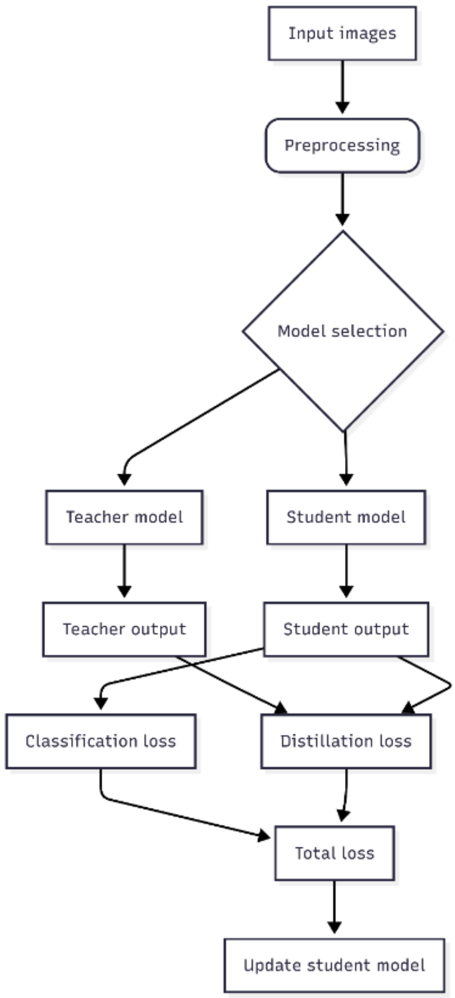

Figure 3. Self-Knowledge Distillation Scheme

Figure 3 illustrates the self-knowledge distillation scheme, depicting the training workflow that begins with input images undergoing a preprocessing stage. The system then performs model selection to define the roles of the teacher and student, which in this study employ the same network architecture. The teacher model generates predictive outputs that represent previously learned knowledge, while the student model produces predictions that are updated during the training process. The outputs of both models are used to compute two loss components: the classification loss, which measures the agreement between the student's predictions and the ground-truth labels, and the distillation loss, which quantifies the alignment between the student's predictions and those of the teacher. These two loss components are subsequently combined into a total loss that is used to update only the student model's parameters. Through this mechanism, knowledge from the teacher model is effectively transferred to the student model, thereby improving performance without increasing architectural complexity.

a. Softened Probability Distribution

The softened probability distributions used in the distillation process are obtained by applying the softmax function to the logits of both teacher and student models with temperature scaling. The formulation is presented in Equation (1).

$$q_s = softmax\left(\frac{z_s}{T}\right), \quad q_t = softmax\left(\frac{z_t}{T}\right) \quad (1)$$

- qt : softened probability distribution produced by the teacher model.

P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5649

- qs : softened probability distribution produced by the student model.
- zs : logits produced by the student model.
- zt : logits produced by the teacher model.
- T : temperature parameter used to soften the probability distribution.

b. Distillation Loss (KL-Divergence)

Distillation loss is used to measure the similarity between the probability distribution predicted by the student model and the soft target distribution produced by the teacher model. In this study, the distillation loss is calculated using the Kullback–Leibler (KL) divergence with temperature scaling, as presented in Equation (2):

$$\mathcal{L}_{\mathcal{KD}} = \mathrm{T}^2 \cdot \mathrm{KL}(\mathrm{q_t} | \mathrm{q_s}) = \mathrm{T}^{2\sum_{c=1}^{C} \mathrm{q_{t,c}} \log\left(\frac{\mathrm{q_{t,c}}}{\mathrm{q_{s,c}}}\right)} \quad (2)$$

- L_{KD} : distillation loss.
- $q_{t,c}$: probability value of class c from the teacher.
- $q_{s,c}$: probability value of class c from the student.
- C : number of classes.
- T^2 : scaling factor used to maintain gradient magnitude during training.

c. Classification Loss

The classification loss is used to measure the discrepancy between the student model's predicted class probabilities and the ground-truth labels. In this study, the classification loss is computed using categorical cross-entropy, as expressed in Equation (3).

$$\mathcal{L}_{\mathcal{CE}} = -\sum_{c=1}^{C} y_c \log(p_{s,c}) \quad (3)$$

- L_{CE} : classification loss (cross-entropy).
- $p_{s,c}$: student predicted probability for class c.
- y_c : ground-truth label (one-hot).

d. Total Loss

The total loss is defined as a weighted combination of the classification loss (cross-entropy) and the distillation loss. This formulation ensures that the student model learns both from the ground-truth labels and from the teacher's softened predictions. The total loss function is shown in Equation (4).

$$\mathcal{L}_{total} = \alpha\mathcal{L}_{\mathcal{CE}} + (1-\alpha)\mathcal{L}_{\mathcal{KD}} \quad (4)$$

- Ltotal : total loss used to update the student model.
- alpha (α) : weighting parameter controlling the contribution of classification loss.
- (1 − alpha) : weighting parameter controlling the contribution of distillation loss.

2.5. Training Configuration

The training configuration stage was designed to ensure that all experiments were conducted in a consistent, fair, and reproducible manner. The training parameters were determined by considering the stability of the learning process as well as the need for an objective performance comparison between the baseline model and the model employing the self-knowledge distillation approach. By establishing a controlled and well-documented configuration, the training process is expected to accurately reflect the impact of the proposed optimization method without being influenced by irrelevant technical variations.

P-ISSN: 2723-3863
https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5649

Table 3. Configurations Parameter

Parameter	Baseline Model	ShuffleNetV2 + Self-KD
Optimizer	Adam	Adam
Learning Rate	0.0001	0.00001
Batch Size	16	16
Epochs	15	15
Weight Decay	-	0.0001
Input Image Size	224×224	224×224
Augmentation	HFlip, VFlip, Rotation(20°), ColorJitter(b=0.2,c=0.2)	HFlip, VFlip, Rotation(20°), ColorJitter(b=0.2,c=0.2)
Library	torchvision.transforms	torchvision.transforms
Pretrained Weights	ImageNet (IMAGENET1K_V1)	ImageNet (IMAGENET1K_V1)
Framework	PyTorch 2.9.0, TorchVision 0.24.0	PyTorch 2.9.0, TorchVision 0.24.0
Hardware	Intel Core i5-1235U (12th Gen), 8 GB RAM, Windows 64-bit.	Intel Core i5-1235U (12th Gen), 8 GB RAM, Windows 64-bit.

Table 3 summarizes the training and implementation configurations used for the baseline ShuffleNetV2 model and the ShuffleNetV2 model optimized using the self-knowledge distillation (Self-KD) scheme. To ensure a fair comparison, the main training settings were kept consistent across both experiments, including the optimizer (Adam), batch size (16), number of epochs (15), input image size (224×224), and the data augmentation strategy implemented through torchvision.transforms. Both models were initialized using ImageNet-pretrained weights (IMAGENET1K_V1) and trained under the same software and hardware environment (PyTorch 2.9.0, TorchVision 0.24.0, Intel Core i5-1235U, 8 GB RAM, Windows 64-bit). The experiments were conducted without GPU acceleration, and all training was executed on the CPU to reflect a resource-constrained deployment setting. No learning-rate scheduler or early stopping was applied. The best model was selected and saved based on the highest validation accuracy during training. Under these controlled conditions, performance differences can be attributed primarily to the learning optimization strategy rather than to differences in preprocessing or computational settings.

The Self-KD configuration introduced two intentional adjustments compared to the baseline setting. First, the learning rate was reduced to improve optimization stability when the distillation loss component was added, as the additional soft-target supervision may increase gradient sensitivity during training. Second, weight decay (0.0001) was applied in the Self-KD training to provide additional regularization and reduce the risk of overfitting, thereby improving the generalization capability of the lightweight model. These modifications are consistent with common practices in knowledge distillation studies, where smaller learning rates and explicit regularization are often adopted to stabilize training and enhance the effectiveness of distillation, particularly for lightweight architectures [31], [47], [48].

2.6. Evaluation Metrics

The model performance in this study was evaluated using several classification metrics to provide a comprehensive assessment of the model's capability in recognizing each cocoa fruit disease class. The metrics employed include accuracy, precision, recall, and F1-score, which are commonly used in image classification evaluations. In addition, the confusion matrix and ROC curves were utilized to analyze model performance in greater detail at the class level. Model performance during the training process was further examined through training loss and validation loss curves, as well as training accuracy and validation accuracy plots, in order to observe the stability and convergence of the learning process.

Vol. 7, No. 3, June 2026, Page. 2670-2689
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5649

a. Accuracy

Accuracy is a metric that indicates the level of correctness of the model in making predictions across the entire test dataset. The formula for calculating accuracy is presented in Equation (5):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

- TP (True Positive): positive instances that are correctly classified.
- TN (True Negative): negative instances that are correctly classified.
- FP (False Positive): negative instances that are incorrectly classified as positive.
- FN (False Negative): positive instances that are incorrectly classified as negative.

b. Precision

Precision is a metric that reflects the model's accuracy in classifying data into a positive class by considering the proportion of correctly predicted positive instances relative to the total number of positive predictions produced by the model. The formula for calculating precision is presented in the corresponding equation. (6):

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

c. Recall

Recall is a metric used to measure the model's ability to identify all actual positive instances by comparing the number of correctly predicted positive samples to the total number of positive instances present in the test dataset. The formula for calculating recall is presented in the corresponding equation (7):

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

d. F1-Score

The F1-score is an evaluation metric used to balance precision and recall into a single measure, thereby providing a more representative assessment of model performance, particularly under imbalanced data conditions. The F1-score is computed as the harmonic mean of precision and recall, as shown in the corresponding equation (8):

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (8)$$

3. RESULT

This section describes the experimental results obtained from the implementation and evaluation of the models across all experimental scenarios conducted in this study. The performance analysis begins with the presentation of several comparative models serving as baselines, followed by a focused discussion on the ShuffleNetV2 model optimized through the self-knowledge distillation approach. Each result is evaluated using relevant classification metrics and supported by visualizations of the training and testing processes, thereby providing a clear overview of the impact of the proposed method on improving accuracy, consistency, and the generalization capability of the model in image-based cocoa fruit disease classification.

3.1. Baseline Model Performance Comparison

The baseline model performance comparison was conducted to obtain an initial overview of the capability of each architecture in classifying image-based cocoa fruit diseases. The evaluation included ShuffleNetV2, MobileNetV2, YOLOv5s Classification, and ResNet18, all of which were tested using

Vol. 7, No. 3, June 2026, Page. 2670-2689
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5649

identical training configurations and evaluation schemes. The experimental results are presented in the form of performance tables, confusion matrices, and training and validation curves to assess accuracy, training stability, and classification error patterns for each model. This analysis serves as an objective basis for determining the most suitable architecture for the subsequent optimization stage.

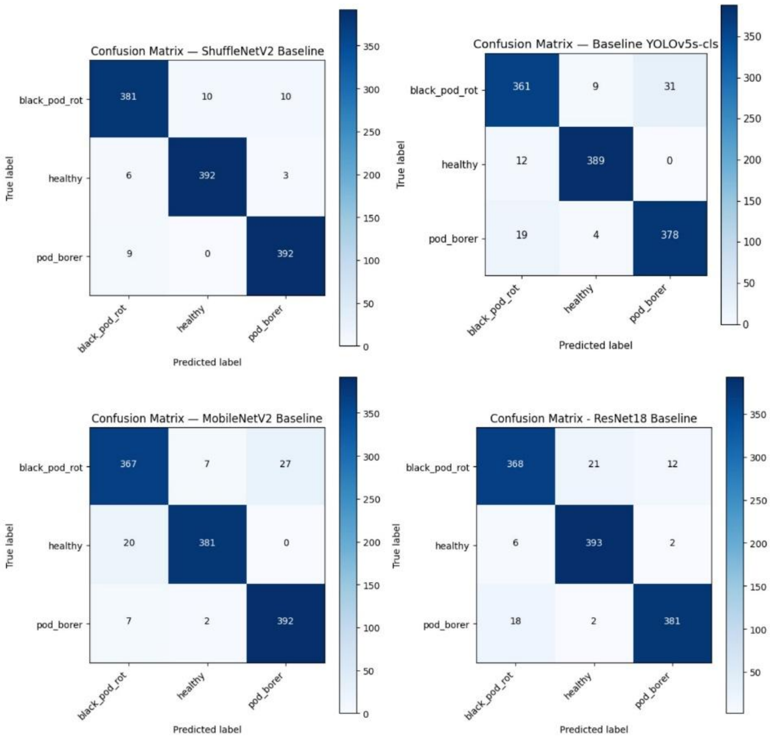

Figure 4. Confusion Matrix Baselines ShuffleNetV2, YOLOc5s-cls, MobileNetV2, and ResNet18

Figure 4 presents the confusion matrices of the baseline models evaluated in this study, including ShuffleNetV2, YOLOv5s Classification, MobileNetV2, and ResNet18, for image-based cocoa fruit disease classification. Overall, all models demonstrate strong classification performance, as indicated by the dominance of values along the diagonal of each matrix, which represents correct predictions. ShuffleNetV2 exhibits the most balanced classification results across all classes, with relatively low misclassification rates, particularly in distinguishing visually similar classes such as black pod rot and pod borer. MobileNetV2 and YOLOv5s show slightly higher misclassification rates in the black pod rot class, suggesting greater sensitivity to variations in disease appearance. ResNet18 demonstrates stable performance with moderate misclassification across classes but exhibits a more gradual discrimination pattern compared to ShuffleNetV2. These results indicate that ShuffleNetV2 achieves a favorable balance between classification accuracy and error distribution, supporting its selection as the baseline model for subsequent optimization.

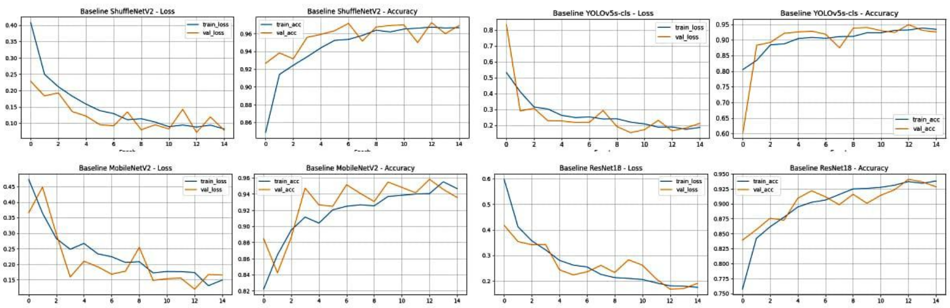

Figure 5. Learning Curve Baselines ShuffleNetV2, YOLOc5s-cls, MobileNetV2, and ResNet18

Vol. 7, No. 3, June 2026, Page. 2670-2689
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5649

Figure 5 presents the learning curves, including training and validation loss as well as training and validation accuracy, for all evaluated baseline models. Overall, the four models exhibit a decreasing loss trend and an increasing accuracy trend as the number of epochs increases, indicating that the learning process proceeds effectively. ShuffleNetV2 demonstrates faster and more stable convergence, as reflected by the relatively small gap between the training and validation curves and more controlled fluctuations compared to the other models. YOLOv5s and MobileNetV2 show more pronounced fluctuations in validation loss at certain epochs, while ResNet18 exhibits a stable learning pattern with a more gradual convergence rate. These curve patterns indicate that ShuffleNetV2 achieves a favorable balance between learning capability and training stability, thereby supporting its selection as the baseline model for the subsequent optimization stage.

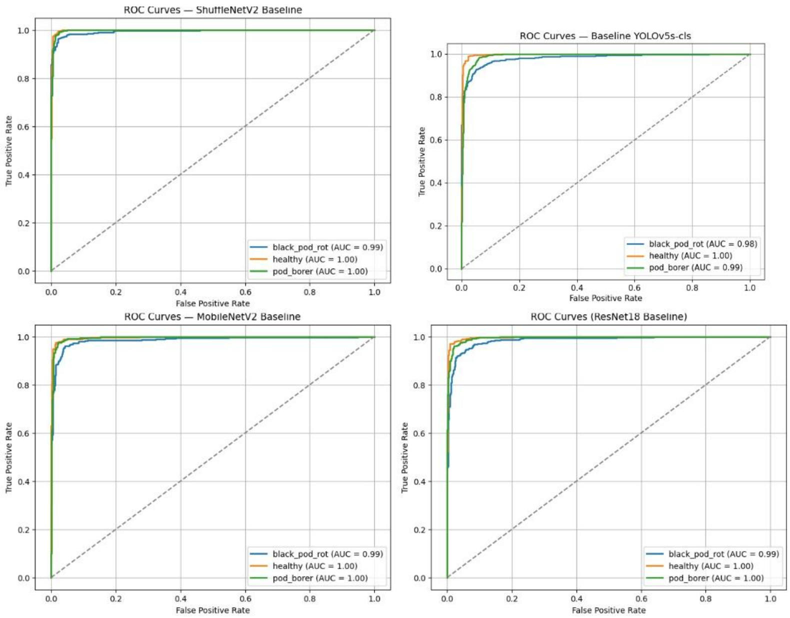

Figure 6. ROC Curve Baselines ShuffleNetV2, YOLOc5s-cls, MobileNetV2, and ResNet18

Figure 6 illustrates the ROC curves of all baseline models employed for cocoa fruit disease classification, including ShuffleNetV2, YOLOv5s Classification, ResNet18, and MobileNetV2. Overall, all models exhibit ROC curves that lie well above the random diagonal line, indicating strong class discrimination capability. The Area Under the Curve (AUC) values for each class are relatively high and close to the maximum, particularly for the healthy and pod borer classes, suggesting consistent class separability. ShuffleNetV2 and ResNet18 demonstrate more stable curves with nearly uniform AUC values across all classes, whereas YOLOv5s and MobileNetV2 show slight variations in the black pod rot class. These patterns further reinforce the previous evaluation results, indicating that ShuffleNetV2 provides balanced and reliable classification performance, thereby justifying its selection as the baseline model for the subsequent optimization stage.

Table 4. Accuracy Comparison of Baseline Models

Model	Accuracy
ShuffleNetV2	96.84%
Yolov5s-cls	93.77%
MobileNetV2	94.76%
ResNet18	94.90%

Table 4 presents a comparison of the accuracy achieved by all baseline models evaluated for image-based cocoa fruit disease classification. Based on the obtained results, ShuffleNetV2 achieves

Vol. 7, No. 3, June 2026, Page. 2670-2689
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5649

the highest accuracy of 96.84%, followed by ResNet18 and MobileNetV2 with accuracies of 94.90% and 94.76%, respectively, while YOLOv5s Classification yields a relatively lower accuracy. These accuracy differences indicate that ShuffleNetV2 is able to extract and represent visual features of cocoa diseases more effectively than the other baseline models. This finding further supports the selection of ShuffleNetV2 as the primary architecture to be optimized in the subsequent stage using the self-knowledge distillation approach.

3.2. Performance of ShuffleNetV2 with Self-Knowledge Distillation

After ShuffleNetV2 was established as the baseline model based on the results of the baseline evaluation, this stage focuses on analyzing the performance of ShuffleNetV2 optimized using the self-knowledge distillation approach. The application of this strategy aims to enhance the model's generalization capability and prediction stability without increasing architectural complexity. The evaluation is conducted by comparing the optimized ShuffleNetV2 with its baseline counterpart using classification metrics, evaluation visualizations, and an analysis of learning patterns during the training process. The results presented in this section are used to assess the extent to which self-knowledge distillation provides a tangible contribution to improving model performance.

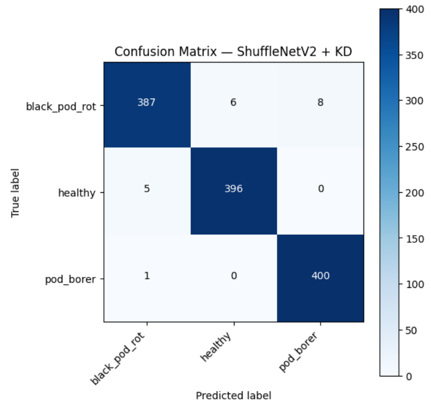

Figure 7. Confusion Matrix ShuffleNetV2 with KD

Figure 7 presents the confusion matrix of the ShuffleNetV2 model optimized using the self-knowledge distillation approach for the task of cocoa fruit disease classification. The visualization indicates a consistent improvement in classification performance across all classes, as reflected by the dominance of values along the diagonal of the matrix. The pod borer class is classified with a very high level of accuracy, evidenced by the maximum number of correct predictions and minimal misclassification errors. For the black pod rot and healthy classes, the number of incorrect predictions is also relatively low, indicating that the model is able to distinguish visual characteristics among classes more accurately. This pattern suggests that the application of self-knowledge distillation helps refine the feature representations learned by ShuffleNetV2, resulting in more stable and reliable predictions compared to the baseline model.

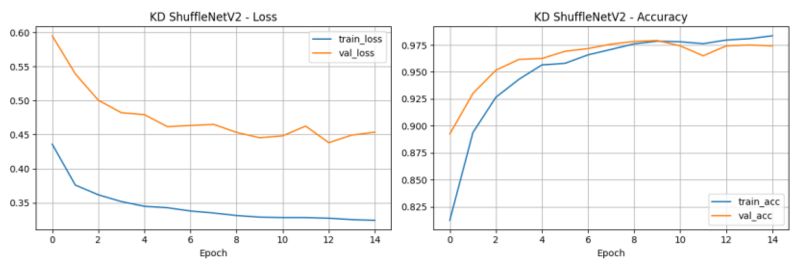

Figure 8. Learning Curves ShuffleNetV2 with Self-KD

Figure 8 illustrates the learning curves of the ShuffleNetV2 model trained using the self-knowledge distillation approach, as observed through changes in training and validation loss as well as training and validation accuracy throughout the training process. The loss curves exhibit a stable decreasing trend on the training data, while the validation loss shows minor fluctuations but remains within a controlled range, indicating a well-balanced learning process. In terms of accuracy, both training and validation accuracy increase consistently until reaching relatively convergent values, with a small gap between the curves at the final epochs. This pattern reflects improved training stability and enhanced generalization capability, supporting the effectiveness of applying self-knowledge distillation to improve ShuffleNetV2 performance without inducing significant overfitting.

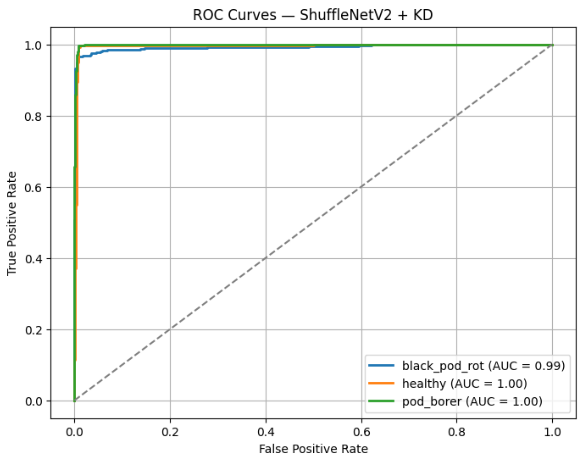

Figure 9. ROC Curve ShuffleNetV2 with Self-KD

Figure 9 shows the ROC curves of the ShuffleNetV2 model optimized using the self-knowledge distillation approach for each cocoa fruit disease class. The resulting curves lie very close to the upper-left corner of the graph, indicating the model's strong ability to effectively discriminate between positive and negative classes. The Area Under the Curve (AUC) values, which approach the maximum for all classes, demonstrate high sensitivity and specificity in the classification process. The consistency of the curve shapes across classes further indicates that the performance improvement is not confined to a single class but is evenly distributed across all evaluated classes. These findings reinforce that the application of self-knowledge distillation enhances the overall discriminative capability of ShuffleNetV2 in image-based cocoa fruit disease classification.

Table 5. Class-wise performance comparison between ShuffleNetV2 baseline and ShuffleNetV2 with Self-Knowledge Distillation

Class	**ShuffleNetV2 Baseline**			**ShuffleNetV2 with Self-KD**		
	Precision	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**
Black Pod Rot	96.21%	95.01%	95.61%	98.41%	96.51%	97.48%
Healthy	97,51%	97.76%	97.63%	98.51%	98.75%	98.63%
Pod Borer	96.79%	97.76%	97.27%	98.04%	99.75%	98.89%
accuracy		96.84%			98.34%	

Table 5 presents the class-wise performance of ShuffleNetV2 before and after the application of self-knowledge distillation in terms of precision, recall, and F1-score. The baseline ShuffleNetV2 achieves an overall accuracy of 96.84%, while the ShuffleNetV2 model trained with Self-KD attains an accuracy of 98.34% on the test dataset. Detailed metric values for each class Black Pod Rot, Healthy, and Pod Borer are reported in the table, providing a quantitative comparison of model performance across all evaluated categories.

Vol. 7, No. 3, June 2026, Page. 2670-2689
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5649

Figure 10. Qualitative comparison of Baseline and Self-KD predictions

Figure 10 presents a qualitative comparison between the Baseline ShuffleNetV2 model and the ShuffleNetV2 with Self-KD model using ten representative sample images from the test set. The figure specifically highlights example cases where the Baseline model produced incorrect predictions, while the Self-KD model successfully classified the same images into the correct ground-truth class. This visual evidence supports the quantitative results reported earlier, showing that self-knowledge distillation improves the model's robustness in recognizing subtle symptom patterns and reduces misclassification across visually similar classes. Therefore, beyond improving overall accuracy, the Self-KD scheme also contributes to more stable and consistent decision-making, strengthening the claim of improved training stability and generalization performance in practical cocoa disease classification scenarios.

In addition to the quantitative improvements summarized in Table 5 and the qualitative comparison provided in Figure 10, the findings of this study also carry important implications for the computer science and informatics domain, particularly for edge AI applications. The results confirm that self-knowledge distillation can serve as an effective learning-level optimization technique for lightweight CNN models such as ShuffleNetV2, enabling measurable improvements in accuracy and prediction stability without altering the network structure or increasing the number of parameters. This highlights the methodological value of Self-KD as a practical approach for strengthening resource-efficient models intended for deployment on mobile or embedded platforms where computational capacity and memory availability are inherently constrained.

Moreover, while the present work is centered on cocoa fruit disease classification, the observed performance trends suggest that the proposed strategy is not commodity-specific and may be transferable to other crops with visually comparable symptom patterns, including coffee, oil palm, chili, and tomato. From the perspective of sustainable computing, the ability to enhance lightweight models without escalating inference complexity is strongly aligned with the principles of Green AI and Sustainable AI, which emphasize balancing predictive performance with reduced computational and energy requirements. Therefore, the contribution of Self-KD in this study extends beyond cocoa disease recognition, supporting the broader development of efficient, scalable, and environmentally responsible deep learning solutions for precision agriculture.

4. DISCUSSION

The findings demonstrate that self-knowledge distillation yields a clear and quantifiable improvement in ShuffleNetV2 for cocoa fruit disease classification. Referring to Table 5, the overall

Vol. 7, No. 3, June 2026, Page. 2670-2689
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5649

test accuracy increases from 96.84% for the baseline to 98.34% with Self-KD, corresponding to a gain of 1.50 percentage points. Consistent improvements are also observed at the class level, where the F1-score rises from 95.61% to 97.48% for Black Pod Rot (+1.87 points), from 97.63% to 98.63% for Healthy (+1.00 point), and from 97.27% to 98.89% for Pod Borer (+1.62 points). The uniform upward trend across all classes indicates that the distillation strategy enhances the model's class-separation capability in a balanced manner, rather than concentrating gains on a single category.

The consistent improvements in precision and recall across all classes suggest that Self-KD does not merely increase overall accuracy, but also enhances the reliability of class-level decision making. Conceptually, the distillation mechanism enables the model to learn from soft targets derived from the predicted probability distribution, so the optimization process is not solely guided by hard one-hot labels but also captures implicit inter-class similarity patterns embedded in the model outputs. As a result, Self-KD can encourage smoother decision boundaries and more stable feature representations, which may reduce overly confident predictions on visually ambiguous samples. This interpretation is in line with the uniform metric gains observed across classes, particularly in categories where visual cues can overlap due to variations in color, texture, and illumination conditions during image acquisition.

Tabel 6. Performance Comparison of the Proposed Method with Related Cocoa Fruit Disease Classification Studies

Reference	**Class**	**Accuracy**
2025 [22]	EfficientNetBo	96.00%
2025 [49]	ResNeXt50+SE+CBAM	97.00%
2024 [38]	Custom CNN	91.79%
2023 [23]	MobileNetV2-SVM	86.04%
Proposed Method	ShuffleNetV2 with Self-KD	98.34%

The results reported in Table 6 indicate that the proposed ShuffleNetV2 model equipped with Self-KD is competitive when viewed alongside related cocoa fruit disease classification studies published during 2023–2025. In this study, the proposed approach attains an accuracy of 98.34%, which is higher than the accuracies reported for EfficientNetB0 (96.00%), ResNeXt50 integrated with SE and CBAM modules (97.00%), a Custom CNN (91.79%), and MobileNetV2–SVM (86.04%). However, because these studies differ in terms of datasets, class taxonomy, preprocessing, and evaluation settings, the comparison should be interpreted as contextual evidence rather than a strict head-to-head benchmark. Even with this caveat, the relative gain suggests that self-knowledge distillation can strengthen the predictive capacity of a lightweight architecture, making it a promising option for deployment in resource-constrained environments where efficiency is a key requirement.

The interpretation of Self-KD performance on the ShuffleNetV2 architecture should be considered in relation to the characteristics of the data and the experimental design adopted in this study. The dataset was prepared through class balancing and a structured data-splitting strategy; therefore, further evaluation on more diverse field images such as those captured under uneven lighting conditions, complex backgrounds, varying acquisition angles, and heterogeneous camera devices is necessary to assess the model's robustness in real-world operational settings. Moreover, the application of Self-KD involves additional loss function formulations and specific regularization schemes, making the reporting of efficiency-related aspects, including inference time, model size, and computational requirements, an important factor in supporting its applicability in resource-constrained environments. Through such expanded evaluation scenarios, the effectiveness of Self-KD on ShuffleNetV2 can be examined more comprehensively, not only in terms of accuracy but also with respect to reliability and implementation readiness. It should be noted that the Self-KD training procedure introduces additional computational cost compared to standard training, since teacher predictions must also be computed during each

Vol. 7, No. 3, June 2026, Page. 2670-2689
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5649

iteration. However, this overhead occurs only during training and does not affect inference efficiency because the deployed model architecture remains unchanged.

5. CONCLUSION

This research presents an optimization of ShuffleNetV2 by applying a self-knowledge distillation (Self-KD) strategy for cocoa fruit disease classification involving three categories Healthy, Pod Borer, and Black Pod Rot. Following baseline evaluations, ShuffleNetV2 was identified as the most suitable lightweight backbone and was subsequently improved through an internal distillation mechanism during the training phase. The experimental findings indicate that the proposed ShuffleNetV2 with Self-KD approach achieved an accuracy of 98.34%, surpassing the standard ShuffleNetV2 baseline (96.84%), while also yielding consistent increases in precision, recall, and F1-score across all disease classes. More importantly, this study contributes methodologically by demonstrating that self-distillation can enhance the capability of lightweight architectures through learning optimization, without introducing additional architectural complexity or increasing inference cost. As a result, the proposed method supports the development of efficient edge-oriented AI for smart agriculture and aligns with Green AI and Sustainable AI principles, which emphasize achieving stronger performance while maintaining computational efficiency during deployment.

For future research, broader validation is recommended using more diverse field-acquired images captured under challenging real-world conditions, such as varying illumination, complex backgrounds, different camera devices, and diverse acquisition angles, in order to further examine robustness in operational scenarios. Additionally, future work may expand the proposed framework by including more disease categories and incorporating deployment-focused efficiency measurements, such as inference latency, energy usage, and on-device performance, to strengthen its readiness for practical adoption within Indonesian smart farming environments.

CONFLICT OF INTEREST

The authors declares that there is no conflict of interest between the authors or with research object in this paper.

ACKNOWLEDGEMENT

The authors would like to express their gratitude to STIKOM Tunas Bangsa for providing institutional support and research facilities. Appreciation is also extended to the dataset owner for granting access to the cocoa fruit image data used in this study.

REFERENCES

[1] A. S. Adyanti, J. M. M. Aji, and E. S. Hani, "Mapping and Future Potential of Cocoa Commodity Base Regions in Indonesia," *Agro Ekonomi*, vol. 36, no. 2, pp. 136–155, 2025, doi: 10.22146/ae.103907.

[2] S. Dröge *et al.*, "From chocolate to palm oil: The future of Indonesia's cocoa plantations," *Ambio*, vol. 54, no. 1, pp. 151–161, 2025, doi: 10.1007/s13280-024-02061-0.

[3] J. Merga, "Epidemiology and Management Strategies of Cocoa black pod (Phytophthora spp.)," *Plant Pathology & Quarantine*, vol. 12, no. 1, pp. 34–39, 2022, doi: 10.5943/ppq/12/1/2.

[4] J. B. Rola *et al.*, "Convolutional Neural Network Model for Cacao Phytophthora Palmivora Disease Recognition," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 8, pp. 986–990, 2024, doi: 10.14569/IJACSA.2024.0150897.

[5] J. Liu and X. Wang, "Plant diseases and pests detection based on deep learning: a review," *Plant Methods*, vol. 17, no. 1, pp. 1–18, 2021, doi: 10.1186/s13007-021-00722-9.

[6] Y. Gai and H. Wang, "Plant Disease: A Growing Threat to Global Food Security," *Agronomy*, vol. 14, no. 8, pp. 1–11, 2024, doi: 10.3390/agronomy14081615.

Vol. 7, No. 3, June 2026, Page. 2670-2689
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5649

[7] M. Shoaib *et al.*, "An advanced deep learning models-based plant disease detection: A review of recent research," *Front. Plant Sci.*, vol. 14, pp. 1–22, 2023, doi: 10.3389/fpls.2023.1158933.

[8] M. Shoaib, A. Sadeghi-Niaraki, F. Ali, I. Hussain, and S. Khalid, "Leveraging deep learning for plant disease and pest detection: a comprehensive review and future directions," *Front. Plant Sci.*, vol. 16, pp. 1–19, 2025, doi: 10.3389/fpls.2025.1538163.

[9] S. Wang *et al.*, "Advances in Deep Learning Applications for Plant Disease and Pest Detection: A Review," *Remote Sens. (Basel).*, vol. 17, no. 4, pp. 1–30, 2025, doi: 10.3390/rs17040698.

[10] A. Upadhyay *et al.*, "Deep learning and computer vision in plant disease detection: a comprehensive review of techniques, models, and trends in precision agriculture," *Artif. Intell. Rev.*, vol. 58, no. 3, pp. 1–64, 2025, doi: 10.1007/s10462-024-11100-x.

[11] M. Shafay *et al.*, "Recent advances in plant disease detection: challenges and opportunities," *Plant Methods*, vol. 21, no. 1, pp. 1–44, 2025, doi: 10.1186/s13007-025-01450-0.

[12] J. P. Nyakuri, C. Nkundineza, O. Gatera, K. Nkurikiyeyezu, and G. Mwitende, "AI and IoT-powered edge device optimized for crop pest and disease detection," *Sci. Rep.*, vol. 15, no. 1, pp. 1–14, 2025, doi: 10.1038/s41598-025-06452-5.

[13] Y. Wu, L. Chen, N. Yang, and Z. Sun, "Research Progress of Deep Learning-Based Artificial Intelligence Technology in Pest and Disease Detection and Control," *Agriculture (Switzerland)*, vol. 15, no. 19, pp. 1–36, 2025, doi: 10.3390/agriculture15192077.

[14] R. Sun *et al.*, "Survey of Image Edge Detection," *Frontiers in Signal Processing*, vol. 2, pp. 1–13, 2022, doi: 10.3389/frsip.2022.826967.

[15] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, "A review of convolutional neural networks in computer vision," *Artif. Intell. Rev.*, vol. 57, no. 4, pp. 1–43, 2024, doi: 10.1007/s10462-024-10721-6.

[16] M. Krichen, "Convolutional Neural Networks: A Survey," *Computers*, vol. 12, no. 8, pp. 1–41, 2023, doi: 10.3390/computers12080151.

[17] M. M. Taye, "Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions," *Computation*, vol. 11, no. 3, pp. 1–23, 2023, doi: 10.3390/computation11030052.

[18] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, pp. 1–74, 2021, doi: 10.1186/s40537-021-00444-8.

[19] Q. Lv, S. Zhang, and Y. Wang, "Deep Learning Model of Image Classification Using Machine Learning," *Advances in Multimedia*, vol. 2022, 2022, doi: 10.1155/2022/3351256.

[20] J. Alvarado, J. F. Restrepo-Arias, D. Velásquez, and M. Maiza, "Disease Detection on Cocoa Crops Based on Computer-Vision Techniques: A Systematic Literature Review," *Agriculture (Switzerland)*, vol. 15, no. 10, pp. 1–27, 2025, doi: 10.3390/agriculture15101032.

[21] K. J. Ayikpa, A. B. Ballo, D. Mamadou, and P. Gouton, "A Hybrid Feature Fusion and Machine Learning Approach for Cocoa Pod Disease Detection," *Human-Centric Intelligent Systems*, vol. 5, no. 4, pp. 531–544, 2025, doi: 10.1007/s44230-025-00117-2.

[22] N. P. Maylianti, I. G. N. L. Wijayakusuma, and I. P. C. Arta Wiguna, "Comparison of EfficientNet-B0 and ResNet-50 for Detecting Diseases in Cocoa Fruit," *Journal of Applied Informatics and Computing*, vol. 9, no. 1, pp. 115–120, 2025, doi: 10.30871/jaic.v9i1.8868.

[23] D. Mamadou, K. J. Ayikpa, A. B. Ballo, and B. M. Kouassi, "Cocoa Pods Diseases Detection by MobileNet Confluence and Classification Algorithms," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 9, pp. 344–352, 2023, doi: 10.14569/IJACSA.2023.0140937.

[24] X. Yue, K. Qi, X. Na, Y. Liu, F. Yang, and W. Wang, "Deep learning for recognition and detection of plant diseases and pests," *Neural Comput. Appl.*, vol. 37, no. 17, pp. 11265–11310, 2025, doi: 10.1007/s00521-025-11125-5.

[25] S. Zhu and H. Gao, "MC-ShuffleNetV2: A lightweight model for maize disease recognition," *Egyptian Informatics Journal*, vol. 27, no. 7, p. 100503, 2024, doi: 10.1016/j.eij.2024.100503.

[26] H. Zhou *et al.*, "Maize Leaf Disease Recognition Based on Improved Convolutional Neural Network ShuffleNetV2," *Plants*, vol. 13, no. 12, pp. 1–18, 2024, doi: 10.3390/plants13121621.

[27] J. Su, M. Yang, and X. Tang, "Integration of ShuffleNet V2 and YOLOv5s Networks for a Lightweight Object Detection Model of Electric Bikes within Elevators," *Electronics (Switzerland)*, vol. 13, no. 2, pp. 1–16, 2024, doi: 10.3390/electronics13020394.

[28] M. Xie *et al.*, "MFFSNet: A Lightweight Multi-Scale Shuffle CNN Network for Wheat Disease Identification in Complex Contexts," *Agronomy*, vol. 15, no. 4, pp. 1–19, 2025, doi: 10.3390/agronomy15040910.

[29] L. Lu, W. Liu, W. Yang, M. Zhao, and T. Jiang, "Lightweight Corn Seed Disease Identification Method Based on Improved ShuffleNetV2," *Agriculture (Switzerland)*, vol. 12, no. 11, pp. 1–18, 2022, doi: 10.3390/agriculture12111929.

[30] H. Zhou *et al.*, "Identification of leaf diseases in field crops based on improved ShuffleNetV2," *Front. Plant Sci.*, vol. 15, pp. 1–17, 2024, doi: 10.3389/fpls.2024.1342123.

[31] H. Zhao *et al.*, "Self-Knowledge Distillation via Progressive Associative Learning," *Electronics (Switzerland)*, vol. 13, no. 11, pp. 1–15, 2024, doi: 10.3390/electronics13112062.

[32] Z. Qin, S. Ni, M. Zhu, Y. Jia, S. Liu, and Y. Chen, "A Feature Map Fusion Self-Distillation Scheme for Image Classification Networks," *Electronics (Switzerland)*, vol. 14, no. 1, pp. 1–15, 2025, doi: 10.3390/electronics14010182.

[33] C. Shi, M. Ding, L. Wang, and H. Pan, "Learn by Yourself: A Feature-Augmented Self-Distillation Convolutional Neural Network for Remote Sensing Scene Image Classification," *Remote Sens. (Basel).*, vol. 15, no. 23, 2023, doi: 10.3390/rs15235620.

[34] J. Yan, H. Li, Y. Bai, J. Liu, H. Lv, and Y. Bai, "Timestamp-Guided Knowledge Distillation for Robust Sensor-Based Time-Series Forecasting," *Sensors*, vol. 25, no. 15, pp. 1–19, 2025, doi: 10.3390/s25154590.

[35] J. Yu and H. Hu, "A Multi-Scale Convolutional Neural Network with Self-Knowledge Distillation for Bearing Fault Diagnosis," *Machines*, vol. 12, no. 11, 2024, doi: 10.3390/machines12110792.

[36] S. Ni *et al.*, "An improved ShuffleNetV2 method based on ensemble self-distillation for tomato leaf diseases recognition," *Front. Plant Sci.*, vol. 15, no. 1, pp. 1–15, 2024, doi: 10.3389/fpls.2024.1521008.

[37] X. Yang *et al.*, "A Lightweight and Efficient Plant Disease Detection Method Integrating Knowledge Distillation and Dual-Scale Weighted Convolutions," *Algorithms*, vol. 18, no. 7, 2025, doi: 10.3390/a18070433.

[38] K. S. Soh, E. G. Moung, J. J. D. Khoo, J. A. Dargham, and A. Farzamnia, "Cocoa Diseases Classification using Deep Learning Algorithm," in *ITM Web of Conferences*, Les Ulis, France: EDP Sciences, 2024, p. 1014. doi: 10.1051/itmconf/20246301014.

[39] S. Kumi, D. Kelly, J. Woodstuff, R. K. Lomotey, R. Orji, and R. Deters, "Cocoa Companion: Deep Learning-Based Smartphone Application for Cocoa Disease Detection," *Procedia Comput. Sci.*, vol. 203, pp. 87–94, 2022, doi: 10.1016/j.procs.2022.07.013.

[40] Q. Huang *et al.*, "Knowledge Distillation Facilitates the Lightweight and Efficient Plant Diseases Detection Model," *Plant Phenomics*, vol. 5, 2023, doi: 10.34133/plantphenomics.0062.

[41] Y. Li, L. Qing, X. He, H. Chen, and Q. Liu, "Image classification based on self-distillation," *Applied Intelligence*, vol. 53, no. 8, pp. 9396–9408, 2022, doi: 10.1007/s10489-022-04008-y.

[42] H. Wang, S. Qiu, H. Ye, and X. Liao, "A Plant Disease Classification Algorithm Based on Attention MobileNet V2," *Algorithms*, vol. 16, no. 9, 2023, doi: 10.3390/a16090442.

[43] X. Xu, Y. Zhang, H. Cao, D. Yang, L. Zhou, and H. Yu, "Recognition of Edible Fungi Fruit Body Diseases Based on Improved ShuffleNetV2," *Agronomy*, vol. 13, no. 6, 2023, doi: 10.3390/agronomy13061530.

[44] O. G. Ajayi, J. Ashi, and B. Guda, "Performance evaluation of YOLO v5 model for automatic crop and weed classification on UAV images," *Smart Agricultural Technology*, vol. 5, p. 100231, 2023, doi: 10.1016/j.atech.2023.100231.

[45] Y. Zhao, Y. Yang, X. Xu, and C. Sun, "Precision detection of crop diseases based on improved YOLOv5 model," *Front. Plant Sci.*, vol. 13, pp. 1–17, 2023, doi: 10.3389/fpls.2022.1066835.

[46] L. Ma, Y. Hu, Y. Meng, Z. Li, and G. Chen, "Multi-Plant Disease Identification Based on Lightweight ResNet18 Model," *Agronomy*, vol. 13, no. 11, 2023, doi: 10.3390/agronomy13112702.

Vol. 7, No. 3, June 2026, Page. 2670-2689
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5649

[47] A. M. Mansourian *et al.*, "A Comprehensive Survey on Knowledge Distillation," *Transactions on Machine Learning Research*, vol. 2025, no. 9, pp. 1–102, 2025, doi: 10.48550/arXiv.2503.12067.

[48] Z. Hao, J. Guo, K. Han, H. Hu, C. Xu, and Y. Wang, "Revisit the Power of Vanilla Knowledge Distillation: from Small Scale to Large Scale," *Adv. Neural Inf. Process. Syst.*, vol. 36, no. NeurIPS, pp. 1–14, 2023.

[49] Robet, J. T. K. Perangin Angin, and T. H. Siregar, "Attention Augmented Deep Learning Model for Enhanced Feature Extraction in Cacao Disease Recognition," *Sinkron : Jurnal dan Penelitian Teknik Informatika*, vol. 9, no. 4, pp. 1965–1977, 2025, doi: 10.33395/sinkron.v9i4.15249.