A Feasibility Study on Synthetic RGB-NIR Image Generation for Oil Palm Fresh Fruit Bunch Grading

Nor Surayahani Suriani[1], Norzali Hj Mohd[2], Shaharil Mohd Shah[3], Siti Zarina Muji[4], Fadilla Atyka Nor Rashid[5]
Department of Electronics Engineering-Faculty of Electrical and Electronics Engineering,
Universiti Tun Hussein Onn Malaysia, Johor, Malaysia[1, 2, 3, 4]
Center for Artificial Intelligence Technology, Faculty of Information Science and Technology,
Universiti Kebangsaan Malaysia, Selangor, Malaysia[5]

***Abstract*—Accurate ripeness of grading oil palm fruit bunches (FFBs) is essential for optimizing oil quality and harvesting decisions. While near-infrared (NIR) imaging provides useful spectral cues for ripeness assessment, its adoption in field conditions is limited by sensor cost and system complexity. This study presents a low-cost alternative by generating synthetic NIR images from RGB inputs using a U-Net-based image translation model and integrating the generated NIR with RGB channels for ripeness classification. Five deep learning models, including a custom CNN, ResNet-50, EfficientNet-B0, DenseNet-201 and MobileNetV3, were evaluated under RGB-only and RGB + synthetic NIR configurations using identical training protocols. Experimental results demonstrate consistent performance improvements when synthetic NIR was incorporated. EfficientNet-B0 achieved the highest overall accuracy of 90.3%, while MobileNetV3 obtained the highest macro-averaged F1-score of 85.4%, indicating strong and balanced classification across ripeness classes. Confusion matrix analysis further revealed complementary strengths between the models, where EfficientNet-B0 showed stronger robustness in late-stage maturity detection, and MobileNetV3 provided improved discrimination of early-stage ripeness. The results demonstrate that synthetic NIR augmentation enhances classification performance and training stability without requiring specialized imaging hardware.**

***Keywords*—Generative AI; deep learning; U-Net image translation; EfficientNet-B0; MobileNetV3**

I. Introduction

The oil palm (Elaeis gueneensis) is one of the most economically important oil-yielding crops globally, whose crude palm oil quality largely depends on the ripeness degree of fresh fruit bunches (FFBs) at harvest. Optimizing the stage of maturity to harvest FFBs from oil palm maximizes oil production, benefits free fatty acid concentration and minimizes post-harvest losses. Nevertheless, commercial plantations still mostly rely on visual inspection by human graders to determine ripeness, which is subjective, labour-intensive and prone to inconsistency due to variations in experience, lighting conditions and field environments.

Many years ago, computer vision and machine learning techniques were increasingly explored to automate oil palm FFB grading [1]. Conventional approaches rely heavily on RGB images captured using standard cameras, extracting visual cues such as color distribution, texture and fruitlet appearance [2]. While RGB-based methods are cost-effective and easy to deploy, their performance is often affected by illumination changes, occlusion and limited spectral information, particularly when distinguishing between adjacent ripeness stages [3]. To address these limitations, near-infrared (NIR) imaging has been introduced as NIR reflectance is sensitive to internal biochemical properties, related to moisture content, oil accumulation and fruit maturity [4].

Despite their advantages, RGB-NIR imaging systems face practical challenges in large-scale plantation deployment. The acquisition of multispectral data typically requires specialized sensors, higher costs and controlled data collection protocols [5]. Moreover, the availability of large, well-annotated RGB-NIR datasets for oil palm FFB grading remains limited, which restricts the effective training of deep learning models and may lead to overfitting. These constraints motivate the exploration of alternative strategies that can enhance model performance without substantially increasing data acquisition complexity.

Synthetic image generation has emerged as a promising solution to address data scarcity in computer vision applications. By generating artificial yet realistic images, synthetic data can augment a limited real dataset to improve feature learning and enhance generalization performance of classification models. Recent advances in deep generative models have enabled the synthesis of images across different spectral domains, including the generation of NIR representations from RGB inputs and vice versa [6]. While synthetic data generation has been successfully applied in domains such as medical imaging, remote sensing and industrial inspection. The feasibility and effectiveness of oil palm FFB grading using combined RGB-NIR information remain underexplored.

Therefore, this study investigates the feasibility of synthetic RGB-NIR image generation for oil palm fresh fruit bunch grading. Specifically, this work evaluates whether synthetically generated RGB-NIR images can enhance classification accuracy when used to augment real image datasets. The study focuses on assessing the impact of synthetic data onmodel performance rather than proposing a new grading standard with emphasis on practical applicability under constrained data conditions. The main contributions of this study are: (1) an investigation of synthetic RGB-NIR image generation as a data augmentation strategy for oil palm FFB grading, (2) a comparative analysis of classification performance using real images versus combined real and synthetic datasets, and (3) an empirical assessment of the feasibility of deploying synthetic RGB-NIR data to improve grading accuracy without additional sensor requirements.

The findings of this work provide insights into the potential role of synthetic multispectral data in advancing automated oil palm ripeness grading systems and support future research toward cost-effective and scalable solutions for plantation level deployment.

The remainder of this study is organized as follows. Section II reviews related works on oil palm ripeness detection, RGB-based computer vision methods and multispectral imaging approaches. Section III describes the proposed methodology, including synthetic NIR image generation from RGB inouts uisng U-Net model and ripeness classification using deep learning architectures, namely a custom CNN, ResNet50, EfficientNet-B0, DenseNet-201 and MobileNetV3. Section IV presents the experimental results and discussion, including qualitative evaluation of synthetic NIR images and comparative analysis of classification performance between RGB-only and RGB combined with synthetic NIR inputs. Finally, Section V concludes the study and outlines directions for future work.

II. Related Works

Automated classification of oil palm Fresh Fruit Bunches (FFBs) ripeness has attracted significant research interest due to its direct impact on palm oil yield and quality. Traditionally, FFB grading has been performed manually by trained field graders. But this approach is subjective, inconsistent and labour-intensive. As a result, computer vision and machine learning methods have been explored to standardize and automate the grading process [7].

Early works in oil palm FFB classification predominantly employed RGB image analysis and handcrafted features, such as colour and texture [8]. For instance, studies have employed variations of colour space transformation and texture measures to extract ripeness cues from surface appearance with classifiers like ANN, SVM, K-Nearest Neighbors and advanced deep learning model achieves from moderate to high classification accuracy under controlled environments [9] . These methods confirm that colour indices and textural differences correlate with FFB maturity, yet performance often degrades under real plantation environments with variable illumination, viewpoints and occlusion. In addition, such handcrafted features require careful tuning and may not generalize well across datasets.

To overcome these constraints, deep learning techniques have been adopted for ripeness detection based solely on RGB imagery. Recent work using convolutional neural networks (CNNs) and real-time object detection frameworks such as YOLO has demonstrated improved robustness by automatically learning hierarchical features from large image datasets [10]. CNN-based classifiers and YOLO-based detection pipelines have achieved competitive ripeness classification performance beyond traditional RGB-feature-based methods [11]. However, despite advantages in representation learning, RGB alone provides only limited spectral information, which can be insufficient for distinguishing close ripeness stages, especially when colour differences are subtle.

In parallel with RGB-based techniques, multispectral and near-infrared(NIR) imaging has been explored to enhance ripeness assessment [12]. NIR reflectance is sensitive to internal biochemical properties such as moisture content, cell structure and oil accumulation to offer complementary information to visible bands. A recent review found that combining visible and NIR spectra significantly improves fruit ripeness classification accuracy compared to visual inspection alone. The optical spectrometers that incorporate visible and NIR bands have demonstrated improved ripeness discrimination among ripe and overripe categories for FFB samples [13]. These findings indicate that multispectral techniques can potentially overcome limitations of RGB imaging, but they typically reuire specialized sensors that are costly and complex to deploy at scale in plantation contexts.

The emergence of hyperspectral imaging (HSI) further extends spectral coverage by capturing continuous reflectance across numerous narrow bands. HSI can reveal subtle biochemical and structural changes during fruit maturation that are not discernible with RGB or multispectral imaging alone [14]. Research on wavelength selection from hyperspectral datasets has shown that optimized spectral bands can effectively differentiate between ripeness levels in oil palm FFBs [15]. Hyperspectral devices have been employed in other fruit quality assessments, such as fruit quality evaluation [16], where reconstruction from RGB images to hyperspectral representations has shown promise in augmenting spectral information using deep learning models.

Despite the potential benefits of multipectral and hyperspectral information, practical data acquisition remains a key barrier. Hyperspectral cameras are expensive and bulky, limiting their use in operational plantation environments. Furthermore, large and well annotated RGB-NIR or RGB-HSI paired datasets for oil palm grading are scarce, posing challenges for training deep models and evaluating their generalization performance under diverse field conditions.

To address data scarcity, synthetic image generation and domain translation models have emerged as promising strategies. Deep generative models such as Generative Adversarial Networks (GANs) [17], [18] and image-to-image translation frameworks (e.g. Pix2Pix) have been used in related agricultural domains [19] to synthesize NIR or hyperspectral representations from standard RGB inputs [20]. Meanwhile, synthetic NIR generation from RGB imagery has been applied successfully in crop monitoring tasks, showing that synthetic NIR can approximate spectral information without requiring expensive hardware. Additionally, generative models have been applied to synthetic hyperspectral data generation to enhance maturity classification in other crops.

Despite progress in agricultural synthetic image generation, there are clear research gaps relevant to oil palm FFB grading:

1) Lack of paired RGB-NIR datasets specific to oil palm plantations, which limits effective learning for cross-domain translation and multispectral fusion techniques. Existing datasets primarily provide RGB imagery with limited spectral diversity, and the acquisition of matched NIR data remains sparse.

2) Limited exploration of synthetic spectral generation for oil palm ripeness tasks. Since studies have explored synthetic NIR in other crops and synthetic hyperspectral data for other agricultural quality assessments, there has been no comprehensive investigation into the feasibility of synthetic RGB-NIR images for enhancing oil palm FFB classification accuracy.

3) Insufficient evaluation under real-world plantation conditions. Most existing work is conducted under controlled lighting or constrained datasets, which do not fully capture the variability and complexity of outdoor agricultural environments.

4) Underutilized hyperspectral to RGB/NIR reconstruction frameworks in the context of oil palm grading. Although hyperspectral reconstruction from RGB has been shown to enhance spectral richness in other produce quality tasks, its applicability and effectiveness for oil palm FFb remain under-examined.

Addressing these gaps, the present study examines the feasibility of synthetic RGB-NIR image generation as a data augmentation strategy for oil palm FFB grading and evaluates whether such synthetic modalities can enhance classification performance without requiring additional hardware. This represents a step toward developing cost-effective, scalable and generalizable automated grading systems suitable for real plantation deployment.

III. Methodology

This study adopts a feasibility-driven methodological framework to investigate whether synthetic near-infrared (NIR) information generated from RGB images can enhance oil palm fresh fruit bunch (FFB) grading. The proposed approach consists of two sequential stages: (1) synthetic NIR image generation from RGB data using a U-Net model, and (2) deep learning based classification of FFB ripeness using RGB images with and without synthetic NIR augmentation. The overall objective is not to replicate true NIR measurements, but to assess whether NIR-like representations inferred from RGB data can provide complementary information that improves classification performance under practical data acquisition constraints.

A. RGB Image Acquisition and Annotation

All image data used in this study consists exclusively of RGB images captured in oil palm plantations under natural outdoor conditions. Image acquisition was performed using a standard digital camera without any near-infrared or multispectral sensors. Images were collected across multiple harvesting sessions to capture variability in illumination, background clutter, fruit bunch orientation, and occlusion. Each image corresponds to a single oil palm FFB and was manually annotated into predefined ripeness categories, such as underripe, unripe, ripe, and overripe. The grading groundtruth were determined by experienced plantation personnel following standard harvesting guidelines. These annotated RGB images constitute the sole real data source for both synthetic image generation and classification experiments. The image dataset was obtained from the Pertubuhan Peladang Parit Raja. The total data set of 3476 images for four categories of oil palm fruits is shown in Fig. 1.

Fig. 1. Oil palm fruit images underripe, unripe, (top) ripe, and overripe (bottom) from left to right.

B. Synthetic NIR Image Generation

The second phase of the simulation work was the generation of synthetic NIR images. Before model development, all RGB images underwent preprocessing to ensure consistency and suitability for deep learning models.

1) Image preprocessing: The images were resized to a fixed spatial resolution to accommodate network input requirements and to standardize feature representation. Pixel intensity values were normalized to improve numerical stability during training. The dataset was then partitioned into training, validation, and testing subsets using stratified sampling to preserve the original class distribution across splits. No NIR images were used or referenced during preprocessing, as the study relies entirely on RGB inputs.

2) U-Net architecture: Synthetic NIR image generation was performed using a U-Net convolutional neural network, which is well-suited for dense pixel-level prediction tasks due to its encoder-decoder architecture and skip connections. The U-Net model was designed to take a three-channel RGB image as input and produce a single-channel synthetic NIR image as output. Fig. 2 shows the U-Net architecture.

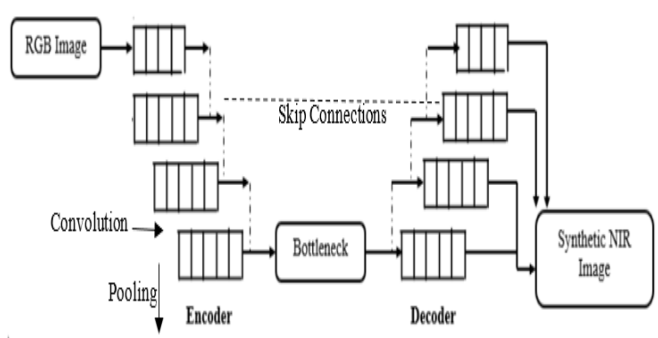

Fig. 2. U-Net model architecture for synthetic NIR image generation, illustrating the encoder-decoder structure and skip connections.

The encoder extracts hierarchical features related to surface texture, color gradients, and spatial structure. Meanwhile, the decoder reconstructs a spatially aligned representation that approximates NIR-like characteristics. Skip connections between corresponding encoder and decoder layers enable the preservation of fine-grained spatial details. The important parts were capturing fruitlet boundaries and structural patterns relevant to ripeness assessment.

Since no real NIR images were collected, the U-Net model was trained using a self-supervised and prior-guided learning strategy rather than direct pixel-wise supervision. The training objective was to encourage the network to generate physically plausible and structurally meaningful NIR-like representations based on known correlations between visible and NIR responses reported in hyperspectral and multispectral studies of agricultural produce.

The learning process was guided by a combination of reconstruction consistency constraints, smoothness regulation, and edge-preserving objectives to maintain spatial coherence and prevent unrealistic artifacts [21]. Spectral priors derived from the hyperspectral to RGB/NIR relationship were implicitly incorporated to guide the network toward a representation that emphasizes reflectance characteristics associated with fruit maturity, such as texture homogeneity and intensity variation. This training strategy aligns with the feasibility-oriented nature of the study, as the aim is to infer complementary information rather than reproduce true NIR measurements.

3) Synthetic NIR augmentation strategy: Once trained, the U-Net model was applied to the RGB images in the training set to generate corresponding synthetic NIR images. These synthetic images were then used to construct different data configurations for classification experiments. In the baseline configuration, only RGB images were used for ripeness classification.

In the augmented configuration, the synthetic NIR images were concatenated with the RGB channels to form four-channel input tensors. Additional experiments examined the use of synthetic NIR images alone and the use of synthetic NIR to augment minority classes, enabling a systematic evaluation of the contribution of synthetic spectral information under different scenarios.

C. Deep Convolutional Neural Network Model

Deep convolutional neural networks (CNNs) were employed for FFB grading due to their demonstrated effectiveness in agricultural image classification and visual feature learning. In this study, five representative CNN architectures were evaluated to comprehensively assess the impact of synthetic NIR augmentation across different network depths, parameter scales and computational characteristics, namely a custom baseline CNN, ResNet50, EfficientNet-B0, DenseNet-201 and MobileNetV3.

A baseline custom CNN architecture was first implemented to establish reference performance using both RGB-only and RGB augmented with synthetic NIR inputs. The network consists of sequential convolutional layers followed by batch normalization, nonlinear activation and max-pooling operations to progressively extract hierarchical spatial features associated with oil palm fruit texture, shape and surface patterns. This model provides a lightweight benchmark to evaluate the effectiveness of spectral augmentation under limited architectural complexity.

ResNet50 was adopted as a deep residual network benchmark for comparative analysis. The architecture employs identity shortcut connections that mitigate the vanishing gradient problem and enable stable optimization of deep networks. Owing to its strong representation capacity and extensive adoption in agricultural vision applications, ResNet50 serves as a reliable reference for evaluating the contribution of synthetic NIR information in a high-capacity model. The network was initialized using ImageNet-pretrained weights, and the final fully connected layer was replaced to match the number of FFB ripeness classes. For RGB + synthetic NIR experiments, the first convolutional layer was modified to accept four-channel input while preserving pretrained knowledge through channel-wise weight initialization.

To explore parameter-efficient and scalable architectures, EfficientNet-B0 was incorporated. EfficientNet-B0 employs compound scaling to jointly balance network depth, width and input resolution. This enables strong accuracy with significantly reduced parameter count compared to conventional CNNs. This characteristic makes EfficientNet-B0 particularly suitable for feasibility studies where both performance and computational efficiency are critical considerations.

DenseNet-201 was included to evaluate the effect of dense feature reuse and enhanced gradient propagation. Dense connections encourage feature sharing across layers and reduce redundant representations, which can be advantageous when training with limited datasets. Its inclusion allows assessment of whether dense connectivity benefits the exploitation of synthetic spectral features.

MobileNetV3 was selected as a lightweight architecture optimized for deployment in resource-constrained environments. It utilised depth-wise separable convolutions, inverted residual blocks and hardware optimization strategies to achieve favourable accuracy to complexity trade-offs. This model reflects potential real-world implementation scenarios in plantation monitoring systems where computational resources and power consumption are limited.

All networks were trained using identical dataset partitions, preprocessing pipelines and training hyperparameters to ensure fair and unbiased comparison. This multi-model evaluation framework enables systematic investigation of the consistency, robustness and generalizability of synthetic NIR augmentation across diverse CNN architectures, ranging from shallow baseline models to deep residual, densely connected and efficient oriented networks.

D. Performance Metrics

The performance of the proposed oil FFB grading framework was evaluated using standard classification metrics derived from the confusion matrix, including accuracy, precision, recall and F1-score. These metrics provide complementary perspectives on classification behavior and are widely adopted in agricultural image analysis to assess both overall performance and class-specific reliability.

Overall classification accuracy measures the proportion of correctly classified samples relative to the total number of samples and is defined as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

where TP, TN, FP and FN denote the numbers of true positives, true negatives, false positives and false negatives, respectively. While accuracy provides a general indication of model performance, it may be biased in the presence of class imbalance, which is common in real-world agricultural datasets.

Precision quantifies the proportion of correctly predicted positive samples among all predicted positives and is expressed as:

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

Recall, also referred to as sensitivity, measures the proportion of correctly predicted positive samples among all actual positives and is given by:

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

The F1-score is defined as the harmonic mean of precision and recall, which provides a balanced assessment of classification performance by jointly considering false positives and false negatives.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{4}$$

In this study, the macro-averaged F1-score was emphasized to ensure equal contribution from each ripeness class, regardless of class frequency.

$$F1_{macro} = \frac{1}{C}\sum_{i=1}^{C} F1_i \tag{5}$$

where C denotes the total number of classes and $F1_i$ represents the F1-score of the i-th class, given by

$$F1_i = 2 \times \frac{Precision_{\ i} \times Recall_i}{Precision_i + Recall_i} \tag{6}$$

where $Precision_{\ i}$ and $Recall_i$ correspond to the precision and recall computed for class i, respectively.

Among the evaluated metrics, the macro-averaged F1-score is considered the most important performance indicator for this feasibility study. Unlike accuracy, macro F1-score is less influenced by dominant classes and more sensitive to improvements in underrepresented or visually ambiguous ripeness categories. This characteristic is particularly relevant for oil palm FFB grading, where misclassification between adjacent ripeness stages can have significant economic implications. Consequently, the macro-averaged F1-score was used as the primary metric for comparing RGB-only and RGB with synthetic NIR configurations, while accuracy and class-wise precision and recall were reported as supporting metrics.

IV. Results and Discussion

The performance of the proposed framework was evaluated by comparing classification results obtained using RGB-only inputs against those obtained using RGB augmented with synthetic NIR images. Across all evaluated models, the inclusion of synthetic NIR consistently improved classification accuracy relative to the RGB-only baseline. This improvement indicates that the synthetic NIR channel provides complementary information that enhances feature discrimination for oil palm fresh fruit bunch (FFB) grading.

Performance gains resulting from synthetic NIR augmentation were observed across all model types, suggesting that the effectiveness of synthetic NIR is not dependent on a specific network architecture. This consistency supports the feasibility of using synthetic spectral representation to enhance classification robustness.

A. Synthetic NIR Generation

Fig. 3 illustrates representative examples of synthetic near-infrared (NIR) images generated by the U-Net model for oil palm FFB at four ripeness stages: underripe, unripe, ripe and overripe. Although no real NIR images were acquired during data collection, the trained U-Net successfully learned a mapping from RGB inputs to synthetic NIR-like intensity representations which produce consistent spatial and textural patterns across different maturity levels.

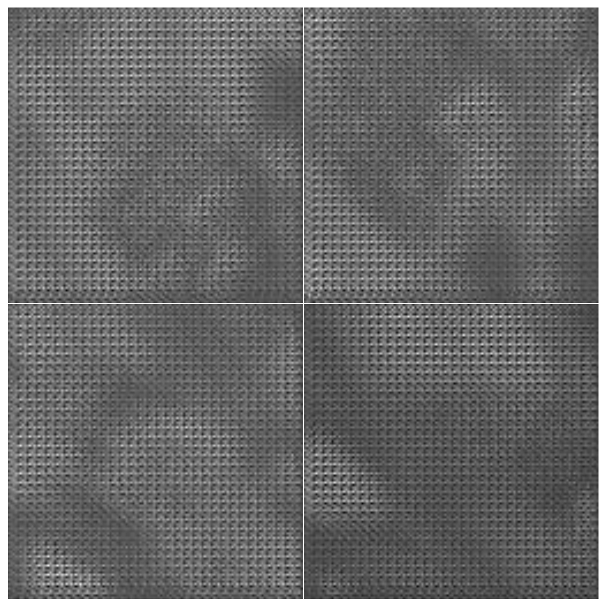

Fig. 3. Synthetic NIR of oil palm fruit images underripe, unripe (top), ripe, and overripe (bottom) from left to right.

Visually, the synthetic NIR images exhibit relatively smooth grayscale distributions with localized intensity variations that correspond to fruit surface structure and cluster density. Brighter regions indicate higher predicted reflectance, while darker areas represent lower reflectance. These intensity variations are not random noise but follow coherent spatial patterns aligned with the shape and arrangement of fruitlets within each bunch. This suggests that the network has learned meaningful morphological cues from RGB inputs.

Across ripeness stages, subtle differences in brightness distribution and texture consistency can be observed. The underripe and unripe samples display more uniform and compact intensity patterns, whereas the ripe and overripe samples exhibit increased contrast and localized brightness variation. This behavior is consistent with expected physical changes during fruit maturation, such as surface texture evolution, moisture redistribution and oil accumulation, which typically influence near-infrared reflectance characteristics. Although the generated images do not represent true spectral measurements, they encode discriminative structural information that complements visible color features.

The synthetic NIR images preserve spatial continuity and edge consistency, indicating that the U-Net effectively maintains geometric fidelity during image translation. The absence of severe artifacts or discontinuities demonstrates stable learning and successful generalization across different fruit appearances. This stability is critical because downstream classification models rely on consistent spatial patterns rather than absolute pixel values. The synthetic NIR channel likely enhances feature separability by introducing additional intensity-based cues that are less sensitive to illumination variations and color ambiguity in RGB images alone. This is particularly beneficial for discriminating adjacent ripeness stages, which often exhibit overlapping color characteristics in visible imagery.

Overall, Fig. 3 confirms that the U-Net is capable of generating structurally meaningful synthetic NIR representations from RGB inputs, providing complementary information that strengthens the robustness and discriminative power of the FFB grading framework without requiring physical NIR sensors.

B. Training Convergence and Optimization Stability

Fig. 4 presents the training and validation loss curve of EfficientNet-B0 under RGB only (left) and RGB + synthetic NIR (right) input configurations. The convergence behavior provides important insight into optimization stability, learning efficiency and generalization capability of the proposed framework.

For the RGB-only configuration, the training loss decreases rapidly during the early iterations indicate effective initial feature learning. However, noticeable fluctuations and spikes are observed throughout the training process, suggesting sensitivity to limited data variability and potential instability in optimization. The validation loss follows a similar decreasing trend but remains relatively higher and more oscillatory implies moderate generalizetion uncertainty. The corresponding accuracy curve exhibits substantial variance across iterations and reflects the difficulty of learning discriminative features using RGB information alone, particularly for visually overlapping ripeness classes.

In contrast, the RGB + synthetic NIR configuration demonstrates smoother and more consistent convergence behavior. The training loss decreases steadily with fewer abrupt fluctuations and reaches a lower final value compared to the RGB-only case. In addition, the validation loss closely tracks the training loss with a reduced gap between the two curves, indicating improved generalization and reduced risk of overfitting. The validation accuracy stabilizes at higher levels with less variance, reflecting more reliable feature learning when synthetic spectral information is incorporated.

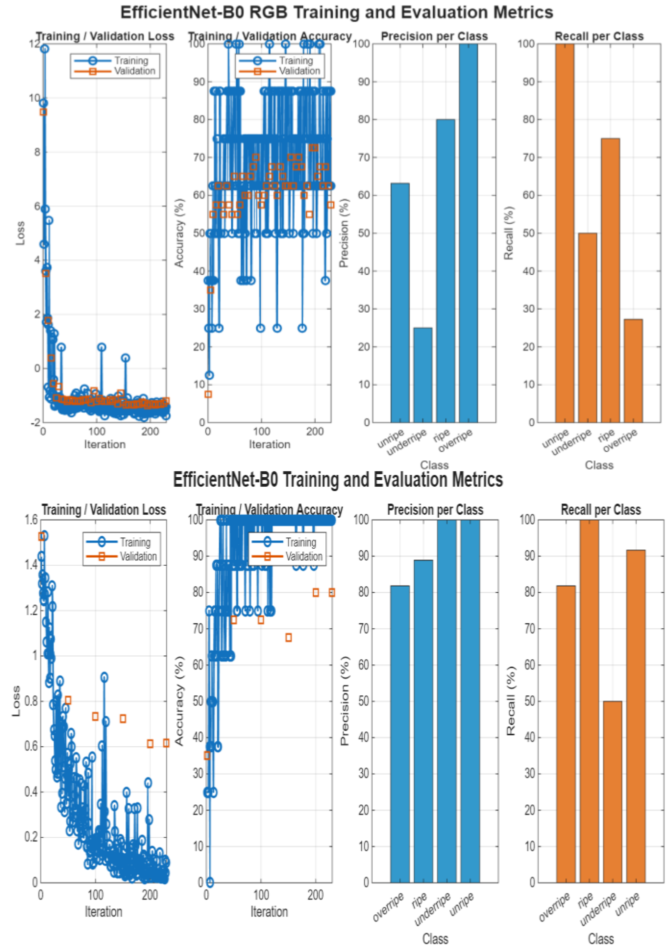

Fig. 4. Training and validation loss curves of EfficientNet-B0 for RGB-only (Top) and RGB + synthetic NIR images (Bottom).

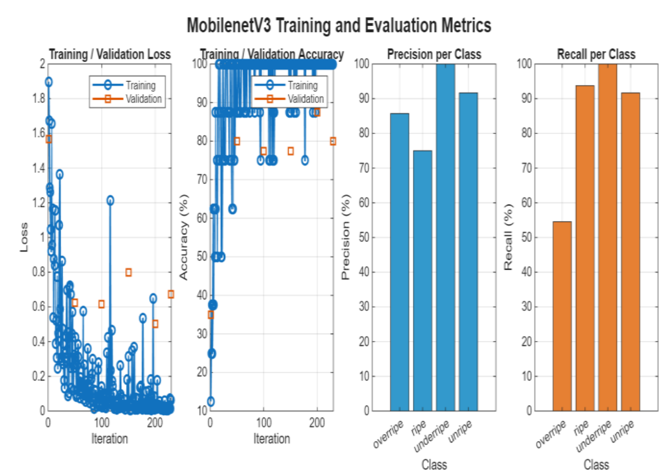

Fig. 5. Training and validation loss curves of MobileNetV3 for RGB + synthetic NIR images.

In addition to EfficientNet-B0, the convergence behavior of MobileNetV3 under the RGB + synthetic NIR configuration is illustrated in Fig. 5. The training loss exhibits a rapid decay during the early iterations. After 50 iterations, the loss stabilizes at a low magnitude with minor fluctuations, suggesting convergence toward a stable minimum and absence of optimization divergence. The gap between training and validation curves remains small, indicating good generalization capability and limited overfitting. This behavior confirms that the lightweight architecture is capable of learning robust representations even with limited training data when augmented with synthetic spectral information.

The per-class precision and recall bars indicate strong discrimination for the underripe and unripe classes, with high recall values approaching 100%, which supports the improved macro-averaged F1-score reported in Table I. The synthetic NIR images likely enhance contrast and structural cues that are less sensitive to illumination variations present in RGB images, enabling the network to converge more efficiently and consistently.

Overall, the convergence characteristics of MobileNetV3 confirm that the optimization process is stable, efficient and generalizable when synthetic NIR is incorporated. The consistent learning behavior observed across both EfficientNet-B0 and MobileNetV3 demonstrates that the benefits of synthetic NIR augmentation are not architecture dependent and supports the robustness and scalability of the proposed framework.

C. Model Performance Comparison

Table II summarizes the classification performance of different deep learning models under two input configurations: RGB-only and RGB augmented with synthetic NIR. Overall, the results demonstrate that the inclusion of synthetic NIR improves both accuracy and macro-averaged F1-score for most evaluated architectures. This confirms the feasibility of using synthetic spectral information to enhance oil palm FFB grading.

TABLE I. THE MODEL PERFORMANCE COMPARISON

Input	Model	Accuracy (%)	Macro F1-score (%)
RGB only	CNN	75.8	70.1
	ResNet50	77.5	75.3
	EfficientNet-B0	85.3	80.5
	DenseNet-201	82.1	80.1
	MobileNetV3	80.3	82.5
RGB + Synthetic NIR	CNN	78.5	74.0
	ResNet50	78.1	75.7
	EfficientNet-B0	**90.3**	84.6
	DenseNet-201	85.4	80.9
	MobileNetV3	82.9	**85.4**

Among the RGB-only models, EfficientNet-B0 achieved the highest accuracy (85.3%) and macro F1-score (80.5%), indicating strong baseline discriminative capability. DenseNet-201 and MobileNetV3 also performed competitively, suggesting that both deep feature reuse (DenseNet) and lightweight architectures (MobileNetV3) are suitable for this task. The custom CNN and ResNet50 achieved lower macro F1-scores, reflecting the limitations of shallow architectures and the sensitivity of deeper residual networks to limited training data.

When synthetic NIR was incorporated, the performance improvements were observed across all models. The most notable improvement was achieved by EfficientNet-B0, whose accuracy increased from 85.3% to 90.3%, and the macro F1-score improved substantially from 80.5% to 84.6%. This demonstrates that synthetic NIR provides complementary information that can be effectively exploited by compound-scaled architectures. Similarly, MobileNetV3 achieved the highest macro F1-score (85.4%) under RGB+synthetic NIR, indicating that lightweight networks are capable of benefiting from synthetic spectral augmentation while maintaining computational efficiency.

Although ResNet50 and CNN exhibited more modest improvements, their consistent upward trends suggest that the benefit of synthetic NIR is not architecture-specific. This consistency strengthens the argument that synthetic NIR contributes meaningful additional representation rather than acting as random noise. In addition, the macro-averaged F1-score was emphasized as the primary metric because it balances class-wise performance and mitigates class imbalance effects. The achieved macro F1-scores of 84.6% (EfficientNet-B0) and 85.4% (MobileNetV3) indicate strong and balanced classification performance across ripeness categories, validating the feasibility of the proposed approach.

1) Class-level performance analysis: Table II and III present per-class precision, recall, and F1-score for EfficientNet-B0 and MobileNetV3 under the RGB + synthetic NIR configuration. These results provide deeper insight into class-wise behavior and error distribution.

TABLE II. PER-CLASS METRICS EFFICIENTNET-B0 FOR RGB+SYNTHETIC NIR IMAGE

Class	Precision	Recall	F1-score
Unripe	100	91.67	95.65
Underripe	100	50	66.67
Ripe	88.89	100	94.12
Overripe	81.82	81.82	81.82

TABLE III. PER-CLASS METRICS MOBILENETV3 FOR RGB+SYNTHETIC NIR IMAGE

Class	Precision	Recall	F1-score
Unripe	91.67	91.67	91.67
Underripe	100	100	100
Ripe	75	93.75	83.33
Overripe	85.71	54.55	66.67

For EfficientNet-B0, the unripe class achieved perfect precision (100%) and high recall (91.67%), resulting in a strong F1-score of 95.65%. This indicates that the model reliably identifies visually distinct unripe bunches. However, the underripe class exhibits a lower recall of 50%, despite perfect precision. This suggests that while predictions labeled as underripe are highly accurate, a significant portion of true underripe samples are misclassified into neighboring classes. This behavior reflects the intrinsic visual ambiguity between the underripe and ripe stages.

The ripe class shows excellent recall (100%) and a strong F1-score (94.12%), indicating that the synthetic NIR channel enhances the model's ability to capture maturity-related features. The overripe class demonstrates balanced precision and recall (both approximately 81.8%), reflecting moderate confusion likely caused by texture degradation and environmental variability.

For MobileNetV3, the underripe class achieves perfect precision and recall (100%), indicating excellent class separability under this architecture. The unripe class also maintains balanced precision and recall (91.67%), that confirm stable discrimination. However, the overripe class shows reduced recall (54.55%), which suggests that lightweight models may struggle to capture subtle degradation patterns associated with over-ripeness, even with synthetic NIR augmentation. The ripe class maintains reasonable performance with an F1-score of 83.33%.

These class-wise results indicate that synthetic NIR improves discrimination, particularly for intermediate maturity stages, but challenges remain for visually overlapping classes such as underripe and overripe.

2) Comparison between EfficientNet-B0 and MobileNetV3: MobileNetV3 achieved the highest macro F1-score (85.4%), EfficientNet-B0 delivered the highest overall accuracy (90.3%). This reveals an important trade-off between balanced class performance and overall correctness. EfficientNet-B0 demonstrates stronger overall predictive power, whereas MobileNetV3 exhibits slightly better balance across classes, particularly for underripe detection.

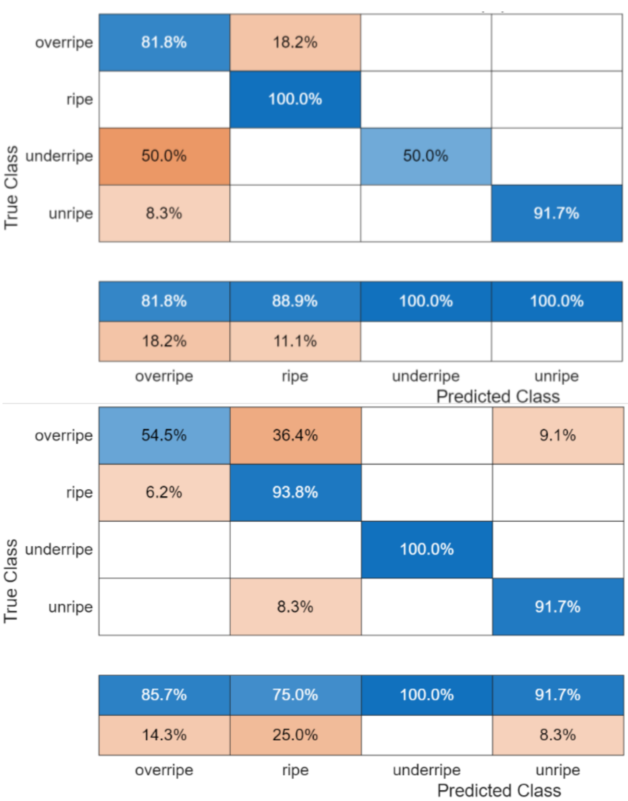

Fig. 6. Confusion matrix for efficientNet-B0 (top) and mobileNet-V3 (bottom).

Fig. 6 presents the normalized confusion matrices for EfficientNet-B0 and MobileNetV3 under the RGB + synthetic NIR configuration, providing insight into class-wise prediction behavior and misclassification patterns beyond overall performance metrics. For EfficientNet-B0, strong discrimination performance is observed for the ripe and unripe classes, where correct classification rates reach 100% and 91.7%, respectively. This indicates that the model effectively captures maturity-related texture and structural cues enhanced by synthetic NIR information. The overripe class also achieves a high recognition rate of 81.8%, with the remaining 18.2% primarily misclassified as ripe.

This confusion is expected due to gradual visual transitions between late ripeness stages, where texture degradation and color saturation differences may be subtle. The underripe class shows comparatively weaker recall, with only 50% correctly classified. While the remaining samples are misclassified as ripe. This reflects inherent ambiguity between early maturity stages and highlights the challenge of separating underripe and ripe fruit bunches based solely on visual cues, even with synthetic spectral augmentation.

In contrast, MobileNetV3 demonstrates excellent recognition for the underripe and unripe classes, achieving 100% and 91.7% correct classification rates, respectively. This suggests that the lightweight architecture effectively leverages synthetic NIR features to enhance early-stage ripeness discrimination. The ripe class also maintains strong performance at 93.8% with only minor confusion toward the overripe category. However, MobileNetV3 exhibits noticeably reduced performance for the overripe class, where only 54.5% of samples are correctly classified. A substantial proportion (36.4%) is misclassified as ripe, and a smaller fraction (9.1%) is confused with unripe. This indicates that MobileNetV3 may lack sufficient representational capacity to fully capture subtle degradation patterns associated with over-ripeness, despite benefiting from synthetic spectral augmentation.

Overall, the confusion matrix analysis reveals complementary strengths between the two architectures. EfficientNet-B0 provides stronger robustness in detecting later ripeness stages, particularly when overripe samples contribute to higher overall accuracy. Conversely, MobileNetV3 excels in discriminating early maturity stages and results in a slightly higher macro-averaged F1-score due to balanced class performance. These observations align with the quantitative results reported in Table I and the per-class metrics in Tables II and III.

From a deployment perspective, MobileNetV3 offers advantages in computational efficiency and suitability for embedded systems, making it attractive for real-time plantation applications. Meanwhile. EfficientNet-B0 may be preferable when maximizing classification accuracy is the primary objective and computational resources are less constrained. The close macro F1-scores between these two models further reinforce the robustness of the synthetic NIR augmentation strategy across different architectural paradigms.

3) Impact of synthetic NIR on feature discrimination: The observed improvements suggest that synthetic NIR provides complementary cues beyond visible color information. Although the synthetic NIR images are not true spectral measurements, they appear to encode structural and intensity variations that enhance feature separability in the learned representations. This is particularly beneficial for distinguishing visually similar ripeness stages, where RGB features alone may be insufficient.

The performance gains achieved without any real NIR acquisition demonstrate the practical value of the proposed approach. By leveraging RGB-only data and learned spectral priors, the system avoids additional hardware cost and deployment complexity. Hence, aligned well with real-world plantation constraints.

4) Limitations and reliability considerations: Despite the encouraging results, certain limitations remain. First, the absence of real NIR ground truth prevents direct validation of the physical accuracy of the synthetic NIR images. Therefore, conclusions are based on downstream classification performance rather than spectral fidelity. Second, class imbalance and limited sample size may influence class-wise variability, particularly for underripe and overripe categories. Nevertheless, the consistent macro F1-score improvements across multiple architectures indicate that the observed gains are systematic rather than incidental. The strong performance of lightweight architectures further supports the feasibility of deploying the proposed framework in resource-constrained environments.

V. Conclusion

This study investigated the feasibility of enhancing oil palm FFB grading performance through synthetic near-infrared (NIR) image generation from RGB inputs using a U-Net architecture. Unlike conventional approaches that require dedicated multispectral or hyperspectral sensors, the proposed framework generates synthetic spectral representations from standard RGB images. This enable cost effective spectral augmentation without additional hardware requirements.

Quantitative evaluation across five deep learning architectures custrom CNN, ResNet50, EfficientNet-B0, DenseNet-201 and MobileNetV3, showed consistent performance gains when synthetic NIR was included. EfficientNet-B0 achieved the highest overall accuracy of 90.3%, while MobileNetV3 obtained the highest macro-averaged F1-score of 85.4%, indicating strong and balanced classification performance across ripeness classes. Confusion matrix analysis revealed complementary strengths between the two models, with EfficientNet-B0 exhibiting stronger robustness in late-stage maturity detection and MobileNetV3 excelling in early-stage discrimination.

The results confirm that synthetic NIR provides complementary discriminative information beyond visible RGB features, particularly for visually ambiguous maturity stages. The improvement is also consistent across diverse architectures, reinforcing the generalizability and robustness of the proposed approach. The strong performance achieved using lightweight architectures further supports the practical feasibility of real-time deployment in plantation environments. Future work will focus on validating synthetic NIR against spectral information and exploring advanced generative models to further enhance the potential of NIR for oil palm FBB grading quality and oil estimation. Overall, the proposed synthetic RGB-to-NIR generation framework demonstrates a viable pathway for improving oil palm FBB accuracy while maintaining low deployment cost and operational simplicity, hence offering strong potential for scalable agricultural monitoring applications.

Acknowledgment

This research was supported by Universiti Tun Hussein Onn Malaysia (UTHM) through TIER1 (vot Q874).

References

[1] R. Kurniawan, Samsuryadi, Mohamad, Fatma Susilawati, Wijaya, Hama Oktafia Lingga and Santoso, Budi, "Advancing palm oil fruit ripeness classification using transfer learning in deep neural networks," Bulletin of Electrical Engineering and Informatics, vol. 14, no. 2, pp. 1126 - 1137, 2025.

[2] A. Septiarini, A. Sunyoto, H. Hamdani, F. Utaminingrum and H. Rahmania, "Machine vision for the maturity classification of oil palm fresh fruit bunches based on color and texture features," Scientia Horticulturae, vol. 286, 2021.

[3] D. Setiawan, Utomo, Pradita Eko Prasetyo and Alfalah, Muksin, "Detection of Oil Palm Fruit Ripeness through Image Feature Optimization using Convolutional Neural Network Algorithm," International Journal on Informatics Visualization, vol. 9, no. 2, pp. 674 - 682, 2025.

[4] M. Shiddiq, Saktioto, Salambue, Roni, Wardana, Fiqra, Dasta, Vicky Vernando, Harmailil, Ihsan Okta, Rabin, Mohammad Fisal, Arpyanti, Nisa and Wahyudi, Dilham, "Multispectral imaging and deep learning for oil palm fruit bunch ripeness detection," Bulletin of Electrical Engineering and Informatics, pp. 4168 - 4181, 2024.

[5] A. S. Safiee, Najib, Muhammad Sharfi, Mohamad, Mujahid and Che Zan, Nazriyah, "Integrated Approaches for Palm Oil Fresh Fruit Bunch Ripeness Classification Using Near-Infrared Optical Sensors, Artificial Data and Case-Based Reasoning," in 6th Symposium on Intelligent Manufacturing and Mechatronics, 2026.

[6] I. Sa, Lim, Jongyoon, Ahn, Hoseok and Macdonald, Bruce, "deepNIR: Datasets for Generating Synthetic NIR Images and Improved Fruit Detection System Using Deep Learning Techniques," Sensors, vol. 22, no. 13, 2022.

[7] R. R. Hutapea, Hanun, Taqiyya and Irwansyah, Edy, "Comparison of Palm Oil Fruit Ripeness Detection and Classification Methods Using Deep Learning," ICIC Express Letters, vol. 16, no. 7, pp. 797 - 804, 2025.

[8] S. C. Valentine, Alfred, Rayner, Fui, Florence Sia, Shamrie Sainin, Mohd and Iswandono, Zamhar, "Classification of Oil Palm Fresh Fruit Bunches (FFB) Based on Its Maturity Colour Using Convolutional Neural Network (CNN) Approach," in 9th International Conference on Computational Science and Technology, 2022.

[9] P. Laddawong, Pianroj, Yutthapong, Chotikawanid, Piyanart, Punvichai, Teerasak, Jumrat, Saysunee, Kham-Ouam, Atitaya and Muangprathub, Jirapond, "Federated Learning for Grading Oil Palm Fruit Ripeness in the Oil Palm Industry," Applied Computational Intelligence and Soft Computing, vol. 1, 2025.

[10] M. Rosbi and Omar, Zaid, "Segmentation of Outdoor Oil Palm Fruit Images for Ripeness Detection using YOLO Algorithm and Features Thresholding," in Sustainable and Integrated Engineering International Conference, 2022.

[11] M. H. Junos, Mohd Khairuddin, Anis Salwa, Abu Talip, Mohamad Sofian, Kairi, Muhammad Izhar and Siran, Yosri Mohd, "Improved hybrid feature extractor in lightweight convolutional neural network for postharvesting technology: automated oil palm fruit grading," Neural Computing and Applications, vol. 36, no. 32, pp. 20473 - 20491, 2024.

[12] Z. Ünal, "Image preprocessing techniques applied on nir images for fruit bruise detection," in 3rd International Conference on Research of Agricultural and Food Technologies, 2024.

[13] N. I., Seminar K.B, Irzaman and Budiastra I.W., "Improving the accuracy of near-infrared (NIR) spectroscopy method to predict the oil content of

oil palm fresh fruits," in IOP Conference Series: Earth and Environmental Science, 2020.

[14] C. Sun, Zhang, Lei, Zhai, Lixiang, Shen, Tingting, Cai, Jianrong, Zou, Xiaobo and Guo, Zhiming, "Automatic early bruise detection in strawberry fruit by hyperspectral imaging and deep learning techniques," Postharvest Biology and Technology, vol. 232, 2026.

[15] B. O.M., Shariff A.M, Mahmud A.B., Shafri H. and Alfatni M, "Oil palm fruit grading using a hyperspectral device and machine learning algorithm," in 7th IGRSM International Conference and Exhibition on Remote Sensing and GIS, 2014.

[16] X. Yu, Liu, Shuangyin and Wang, Chenjiaozi, "Detection of fungal disease in citrus fruit based on hyperspectral imaging," Information Processing in Agriculture, vol. 12, no. 4, pp. 456 - 465, 2025.

[17] R. M. Desai, Bhole, Varsha Yogesh and Deshpande, Manoj M., "Generative Adversarial Networks and Their Variants for Precise Fruit Recognition," in 4th International Conference on Sustainable Expert Systems, 2024.

[18] S. V and Kasturi K, "Structural Invariant Feature Segmentation Based Apple Fruit Disease Detection Using Deep Spectral Generative Adversarial Networks," SN Computer Science, vol. 5, no. 5, 2024.

[19] D. Doria Usta, Hundelshaussen, Ricardo, Martínez López, Carlos, Salgado Chamorro, Delio, López Martínez, César, Coimbra Leite Costa, João Felipe and Arcari Bassani, Marcel, "From RGB to Synthetic NIR: Image-to-Image Translation for Pineapple Crop Monitoring Using Pix2PixHD," Technologies, vol. 13, no. 12, 2025.

[20] Q. Wang, Lu, Jinzhu, Wang, Yuanhong and Miao, Fajun, "In situ nondestructive identification of citrus fruit ripeness via hyperspectral imaging technology," Plant Methods, vol. 21, no. 1, 2025.

[21] Q. Jiang, Zhou, Tao, Yu, Dongjian, Song, Yupeng, Yao, Shaowen, Wang, Punming and Jin, Xin, "GSU-Fusion: A Text-guided Symmetric U-Net Framework for Infrared and Visible Image Fusion," Information Fusion, vol. 127, 2026.