

# Deep Learning for Ripeness Grading of Oil Palm Fresh Fruit Bunches: A Comprehensive Review of Convolutional Neural Network Approaches

Wahyu Nurkholis Hadi Syahputra<sup>1,2</sup>, Chatchawan Chaichana<sup>2,\*</sup>, Damorn Bundhurat<sup>2</sup>,  
Patiwet Wuttisarnwattana<sup>3</sup> and Bayu Taruna Widjaja Putra<sup>4</sup>

<sup>1</sup>Graduate PhD Degree Program in Mechanical Engineering, Department of Mechanical Engineering,  
Faculty of Engineering, Chiang Mai University, Chiang Mai 50200, Thailand

<sup>2</sup>Department of Mechanical Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai 50200, Thailand

<sup>3</sup>Department of Computer Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai 50200, Thailand

<sup>4</sup>Laboratory of Precision Agriculture and Geoinformatics, Faculty of Agricultural Technology, University of Jember,  
Jember 68121, Indonesia

(\*Corresponding author's e-mail: [chatchawan.c@cmu.ac.th](mailto:chatchawan.c@cmu.ac.th))

Received: 31 December 2025, Revised: 12 February 2026, Accepted: 25 February 2026, Published: 30 March 2026

## Abstract

Palm oil is a strategic agricultural crop in Indonesia, Malaysia, and Thailand, contributing significantly to national economies and requiring continuous improvements in harvesting efficiency and mill operations. The growing demand for higher efficiency and consistent quality in palm oil mills has accelerated the adoption of advanced technologies, particularly artificial intelligence (AI), which is increasingly applied across agricultural sectors, including oil palm production. This review aims to examine the development of convolutional neural network (CNN)-based approaches for ripeness grading of oil palm fresh fruit bunches (FFB) using CNN techniques. It provides an overview of research trends and technical progress in this field, showing that Malaysia leads scientific publications related to palm oil ripeness detection, followed by Indonesia. Most existing studies employ 1-stage object detectors, especially YOLO-based architectures, due to their real-time capability and relatively high performance. However, these methods are often trained and evaluated using datasets limited to specific environments, plantation conditions, or fruit varieties, which constrains generalization and large-scale deployment. Key research gaps are identified, including limited dataset diversity, high computational requirements, insufficient integration with Internet of Things (IoT)-based plantation and mill management systems, and the lack of real-time estimation of quality indicators such as free fatty acid (FFA) content and kernel-related attributes. Future research directions highlight the need for multimodal sensing, multi-camera systems, and multi-task learning frameworks that integrate ripeness grading with oil extraction rate (OER) estimation to support more effective operational decision-making in palm oil production systems.

**Keywords:** Artificial intelligence, Image-based sensing, Industrial deployment, Plantation automation, Precision agriculture, Quality assessment, Sustainable palm oil

## Introduction

Palm oil has been considered as one of the most influential crops in the global bioeconomy, driven by its exceptionally high oil yield per hectare, broad functional properties, and cost-effective production compared to other oil-seed crops such as soybean, sunflower, and rapeseed [1]. It is now the most widely produced and

consumed vegetable oil, supplying a substantial proportion of global edible oil demand while also serving as a key raw material for oleochemical industries, pharmaceuticals, cosmetics, and increasingly for biodiesel and renewable energy applications. The dominance of palm oil in international markets is underpinned by the production capacity of Indonesia,

Malaysia, and Thailand, which together account for more than 85% of global output [2]. Indonesia remains the largest producer and exporter, contributing nearly half of the world's supply and playing a pivotal role in meeting global demand. Malaysia, as the second-largest producer, significantly influences international trade flows and sustainability certification frameworks through its well-established plantation and processing sectors. Thailand, ranking third, continues to strengthen its position through expanding smallholder participation, improved cultivation practices, and investments in downstream palm oil industries [3,4]. The socio-economic reliance on palm oil across these countries highlights not only its strategic importance to rural livelihoods and national export revenues but also the pressing need for technological innovations that enhance plantation productivity, ensure consistent fruit quality, and support sustainable agricultural development.

The ripeness stage of oil palm fresh fruit bunches (FFBs) is a critical determinant of both oil yield and oil quality, as the biochemical composition of the mesocarp undergoes substantial changes during maturation, influencing the accumulation of palm oil and the formation of free fatty acids (FFA) [5]. Harvesting FFBs at the optimal ripeness level maximizes oil extraction efficiency and ensures compliance with quality standards required by the food and oleochemical industries [6]. Underripe bunches contain lower oil content and higher moisture, reducing processing efficiency, whereas overripe bunches exhibit elevated lipolytic activity, leading to increased FFA concentrations that compromise the stability and commercial grade of crude palm oil (CPO) [7]. Grading practices predominantly rely on visual assessment of external indicators such as fruit exocarp color progression, the detachment of loose fruits, and the dryness of spikelets [8]. These criteria, while widely adopted across plantations, are inherently variable due to differences in cultivar physiology, environmental conditions, and harvesting systems, and their subjective interpretation often results in inconsistent grading decisions. The reliance on manual inspection also imposes considerable labor demand and operational fatigue, particularly in large-scale estates, making accurate and standardized ripeness evaluation a continuing challenge for the oil palm industry.

Manual ripeness grading of oil palm FFBs is affected by substantial variability arising from differences in human visual perception, grader experience, and environmental conditions in the field [9]. The evaluation process relies primarily on external indicators such as exocarp color transitions, the presence of loose fruits, and spikelet dryness, yet these visual indicators are highly sensitive to illumination changes, background complexity, and the natural heterogeneity of FFB morphology [10,11]. Considerable grader-to-grader and within-grader inconsistencies have been reported, reflecting the subjective nature of manual assessment and the difficulty of achieving uniform interpretation across plantation operations. These limitations are further intensified by the high labor demand required to inspect large volumes of FFBs daily, often under time pressure and physically demanding conditions, contributing to operational fatigue and reduced accuracy. Variations among oil palm varieties, uneven ripening patterns across the surface of the bunch, and differing harvesting practices add further complexity, challenging the reliability of visual grading systems. The lack of standardized grading protocols across estates and regions contributes to inconsistencies in fruit quality entering processing mills, thereby affecting oil extraction efficiency and final product quality.

The rapid advancement of computer vision (CV) and artificial intelligence (AI) has catalyzed significant progress in automated agricultural imaging, enabling more precise and data-driven approaches to crop monitoring, quality assessment, and decision support [12-14]. Within this technological landscape, deep learning (DL) has emerged as the dominant paradigm due to its capacity to model complex, nonlinear relationships in visual data with minimal reliance on handcrafted features [15]. Among DL methods, convolutional neural networks (CNNs) have demonstrated remarkable effectiveness in extracting multilevel spatial features directly from raw images, allowing robust discrimination of subtle morphological or color variations that are often imperceptible to conventional image-processing techniques [16]. The application of CNNs has expanded rapidly across agricultural domains, including fruit maturity estimation, disease identification, stress detection, weed recognition, and yield forecasting, reflecting their strong

generalization capabilities under diverse field conditions characterized by occlusion, irregular lighting, and heterogeneous backgrounds [17-20]. CNNs leverage large annotated datasets, hierarchical feature representations, and end-to-end learning structures that collectively enhance their ability to address the inherent variability present in natural agricultural environments [21]. The increasing incorporation of CNN-based imaging solutions into precision agriculture frameworks underscores a broader transition toward automation, real-time monitoring, and intelligent decision-making across modern crop production systems.

Research on the application of CNNs for automated ripeness classification of oil palm FFBs has expanded considerably over the past decade, reflecting growing interest in deploying DL-based vision systems within plantation environments. Early studies primarily adopted established CNN architectures such as AlexNet, VGGNet, and GoogLeNet to classify FFB ripeness stages using images captured under controlled conditions, demonstrating the feasibility of end-to-end feature learning for distinguishing subtle color transitions associated with fruit maturation [21-24]. Subsequent work incorporated more advanced architectures, including ResNet, Inception, DenseNet, MobileNet, and EfficientNet, which provided improved accuracy, reduced computational load, and enhanced robustness to variations in lighting, occlusion, and fruit orientation commonly encountered in field settings [25,26]. In addition to static image classification, several studies explored object detection frameworks such as YOLO and Faster R-CNN to simultaneously localize and classify FFBs, enabling applications in automated harvesting systems and real-time field monitoring [27,28].

Despite the substantial progress achieved through the application of CNNs for automated ripeness grading of oil palm FFBs, several critical research gaps remain across methodological, operational, and evaluative dimensions. A major limitation is the lack of large, diverse, and standardized datasets that capture variations in cultivar, geographic location, seasonal effects, and field illumination, which constrains model generalizability beyond the conditions represented in training data. Many studies rely on narrowly scoped datasets with limited environmental diversity, often

collected under controlled or semi-controlled conditions, resulting in performance inflation that does not reflect real-world deployment challenges [29,30]. From a processing perspective, the computational demands of many CNN architectures pose challenges for deployment on edge devices commonly used in the field. Furthermore, existing review papers primarily provide high-level summaries of DL applications in agriculture or general imaging techniques for oil palm [15,31].

This review aims to synthesize and critically evaluate existing research on the application of CNNs and vision based for automated ripeness grading of oil palm FFBs, drawing on global developments while also incorporating a dedicated examination of technological progress in Indonesia, Malaysia, and Thailand, the 3 leading palm-oil-producing countries in Southeast Asia. The review analyzes CNN architectures, datasets, imaging methods, annotation practices, and performance metrics reported in previous studies, and assesses their suitability for real-world plantation and mill environments. It also identifies methodological limitations, data constraints, and deployment challenges that affect the scalability and generalizability of current systems. By integrating broader international findings with region-specific contributions from key producer countries, the review establishes a clearer understanding of the current state of FFB ripeness assessment using CNNs and vision-based DL outlines essential research directions for advancing DL-based grading technologies.

### **Research trends and bibliometric in CNN applications for palm oil**

Research on CNN applications in the palm oil sector has increased steadily, showing the growing role of AI and CV in agricultural engineering. A bibliometric approach allows this development to be examined in a systematic and objective manner by analyzing publication volume, citation impact, and collaboration patterns. The bibliometric method offers a structured and objective approach to examine research development and to analyze citation relationships among authors, institutions, countries, and academic journals [32]. Bibliographic information is generally obtained from well-established citation databases, such as Scopus and Web of Science (WoS), which supply

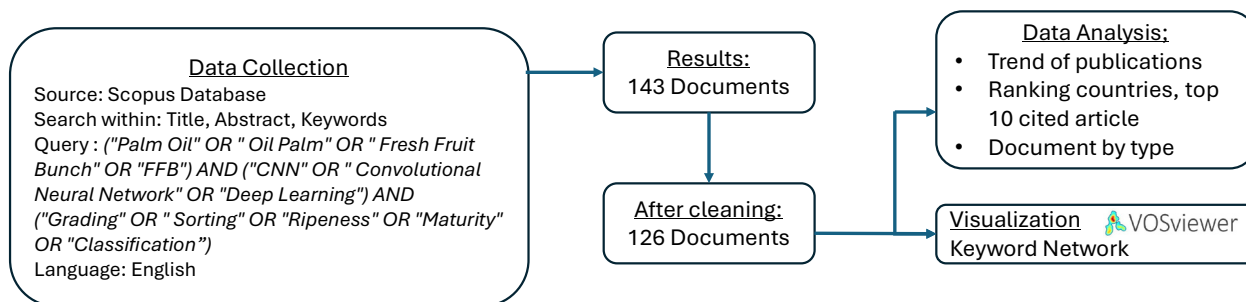
comprehensive publication and citation records required for reliable quantitative assessment [33]. In this study, Scopus was selected because it provides broader coverage of interdisciplinary research in agricultural engineering, AI, DL and CNN, and indexes more conference proceedings where many technological innovations are first reported. Scopus also offers more consistent metadata for bibliometric mapping [34], while access to local agricultural repositories is often limited. Furthermore, several reputable local agricultural publications are already indexed in Scopus. Through science mapping techniques, the relationships and interaction strength among different elements of scholarly publications can be explored. These techniques include co-authorship analysis to evaluate collaboration patterns, keyword co-occurrence analysis to identify dominant and emerging research topics, and citation analysis to determine influential studies and sources. To support interpretation of these complex relationships, network-based methods, including clustering and visualization, are commonly applied to reveal structural patterns and thematic connections within the research field. Among the available tools, VOSviewer is widely used for bibliometric visualization due to its ability to construct and analyze networks based on co-authorship, citation, and bibliographic coupling [35].

In this study, a bibliometric analysis was conducted to examine research trends on CNN applications for oil palm FFB grading and ripeness assessment. The overall workflow of the bibliometric analysis in this study is illustrated in **Figure 1**. The literature dataset was retrieved from the Scopus database using a structured search query as follows:

*(TITLE-ABS-KEY (“Palm Oil” OR “Oil Palm” OR “Fresh Fruit Bunch” OR “FFB”) AND TITLE-ABS-KEY (“CNN” OR “Convolutional Neural Network” OR “Deep Learning”) AND TITLE-ABS-KEY (“Grading” OR “Sorting” OR “Ripeness” OR “Maturity” OR*

*“Classification”)) AND (LIMIT-TO (LANGUAGE, “English”))*

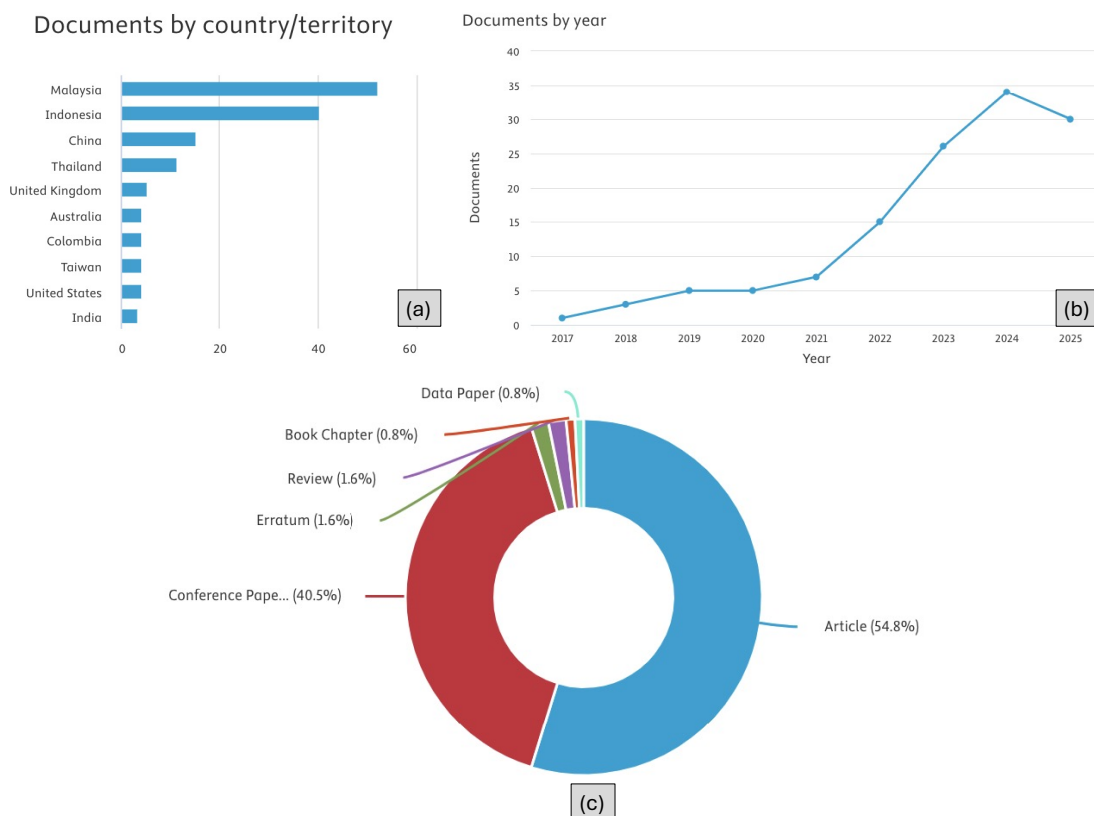
The Scopus search query was structured to comprehensively capture publications related to DL-based ripeness assessment of oil palm FFBs while minimizing irrelevant records. The first component of the query targeted the application domain by including key terms related to palm oil and FFBs, namely “Palm Oil”, “Oil Palm”, “Fresh Fruit Bunch”, and “FFB”, ensuring coverage of different terminologies commonly used in the literatures. The second component focused on methodological approaches by incorporating deep learning related keywords, including “CNN”, “Convolutional Neural Network”, and “Deep Learning”, which allowed retrieval of studies employing modern neural network architectures. The third component addressed the application objective by specifying terms associated with grading and maturity evaluation, such as “Grading”, “Sorting”, “Ripeness”, “Maturity”, and “Classification”. All terms were searched within titles, abstracts, and author keywords to maximize retrieval sensitivity. Finally, the query was restricted to English-language publications to ensure consistency in analysis and interpretation. The workflow began with an initial Scopus retrieval yielding 143 documents, followed by content-level screening to exclude records using CNNs for non-grading purposes such as oil-palm tree detection or plantation mapping, resulting in 126 relevant studies, with **Figure 2** detailing the reduction summary alongside publication contribution by country, Scopus-labelled document families (journal articles, conference papers, reviews, book chapters, and other indexed types), and publication counts per category, after which keyword co-occurrence analysis was performed in VOSviewer to generate the thematic network of research clusters and term linkages shown in **Figure 3**, supporting identification of major research concentrations, topical focus, and emerging directions in CNN-based oil-palm FFB ripeness grading.



**Figure 1** Bibliometric analysis workflow.

**Figure 2** summarizes the publication results obtained from the Scopus database using the defined search query on CNN-based ripeness grading of oil palm FFBS. The distribution by country shows that Malaysia and Indonesia account for the largest share of publications, which is consistent with their positions as major palm oil producers and active research hubs in oil palm technology. Contributions from China and Thailand indicate increasing research involvement from other regions, while publications from Europe, the United States, and additional countries appear at a lower frequency, reflecting more limited engagement in this specific application domain. The annual distribution of documents shows that the trend begins around 2017, which corresponds with the shift from traditional image-processing methods to CNN-based approaches enabled by more accessible deep-learning frameworks and affordable graphical processing units (GPU) computing during 2016 - 2017. This observation also aligns with findings in previous report, which indicate that the adoption of deep learning in the palm-oil sector began to grow more noticeably around 2018 [36]. The observed rise continues gradually until 2020 and then

grows more strongly after 2021, matching the rapid adoption of DL and object-detection techniques in agricultural and computer-vision research. The peak observed around 2024 suggests heightened research activity in recent years, while the slight decrease in the most recent year is likely influenced by incomplete indexing of newly published articles in the Scopus database. The distribution by document type shows that journal articles represent the largest proportion of retrieved records. The increasing publication trend also reflects the gradual industrial uptake of automation technologies in major producing nations such as Malaysia, Indonesia, and Thailand, where larger plantations have begun adopting machine-vision systems for grading and monitoring. However, adoption among small- and medium-scale growers remains limited due to socio-economic barriers, including high initial investment costs, restricted access to technical expertise, and uneven digital infrastructure [37,38]. These contextual factors help explain why research output has grown rapidly while practical implementation still varies across stakeholder groups.



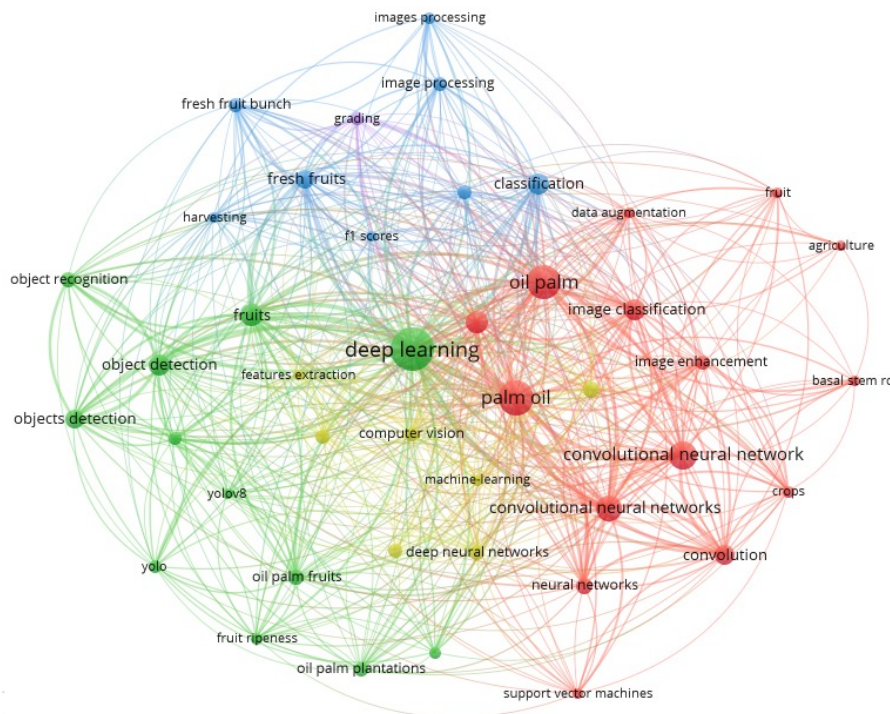
**Figure 2** Distribution of Scopus search results for CNN-based oil palm FFB ripeness research by (a) country; (b) publication year, and (c) document type.

**Figure 3** depicts the keyword co-occurrence network generated using VOSviewer from Scopus records on CNN-based ripeness grading of oil palm FFBs. In this visualization, each node represents a keyword, the node size reflects its frequency of occurrence, and the links indicate co-occurrence relationships between keywords within the same publications. Keywords are grouped into clusters based on their co-occurrence strength, with different colors representing distinct thematic clusters. The red cluster is centered around methodological terms such as *palm oil*, *convolutional neural network*, and *classification*, indicating that CNN-based approaches form the core research focus in this domain. Closely connected terms such as *image enhancement*, *feature extraction*, and *machine learning* suggest strong methodological integration within computer vision pipelines for ripeness assessment. The green cluster is dominated by application-oriented keywords, including *deep learning*, *object detection*, *fruit ripeness*, *fresh fruit bunch*, highlighting research emphasis on detection and maturity evaluation at the field and bunch levels. This cluster reflects practical deployment goals such as

localization and grading under plantation conditions. The blue cluster is associated with data acquisition and processing aspects, including *image processing*, *classification*, and *fresh fruits*, indicating foundational steps in visual data preparation and analysis. The presence of terms related to *harvesting* and *oil palm plantations* across clusters demonstrates the close linkage between algorithm development and agricultural application contexts. Figure 3 also shows that more specialized architectural terms such as “1-stage detector” or “lightweight network” do not appear in the co-occurrence map. This absence reflects the historical development of the field, where early publications relied mainly on general CNN-based classification methods, which generate high-frequency keywords and therefore dominate the clusters. Because lightweight models and advanced detection frameworks were adopted more widely only in later years, their lower keyword frequency prevents them from forming visible clusters in **Figure 3**. One-stage detectors gained earlier attention due to their strong performance and the availability of open-source implementations that were easy to deploy in plantation settings. Lightweight

networks appeared more slowly because they require more extensive architectural redesign and optimization for low-power devices [39]. This contrast shows that the

research community tends to adopt readily deployable models before shifting toward more efficiency-oriented architectures.



**Figure 3** Keyword co-occurrence network generated using VOSviewer based on Scopus search results.

Following the analysis of publication output, country contributions, document types, and keyword co-occurrence derived from the Scopus database, **Table 1**

top ten most cited publications were identified to highlight influential studies in CNN-based ripeness grading and detection of oil palm FFBS.

**Table 1** Top 10 most cited publications on CNN-based ripeness grading and detection of oil palm FFBS retrieved from the Scopus database.

No	Author	Title	Year	Source	Citation	Finding / Contribution	Ref.
1	Mamat <i>et al.</i>	Enhancing Image Annotation Technique of Fruit Classification Using a Deep Learning Approach	2023	Sustainability Switzerland	120	This study applied YOLO-based DL for automatic image annotation in oil palm fruit ripeness classification, attaining 98.7% mAP.	[40]
2	Suharjito <i>et al.</i>	Oil palm tree detection and health classification on high-resolution imagery using deep learning	2021	Computers and Electronics in Agriculture	90	This study developed a mobile-based oil palm FFB ripeness classifier using lightweight CNNs, achieving the best performance with EfficientNetB0 at 0.893 accuracy on TensorFlow Lite with 96 ms inference time per image.	[41]
3	Ibrahim <i>et al.</i>	Palm oil fresh fruit bunch ripeness grading recognition using convolutional neural network	2018	Journal of Telecommunication Electronic and Computer Engineering	66	This study compared CNN-based and hand-crafted feature methods for oil palm FFB ripeness grading. The pre-	[42]

No	Author	Title	Year	Source	Citation	Finding / Contribution	Ref.
						trained AlexNet model achieved the best performance when training data were limited.	
4	Lai <i>et al.</i>	Real-Time Detection of Ripe Oil Palm Fresh Fruit Bunch Based on YOLOv4	2022	IEEE Access	41	This study implemented YOLOv4 for real-time detection of FFBS as part of a robotic harvesting system using images captured at 1920 × 1080 resolution. The model shows a mAP of 87.9%, a recall of 82% at IoU greater than 0.5 after 2000 iterations, and operated at approximately 21 FPS	[43]
5	Herman <i>et al.</i>	Deep learning for oil palm fruit ripeness classification with Densenet	2021	Proceedings of 2021 International Conference on Information Management and Technology Icimtech	34	Using a dataset of 7 ripeness levels with 400 images, the results show that DenseNet outperforms AlexNet by 8.5% in accuracy and 8% in F1-score.	[44]
6	Lai <i>et al.</i>	Oil Palm Fresh Fruit Bunch Ripeness Detection Methods: A Systematic Review*	2023	Agriculture Switzerland	31	Based on 51 reviewed papers covering 11 unique approaches, the study finds that combining CV and DL is the most feasible on-field method, offering noncontact operation, low cost, real-time capability, and high accuracy.	[5]
7	Suharjo <i>et al.</i>	Annotated Datasets of Oil Palm Fruit Bunch Piles for Ripeness Grading Using Deep Learning	2023	Scientific Data	30	This study introduces a novel annotated FFB dataset collected from palm oil mills. The dataset comprises 45 single-category videos and 56 multi-category videos at 1,280×720 px in .mp4 format, with labels across 6 maturity categories.	[45]
8	Suharjo <i>et al.</i>	Real-Time Oil Palm Fruit Grading System Using Smartphone and Modified YOLOv4	2023	IEEE Access	25	The YOLOv4 model with 16 quantization achieves 12% higher mAP than YOLOv4 Tiny and can efficiently detect FFBS that do not meet quality standards in real grading conditions.	[46]
9	Mohd Basir Selvam <i>et al.</i>	Real Time Ripe Palm Oil Bunch Detection using YOLOv3 Algorithm	2021	19th IEEE Student Conference on Research and Development Sustainable Engineering and Technology Towards Industry Revolution Scored 2021	17	This study presents a real-time FFB detection system using the YOLOv3. The model successfully differentiates ripeness levels, reaching learning saturation at the 6,000 <sup>th</sup> iteration and showing potential for mobile and IoT integration.	[47]
10	Wong <i>et al.</i>	Computer vision algorithm development for classification of	2020	AIP Conference Proceedings	17	Implementing a CV algorithm for FFB classification on a low-cost portable device using a	[48]

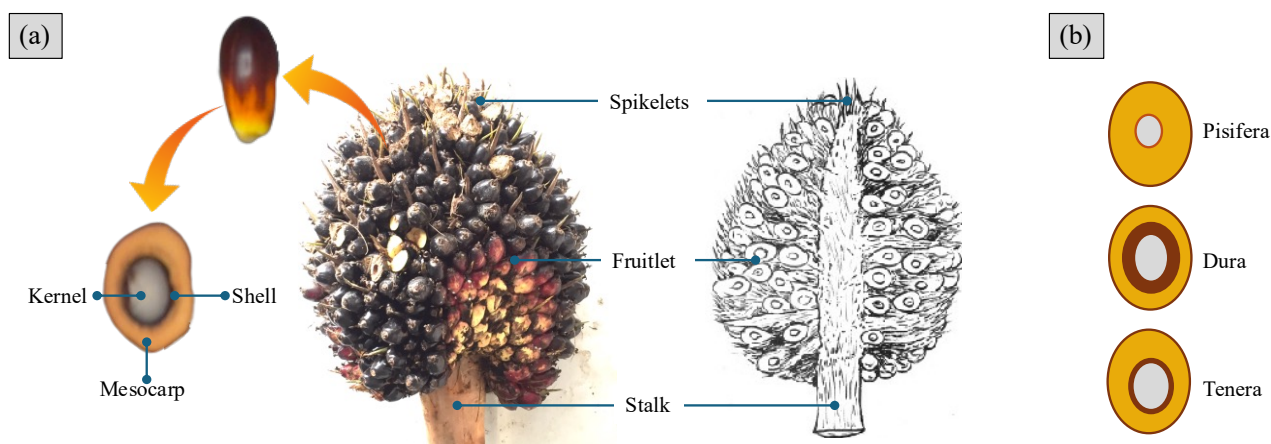
No	Author	Tittle	Year	Source	Citation	Finding / Contribution	Ref.
		palm fruit ripeness				retrained AlexNet CNN and HSV color analysis. Using 100 palm fruit images from different ripeness categories, the system achieves 85% accuracy with 85 correctly classified images and is deployed on a Tinker Board equipped with a GPU and camera.	

Remark: \*Review paper

### Oil palm FFB ripeness indicators

The ripeness of oil palm FFBs is influenced by a series of external and easily observable morphological changes that appear as the bunch matures. As shown in **Figure 4(a)**, an FFB consists of a central bunch stalk with many spikelets, each containing a densely packed group of fruitlets [49]. The structural characteristics of the bunch vary depending on palm type, particularly among the commonly cultivated *dura*, *tenera*, and *pisifera* varieties as shown in **Figure 4(b)**. *Dura* palms

produce fruitlets with thick shells and lower oil content. *Pisifera* palms produce shell less fruitlets that are mostly infertile. *Tenera* palms, obtained from a *dura* and *pisifera* cross, are the dominant commercial type because they have thinner shells and higher oil yield. Across these varieties, a typical mature FFB weighs approximately 10 to 40 kg, contains around 1,000 to 2,000 fruitlets, and individual fruitlets usually weigh between 6 and 20 g [50,51].



**Figure 4** (a) Structural components of an oil palm FFB, showing the bunch stalk, spikelets, and fruitlets, and (b) oil palm varieties, including *dura*, *tenera*, and *pisifera*.

The most widely used indicator of ripeness is the change in fruitlet skin color. The exocarp gradually shifts from dark purple or black in immature stages to orange and red hues at optimal maturity. This visible transformation is a practical indicator for field graders, although it is influenced by lighting conditions, fruit variety, and surface contamination. Another important ripeness signal is the natural detachment of fruitlets from the spikelets. As the bunch matures, the abscission zone weakens, and more loose fruits appear around the tree. Loose fruit count is commonly used in plantations,

but it can be affected by harvesting delays, strong winds, or handling, and it may not always represent the average maturity of all fruitlets inside the bunch.

Graders also observe the dryness and brittleness of spikelets. Immature bunches usually have firm and moist spikelets, while ripe bunches tend to show drier and more fragile structures. Spikelet morphology is shaped not only by maturity stage but also by external conditions, including rainfall, humidity, and sunlight exposure. Internal biochemical indicators, including oil content and FFA concentration, provide more precise

information on physiological maturity, but these require destructive sampling and laboratory analysis, making them unsuitable for routine grading in plantation conditions. Ripening does not occur uniformly throughout the bunch. Fruitlets located on the outer surface usually mature earlier than those deeper inside

the spikelet clusters, which further complicates manual evaluation. **Table 2** presents the standard ripeness indicators adopted by the Malaysian Palm Oil Board (MPOB) and the Indonesian Oil Palm Research Institute (IOPRI), which are commonly used as reference guidelines in plantation practices.

**Table 2** Standard ripeness indicators for oil palm FFBs according to MPOB and IOPRI [52].

Ripeness Level	Detached fruitlets	FFB Color	Mesocarp color
Under raw	None	Black	Pale
Raw	Up to 12.5% outer fruits	Purple black	Yellowish
Under ripe	12.5% - 25% outer fruits	Reddish purple	Yellowish orange
Ripe	25% - 50% outer fruits	Reddish orange	Orange
Over ripe	75% - 100% outer fruits	Darkish red	Orange
Empty or rotten	Most fruits detached	Withered	None

Although the MPOB standard provides a valuable guideline, its reliance on human inspection introduces considerable variability. Lighting, background contrast, and grader fatigue affect perception of color and detachment, often causing discrepancies between graders assessing the same bunch. Differences in training, cultivar type, and local interpretation of color intensity further amplify subjectivity. These inconsistencies can lead to premature or delayed harvesting, affecting OER and increasing FFA levels. The visual thresholds described in the MPOB standard were designed for manual field conditions, yet they lack the quantitative precision required for automated grading. Translating these qualitative descriptors into machine-interpretable labels poses a major challenge in developing CNN-based systems, which require consistent and numerically reliable annotations for model training and validation

#### Vision modalities for ripeness observation

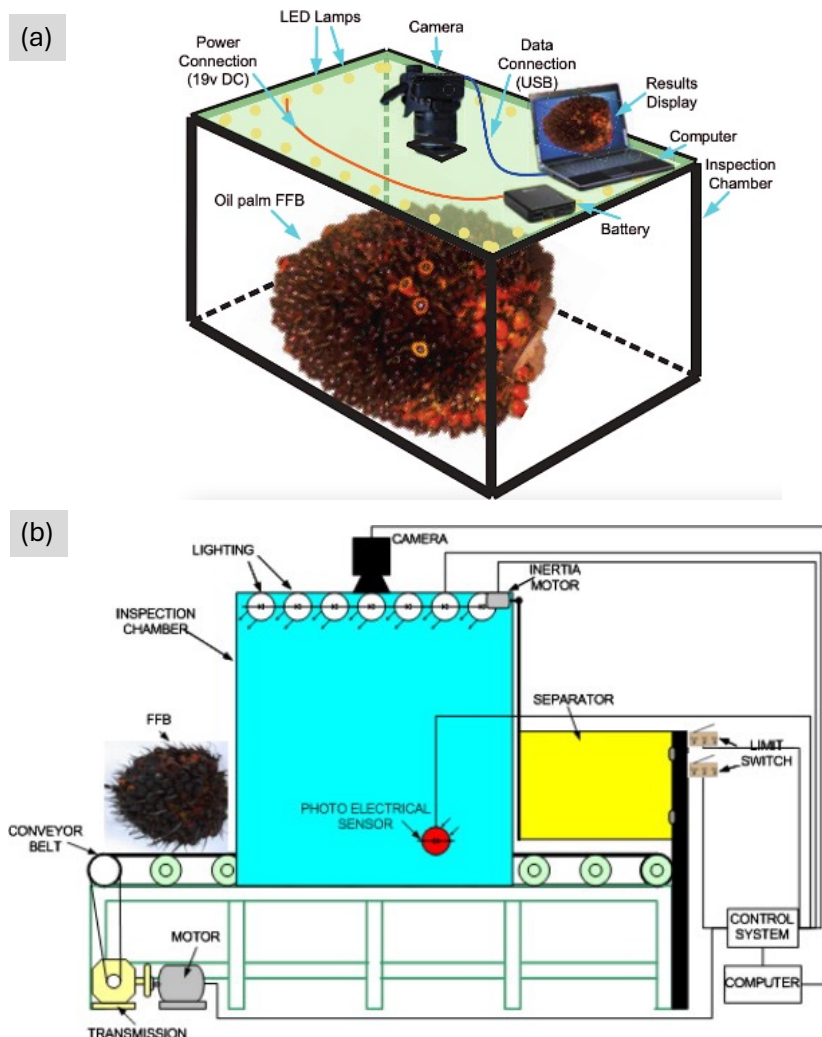
Vision-based observation forms the foundation of automated ripeness assessment in oil palm FFB evaluation, as ripeness is primarily expressed through visible and measurable changes on the fruit surface and bunch structure [53]. Vision modalities refer to the types of imaging systems and visual data sources used to

capture these changes. An imaging system is the physical hardware or sensor configuration used to record FFB appearance. A visual data source is the image dataset generated by the imaging system and used as input for CNN model training, testing, and evaluating. In oil palm ripeness grading, a range of optical and sensing technologies has been employed to capture maturity-related information from FFBs. These sensors include conventional cameras, multispectral and hyperspectral cameras, depth sensors, thermal cameras, and point-based optical instruments such as spectrometers. Spectroscopy is a non-destructive method that analyzes how a sample interacts with light or electromagnetic waves to reveal its chemical and physical properties [54]. For palm-oil sensing, common spectral techniques include Visible-Near Infrared (VIS-NIR) [55], Mid-Infrared (MIR) [56], Raman spectroscopy [57], and fluorescence spectroscopy [58]. Spectrometers are widely used in research to measure reflectance at specific wavelengths and to analyze chemical and physical properties associated with ripening, such as pigment variation and moisture change. However, spectrometers operate at a single point or over a very small surface area and do not capture spatial or structural information across the bunch. For this reason, spectrometers are typically used

as analytical or reference tools to support calibration, wavelength selection, and validation of grading systems, rather than direct vision-based grading instruments. Vision-based modalities, in contrast, acquire spatially resolved images that preserve the arrangement, distribution, and appearance of fruitlets and spikelets across the bunch. These include RGB imaging for surface color and texture observation, multispectral and hyperspectral imaging for spatially distributed spectral analysis, depth imaging for 3-dimensional structure and fruitlet distribution, thermal imaging for surface temperature patterns, and multi-view or aerial imaging for improved coverage and occlusion reduction. The selection of an appropriate modality directly influences the reliability of ripeness interpretation, the robustness of the resulting model, and the feasibility of deployment under plantation or mill conditions. Different modalities capture different aspects of ripeness, ranging from surface color and texture to spatial arrangement and spectral response, each offering specific advantages and limitations.

The most common method is the use of red, green and blue (RGB) cameras [59,60]. RGB images are popular because they capture visual information that graders already use in the field, such as fruit color, loose fruits, and spikelet conditions. RGB images collected in plantations often show variation due to different cameras, backgrounds, and natural surface conditions of the bunches. Pre-processing usually includes removing background, resizing, and adjusting colour balance to make the overall colour appearance more consistent

across samples [61]. Brightness and contrast adjustment may also be applied to stabilize visual information before model training. Several studies on oil-palm FFB imaging report that these steps improve the stability of colour and texture features used for ripeness assessment [62,63]. Study conducted by Makky [60], assessed oil palm FFB ripeness using a portable inspection system similar to the setup shown in the **Figure 5(a)**, which includes a controlled LED-lit chamber, a camera positioned above the bunch, and a computer for RGB color analysis. Using 90 samples, the system achieved 85% correct ripeness classification, produced a strong oil-content prediction ( $R^2 = 0.931$ ;  $SEP = 0.821$ ), and offered faster, lower-cost on-site measurements compared to manual laboratory methods. Alfatni *et al.* [64] developed a real-time grading system that uses image acquisition, segmentation, and colour-feature extraction including histograms and statistical colour metrics to classify ripeness. Using an ANN classifier evaluated with ROC-AUC, their system achieved 94% classification accuracy, showing higher performance than the manual method and outperforming the other colour-feature techniques tested. More recent works used static camera installed in conveyor as shown in **Figure 5(b)**, FFBs move along conveyor belts and require rapid, continuous evaluation [52]. Although RGB imaging is easy to use and inexpensive, the quality of the images is often affected by varying light conditions, shadows, and background elements in the field [65].



**Figure 5** Use of an RGB camera for oil palm FFB ripeness assessment: (a) portable low-cost setup [60], and (b) conveyor-integrated system for continuous grading [52].

Advanced imaging techniques such as Hyperspectral imaging (HSI) and Multispectral imaging (MSI) have shown strong promise for assessing the maturity of oil-palm FFBs by capturing spectral reflectance signals including in near-infrared and visible bands that correlate with internal fruit properties (e.g. oil content, ripeness) beyond what human vision can detect [66-68]. Certain wavelengths especially in the NIR range, have strong relationships with internal maturity and can improve classification performance. An MSI-based grading approach was introduced, capturing oil palm FFB under 19 different LED wavelengths and assessing ripeness using HSV color features. HSI/MSI data often require early cleaning because the large number of wavelength bands can contain random

variation [69]. To address these variations, several preprocessing methods such as normalization, multiplicative scatter correction, and Savitzky Golay smoothing are commonly applied. These techniques help produce cleaner and more stable spectral curves and reduce the overall complexity of high dimensional data [25,70]. The study found that the 940 nm wavelength provided the strongest ripeness discrimination, offering a more objective grading method for improving mill productivity [71]. An hyperspectral approach was also explored to estimate internal qualities of oil palm FFB, focusing on visible, NIR, and reflectance-based indicators to classify ripeness and predict oil content and FFA. Using reflectance values and chlorophyll-carotenoid ratios for

ripeness classification, the method achieved highly accurate predictions of oil content ( $R^2 = 99.7\%$ ,  $SEP = 0.421$ ) and FFA ( $R^2 = 99.5\%$ ,  $SEP = 0.190$ ), further demonstrating the strong potential of spectral imaging for reliable, non-destructive FFB quality assessment [72]. However, HSI and MSI equipment is expensive, sensitive to lighting variations, and not yet practical for large plantation operations. Even so, these techniques show the scientific potential of spectral analysis for more accurate assessment of fruit maturity.

Other imaging technologies beyond spectral/Hyperspectral methods can supply valuable structural or physical information about oil-palm FFBs, helping to overcome limitations of simple color or visual inspection. For example, a study using thermal imaging investigated 297 oil-palm FFBs (sorted as under-ripe, ripe, and over-ripe) and found that the difference between the average bunch surface temperature and ambient temperature ( $\Delta Temp$ ) decreased consistently from under-ripe to over-ripe, and that  $\Delta Temp$  could serve as a reliable maturity index. The authors used this parameter together with ANN and reported high classification accuracy (99.1% training, and 92.5% testing) for maturity categories [73]. In addition, 3-

dimensional (3D) imaging using stereo, or depth sensors has been explored as part of automated or robotic harvesting systems. By providing spatial layout of fruitlets on each spikelet, 3D data helps resolve issues of overlapping fruitlets or occlusion from leaves and branches. Imaging from aerial and mobile platforms is emerging as a promising approach for large-scale monitoring and ripeness classification of oil-palm FFBs. For instance, some studies have used unmanned aerial vehicles (UAVs) equipped with RGB or MSI sensors to capture overhead images of palm plantations. These UAV-based systems can detect loose fruits or fruit bunches from above canopy and help locate ripe bunches over large areas, supporting early harvest planning and plantation-wide monitoring rather than relying solely on spot checking [22]. On the ground, robotics and automated vehicles fitted with cameras and multispectral imagers have also been developed; combining navigation capability with ripeness detection along plantation paths, these mobile platforms enable continuous inspection of trees under real plantation conditions [74]. **Table 3** provide the summary of vision and optic-based sensor for detecting the palm oil ripeness for the grading purpose.

**Table 3** Summary of several study in palm oil grading using vision-optic-based sensing.

No	Sensor	Objective	Method	Results	Region	Ref.
1	MSI	Real-time detection of oil palm FFB ripeness; evaluated Oil Content (OC) and FFA	Tenera; 2,000 images and 30 videos; identified 2 classes (ripe and unripe); using PCA, ANN model, PLS, YOLO model; Soxhlet extraction method	the average FFB detection was 2 seconds; Acc images = 98.60%; Acc videos = 99.66%.	Indonesia	[75]
2	HSI (Imperx IPX-2 M 30)	Classify oil palm FFB ripeness	3 species (nigrescens, virescens, oleifera); 3 category detection (underripe, overripe, and ripe); 469 samples; split data (75:25); evaluation ANN	Classification accuracy: Nigrescens = 94.54%, Virescens = 98.67%, Oleifera = 97.89%; overall = 98.67%	Malaysia	[68]
3	HSI	Real-time classification and counting of oil palm FFB ripeness	440 record video samples; 5 category detection (unripe, underripe, ripe, flower, and abnormal); split data 70% (train) 20%(test) 10%(val); augmentation; DL algorithm (YOLOv6, YOLOv7, YOLOv8, Faster	YOLOv8s Depthwise <ul style="list-style-type: none"> <li>• mAP50 = 0.75</li> <li>• mAP50-95 = 0.481</li> <li>• MAE = 0.164</li> <li>• RMSE = 0.4</li> </ul> Inference time = 0.027s	Indonesia	[76]

No	Sensor	Objective	Method	Results	Region	Ref.
			R-CNN, SSD MobileNetV2, EfficientNet; evaluation mAP50, mAP50-95, MAE, RMSE, Inference Time			
4	RGB camera	To detect FFB ripeness	100 image samples; 3 levels category; DL analysis (CNN AlexNet)	85% classification accuracy	Malaysia	[48]
5	Camera (smartphone samsung A50)	To classify oil palm FFB maturity levels	240 samples image; 3 level detection (raw, ripe, half-ripe); classifier6model naïve bayes, SVM, and ANN; confusion matrix	Best model: ANN Accuracy: 98.3% Precision: 98.4% Recall: 98.3%	Indonesia	[77]
6	Camera (Tefcon, webcam 2.0 16MP, Taiwan)	To develop and evaluate an on-site automatic grading machine for FFB	Tenera species; detection category (Under raw bunch, Raw bunch, Under ripe bunch, Ripe bunch I, Ripe bunch II, Over ripe bunch I, Over ripe bunch II, Empty); Regression for weight estimation	Grading accuracy: 93.53% Fraction classification: 88.7% Weight estimation R <sup>2</sup> : 0.96	Indonesia	[78]
7	Camera (thermal camera)	To observe oil content (OC) oil palm	Tenera variety; 5 level FFB; soxhlet method as validation; ANN-MLP model; analysis regression (R <sup>2</sup> , SEC, SEP)	R <sup>2</sup> = 0.9058 (OC x ripeness) R <sup>2</sup> = 0.8039 (OC x temp). R <sup>2</sup> train = 0.7818 SEC = 0.0831. R <sup>2</sup> validation = 0.9535 SEP = 0.0003.	Indonesia	[79]
8	Camera (0 Lux, 1.0 M pixels, F1.8)	To develop an automated ripeness classification FFBS	Pisifera; 4 level (unripe, underripe, ripe, overripe), 208 samples; PCA for reduce features; ANN-MLP models	From 59 features to 6 features Acc = 93.3% (6 features)	Malaysia	[80]
9	Thermal camera (Pseudo color thermal vision)	To correlate surface temperature of oil palm fruits with ripeness level and storage condition	Ripeness level (ripe and unripe); Level storage (120, 140, 160, 180, 200 DAA); Days Storage (-20 °C) (200, 200, 210, 180, 90); analysis statistic (R <sup>2</sup> )	R <sup>2</sup> = 0.973 - 0.979 with thermal vision R <sup>2</sup> (R) = 0.8037 (180 DAA, -180 °C) R <sup>2</sup> (G) = 0.8574 (200 DAA, -90 °C) R <sup>2</sup> (B) = 0.5610 (120 DAA, -200 °C)	Indonesia	[81]
10	Camera CCD (digital camera DFK 41BF02.H FireWire)	To develop a real-time, non-destructive FFB maturity classification	270 image samples; 3 types (Nigrescens, Oleifera and Virescens); 3 level maturity (under ripe, ripe and over ripe); feature extract technique (statistics, histogram, Gabor wavelets, GLCM, BGLAM); built	Best model: ANN + BGLAM Acc = 93% Time detect = 0.44 s/image	Malaysia	

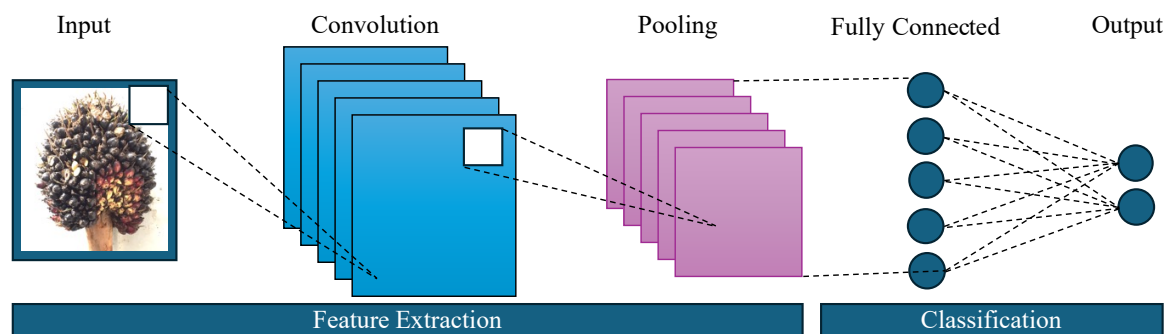
No	Sensor	Objective	Method	Results	Region	Ref.
			best models ML (KNN, SVM, ANN); confusion matrix as evaluated models			
11	Camera (digital camera by smartphone)	To predict maturity stage of oil palm FFBs and create mobile application	Level ripeness (completely ripe, medium ripe, and unripe); using graphics RGB; preprocessing (brightness, deblurring, and remove background)	Accuracy 96%	Thailand	[63]

A balanced evaluation of vision modalities requires considering both the amount of spectral information they provide, and the computational effort required to process them. HSI/MSI systems generate many spectral bands, which increases data size and processing time, and several studies report that full-spectrum models often depend on GPU acceleration to reach practical inference speeds [70]. However, the performance seen in laboratory conditions does not always transfer to plantations or mill environments. In real-world settings, spectral signatures become less stable because fruit orientation, surface moisture, dust, and small changes in sensor position introduce variability that is not present during controlled data collection [71]. CNN architectures also respond differently to illumination and chromatic variation [82]. Deeper backbones may extract more robust semantic features, but high dynamic range and shifting colour temperatures can still degrade performance unless extensive augmentation or adaptive normalization is used [83]. As spectral resolution increases, the additional data and higher model complexity slow processing and limit feasibility for real time use. Compared with HSI/MSI, RGB imaging is more practical for real-time mill applications because it produces smaller data volumes and supports fast inference on embedded devices [5]. Studies in high-throughput fruit sorting show that RGB-based CNNs maintain reliable classification accuracy when illumination is standardized, and they can operate at conveyor speeds without compromising processing time [84,85]. Although RGB imaging cannot detect early

chemical changes due to its limited spectral range, it performs well for surface-based ripeness feature such as colour progression, texture changes, and fruitlet exposure patterns, which are adequate for mill-level grading.

#### Convolutional neural networks architecture

A convolution-neural-network-model, commonly referred to as a CNN, is a specialized NN architecture widely used in ML and AI for processing image data. The primary motivation behind the development of CNNs was to address the limitations of traditional neural networks in handling spatial structure information, making CNNs particularly effective for visual imagery analysis. The architecture of a CNN is inspired by the connection patterns of nerve cells in the human visual cortex, enabling the model to capture spatial patterns in data. CNNs have become one of the leading approaches for image-based analysis in agriculture, including the assessment of oil palm FFBs ripeness [42]. CNNs can automatically learn important visual features from images, such as color patterns, texture differences, and structural shapes, making them highly suitable for tasks that rely on visual feature [86]. CNNs are a type of DL model designed to process image data. They consist of several key components, such as convolution layers that detect patterns, pooling layers that reduce image size while keeping useful information, and fully connected layers that perform classification. In contrast to traditional image-processing methods, which require manual feature extraction, CNNs automatically learn useful patterns directly from examples [87].



**Figure 6** Main components of a CNN, including convolution and pooling layers for feature extraction and fully connected layers for classification, used as the basis for automated oil palm FFB ripeness assessment.

**Figure 6** presents a general architecture of a CNN, showing the main layers commonly used in image classification tasks. A CNN is typically composed of 3 main types of layers [88]. A CNN is composed of several fundamental building blocks: convolutional layers, pooling layers, activation functions, and fully connected layers, which are typically stacked to form the overall architecture. The convolutional layer is the first layer in CNN, consisting of learnable filters (kernels) that are small in spatial dimensions but extend through the full depth of the input image. For example, a typical filter might have a size of  $5 \times 5 \times 3$ , corresponding to the width, height, and depth (e.g., RGB channels) of the input [89]. In the context of oil palm FFB ripeness assessment, these learned features correspond to visual including fruitlet color variation, surface gloss, and spikelet texture, which are closely associated with maturity progression. The second layer is the pooling layer, which reduces the size of the feature maps while keeping the most important information. The convolution operation involves sliding these filters across the input image, producing activation maps that capture local features at each spatial position. This step makes the model more efficient and helps it generalize better to different image conditions, including variations in lighting and fruit orientation. Pooling layers, such as max pooling and average pooling, perform down-sampling operations on the feature maps, reducing their spatial dimensions while retaining significant information. This process helps control overfitting and reduces computational cost by decreasing the number of parameters. Activation layers embed non-linear behavior in neural models, allowing the network to learn complex, non-proportional patterns; frequently adopted

activation functions include the rectified linear unit (ReLU), its trainable variant parametric ReLU (PReLU), the exponential linear unit (ELU), and the hyperbolic tangent (TanH) [89,90]. CNNs employ sparse connectivity and weight sharing, meaning each neuron in a convolutional layer process only a local region (receptive field) of the input, and all neurons in the same layer use the same filter weights. This approach sharply reduces the number of parameters compared to fully connected networks. The final part of the architecture is the fully connected layer, where the extracted features are combined and used to predict the final class, such as underripe, ripe, or overripe. CNNs learn these features automatically through training, which involves feeding the model many labeled images and adjusting the network parameters to improve accuracy. This ability allows CNNs to identify complex ripeness feature, including color changes, the appearance of loose fruits, and variations in spikelet texture.

The training process of a CNN involves forward propagation, where input data is passed through the network to generate predictions, followed by computation of a loss function such as cross-entropy, which quantifies the difference between predicted and actual outputs. Backpropagation is then used to adjust the network's weights by minimizing the loss function, typically employing gradient descent optimization [91]. Large labeled datasets are essential for effective training due to the high learning capacity of CNNs, and preprocessing steps include splitting data into training, validation, and testing sets, feature scaling, and handling missing data. Data augmentation is also an important part of preprocessing [92]. It increases the size and

diversity of the dataset by applying simple changes such as rotation, flipping, cropping, and brightness adjustment. The main purpose of augmentation is to help the model handle the natural variability found in agricultural images, which often show uneven lighting, irregular fruit textures, dust, moisture, and shadows [93]. In oil-palm FFBs, fruitlets have different surface patterns and color levels, so texture-based augmentation helps the model generalise better and avoid overfitting. A recent study on unharvested FFB detection applied rotation, horizontal flipping, and scale adjustments to better simulate field conditions and improve robustness [22]. Another study involving oil-palm FFB images also used augmentation to address class imbalance and increase variation in conditions such as partial visibility, low contrast, occlusion, and blurriness [94]. Regularization techniques such as dropout and batch normalization are implemented to prevent overfitting, with dropout randomly disconnecting neuron connections and batch normalization stabilizing training [95]. Hyperparameter tuning involves adjusting learning rate, batch size, number of epochs, and optimizer choice, with Adam optimizer and batch sizes of 16, 32, and 64 commonly used; a batch size of 32 and learning rate of 0.0001 have shown good performance in many studies. Transfer learning leverages pre-trained CNNs to improve training efficiency and generalization, allowing models to converge faster and perform well with limited data. Balanced and high-quality datasets remain critical, as models trained on reliable data yield more robust predictions [96].

CNN performance for oil-palm ripeness classification and fruit-object detection is typically quantified through numerical indicators calculated from a confusion matrix, which summarizes prediction outcomes as true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), representing the correspondence between model outputs and verified reference labels; these core values support the calculation of key metrics including accuracy, precision, recall, and the F1-score, each reflecting a different dimension of model reliability, where accuracy expresses the total fraction of correctly assigned samples, precision measures the share of correct positive predictions relative to all predicted positives, recall (sensitivity) quantifies the rate of positive instances correctly retrieved from all real positives, and the F1-

score consolidates precision and recall through a harmonic averaging scheme to provide a balanced overall score, with the governing equations provided subsequently.

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

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

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

For CNN-based object detection tasks, evaluation extends beyond classification accuracy to include spatial localization quality. Intersection over Union (IoU) measures the overlap between predicted and ground truth bounding boxes. Mean Average Precision (mAP) is the most used detection metric and is computed by averaging the precision values over different recall levels and object classes. In real-time applications, computational efficiency is also evaluated using frames per second (FPS). FPS indicates whether a CNN model is suitable for deployment in real-time plantation or mill environments

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

$$\text{mAP} = \frac{1}{N} \sum_{k=1}^N AP_k$$

$$\text{FPS} = \frac{\text{Number of processed frames}}{\text{Total processing time}}$$

Early CNN research often used simple models such as AlexNet and VGG, which already demonstrated strong potential for agricultural tasks like fruit classification or disease detection. AlexNet is used to classify palm oil fruit ripeness using a dataset of 1,500 images grouped into 3 classes: Raw, ripe, and rotten. The AlexNet model was trained, and its performance was compared with a conventional CNN using validation loss, accuracy, precision, recall, and F1-score as evaluation metrics. Results showed that AlexNet achieved a validation loss of 0.0261 and an accuracy of

0.9962, outperforming the conventional CNN, which recorded a validation loss of 0.0377 and an accuracy of 0.9925, with precision, recall, and F1-scores of 0.99 for AlexNet compared with 0.98 for the conventional CNN [24]. AlexNet is implemented on an Android platform for automated tomato leaf disease classification using RGB leaf images resized to 64×64 pixels. The model was trained on 18,345 images and tested on 4,585 images across ten disease classes, using the Adam optimizer with a learning rate of 0.0005, 75 epochs, a batch size of 128, and cross-entropy loss. The best-performing model achieved an average accuracy of 98%, precision of 0.98, recall of 0.99, an F1-score of 0.98, and a loss value of 0.1331, demonstrating high classification reliability for mobile-based deployment [97]. The VGG-CNN is applied to classify cucumber ripeness using image-based analysis. A dataset of 800 images was used for training and 200 images for testing, with cucumbers categorized into ripe and unripe classes. The VGG model achieved an accuracy of 98.5%, demonstrating that deep CNN architectures can effectively support objective ripeness assessment and assist farmers in decision making [98]. Over time, deeper and more capable architectures were developed, including ResNet, Inception, DenseNet, MobileNet, and EfficientNet. These models improved accuracy, reduced model size, and became more reliable in real-world conditions. Agricultural imaging has benefited from this progress. For example, lightweight architectures like MobileNet and EfficientNet are useful for mobile or field-based systems, while deeper networks like ResNet or DenseNet provide strong performance in controlled environments where computational resources are available. The increasing diversity of CNN architectures has made it possible to design systems that are suited for both laboratory experiments and field deployment in plantations or mills. Most CNN studies on oil palm focus on classification, where the model predicts the ripeness stage of the FFB based on a single RGB image. Studies using AlexNet, VGG, and ResNet demonstrated that CNNs can outperform traditional ML methods that rely on manual feature extraction. These classification models typically use image inputs showing the whole bunch or close-up views of fruitlets. The main indicators learned by the CNN include exocarp color, fruitlet arrangement, and surface texture.

### ***CNN-based object detection for FFB***

#### ***localization***

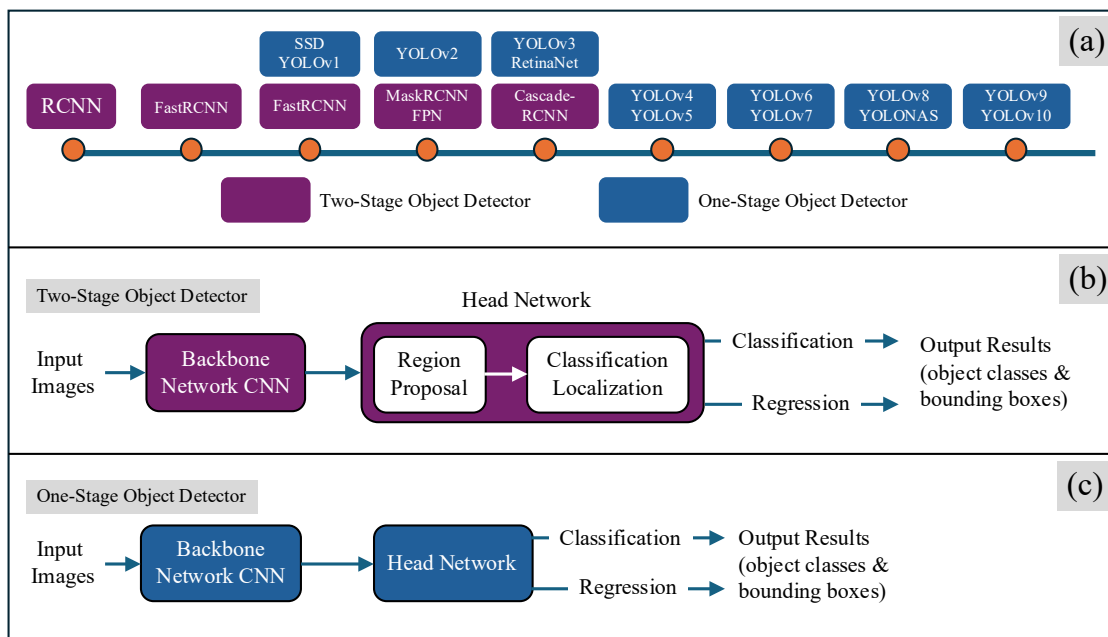
Previous section discussed CNN studies in agricultural imaging used backbone algorithms such as AlexNet and VGG to learn features from entire RGB images for ripeness classification. In early oil palm FFB grading research, these algorithms analyzed the whole image frame without first detecting or isolating the FFB object, which limited reliability in plantation scenes where leaves, shadows, and occlusion affected the learned feature maps. Later work addressed this by extending CNN backbones into object detection algorithms, which localise FFBs through bounding box regression so that ripeness grading can focus only on the detected bunch region, improving robustness under variable field conditions. Localization refers to identifying the position of a target object within an image, commonly represented using bounding boxes or region coordinates. Alfatni *et al.* [99] developed a rule-based expert system using region-of-interest (ROI) image processing to classify FFB ripeness, motivated by the need for a fast, simple, and accurate grading method to reduce disputes between farmers and mills. Three different ROIs were evaluated, namely ROI1 with a size of 300×300 pixels, ROI2 with a size of 50×50 pixels, and ROI3 with a size of 100 × 100 pixels, from which statistical colour features were extracted. These features were first interpreted through expert-defined rules and subsequently classified using ML. The results showed that the rule-based system combined with a k-nearest neighbors classifier achieved an accuracy of 94% for the selected ROI, represents a conventional vision-based approach where handcrafted colour features and predefined rules formed the core of ripeness interpretation. A non-destructive imaging technique was investigated to quantify oil palm FFB ripeness by analyzing surface colour and its relationship with oil content. RGB images were captured using a digital camera, and hue histograms were generated from mathematically transformed RGB data, with dominant hue peaks correlated to measured oil content. The correlation between hue peak and oil content across mixed ripening patterns was  $r = 0.7933$ , while analysis limited to FFBs exhibiting similar colour change patterns produced a much stronger correlation of  $r = 0.9519$ , demonstrating that RGB-based imaging can effectively estimate oil content under homogeneous

ripening conditions [100]. The introduction of CNN-based object detectors marked a major change, as these models learn features directly from diverse datasets and can adapt to complex visual conditions. CNN-based detectors can learn discriminative spatial and contextual features that separate bunches from background vegetation [101].

Following early conventional image processing approaches based on handcrafted colour features, rule-based decision systems, and shallow learning models, research on oil palm FFB analysis gradually shifted toward more data-driven algorithms. While traditional methods demonstrated reasonable performance under controlled conditions, their reliance on fixed thresholds and predefined features limited robustness when faced with illumination variation, occlusion, and complex plantation backgrounds. This limitation motivated the adoption of ML classifiers and, subsequently, DL architectures capable of learning discriminative features directly from image data. Initial DL efforts focused on image-level classification, but these approaches lacked spatial awareness and could not identify the location of bunches within cluttered scenes. To address this, 2-stage object detection frameworks were introduced, enabling the separation of region proposal generation from classification and providing improved localization accuracy in complex environments. Although effective, 2-stage detectors imposed high computational costs and slower inference speeds, which constrained their use in real-time field applications. As a result, research progressed toward single-stage detection architectures that perform localization and classification simultaneously, offering faster inference while maintaining acceptable accuracy.

**Figure 7** illustrates the evolution and structural differences of CNN based object detection algorithms

relevant to oil palm FFB localization. **Figure 7(a)**, detection methods have progressed from early region-based approaches such as R-CNN, Fast R-CNN, and Faster R-CNN, which operate as 2-stage detectors using a separate region proposal process followed by classification, to more advanced 1-stage detectors including SSD and the YOLO family that unify localization and classification within a single network. While 2-stage detectors provide strong localization accuracy, they require higher computational resources and longer inference time. Continuous development of 1-stage detectors, particularly across successive YOLO versions, has improved detection accuracy, robustness, and efficiency, making them suitable for real-time deployment. **Figure 7(b)** presents the general architecture of a 2-stage detector, where input images are processed by a backbone CNN, followed by region proposal generation and refinement through classification and bounding box regression, whereas **Figure 7(c)** illustrates a 1-stage detector in which feature extraction and prediction are integrated into a single pipeline, enabling faster inference and practical deployment on edge devices, mobile platforms, and plantation systems. In addition to these differences, the selection of the backbone network strongly affects detection quality because deeper architectures improve semantic representation but may lose fine spatial detail needed for identifying small or partially occluded fruitlets, whereas wider networks can preserve local structures that are important for detecting subtle ripeness differences and densely packed fruitlet clusters [102]. Performance is also influenced by training components such as anchor box design, IoU based loss functions and assignment strategies, which determine how accurately models can localise small targets on irregular bunch surfaces [103].



**Figure 7** (a) Evolution and architectural comparison of CNN-based object detection models; (b) structure of a 2-stage detector; (c) structure of a 1-stage detector [104].

Recent work has placed strong emphasis on 1 stage detectors, but this focus can mask important limitations when applied to the structure of oil palm fruit FFBS. The FFBS surface contains tightly packed fruitlets of varying sizes and colours, irregular spacing, and frequent occlusion from spikelets and fronds [8]. Many 1 stage models achieve their speed by reducing feature resolution or relying on coarse prediction grids, which makes it difficult to capture the subtle visual transitions that signal intermediate ripeness levels. High performance reported in some studies often reflects the simplicity or uniformity of the dataset rather than the suitability of the architecture for complex field conditions. Two stage detectors deserve stronger consideration because their stepwise refinement allows more precise separation of overlapping fruitlets and more careful interpretation of fine grained visual features [105]. Their slower inference is a disadvantage for rapid field deployment, but many agricultural workflows prioritise precision and reliability over speed, such as breeding programs, controlled seed production, or mill side quality inspections. Oil palm grading also aligns with broader developments in smart agriculture, where neural network based monitoring is increasingly used for soil contamination assessment, pollutant detection, and crop stress analysis [106,107]. Most published work in oil-palm vision research relies

on existing architectures such as YOLO, SSD or lightweight CNNs, and the majority of these studies focus on adjusting training settings rather than introducing structural innovations [28,41]. Common modifications include changing the number of epochs, tuning hyperparameters, altering input resolution or comparing several standard detectors on the same dataset [108].

Region-based CNN architectures, including R-CNN, Fast R-CNN, and Faster R-CNN, have been implemented in oil palm applications primarily for tree detection and counting tasks, where precise localization of individual palm crowns is required in aerial or ground-based imagery [23,109,110]. These 2-stage detectors are well suited for such tasks because the region proposal mechanism enables accurate separation of palm objects from complex backgrounds. However, their application has been largely limited to structural detection problems rather than ripeness assessment. Ripeness evaluation requires fine-grained interpretation of fruitlet color variation, surface texture, and local visual feature at the bunch level, which are not optimally addressed by region-based frameworks designed for generic object localization. In addition, the high computational cost and slower inference speed associated with R-CNN-based architectures reduce their practicality for close-range, real-time ripeness analysis

in plantation environments, where faster and more lightweight detection models are generally favored. In contrast, 1-stage object detection architectures have been more actively explored for oil palm FFB localization and ripeness-related tasks because they offer a better balance between detection accuracy and computational efficiency. Frameworks such as Single Shot MultiBox Detector (SSD) and the You Only Look Once (YOLO) family perform object localization and classification in a single forward pass, allowing faster inference that is suitable for field deployment [111].

Selvam *et al.* [47] employed the YOLOv3 object detection algorithm for real-time detection of ripe palm oil bunches, using labeled images and videos of palm oil bunches to train a detection model within the Darknet framework based on a pre-trained network. Model training reached learning saturation at the 6,000<sup>th</sup> iteration, indicating stable detection performance for differentiating maturity levels during real-time operation. A noted limitation of the approach is the tendency toward overfitting at higher training iterations, suggesting that model generalization may be affected when applied to more diverse plantation conditions or unseen environments. Different YOLOv8 variants, ranging from YOLOv8n to YOLOv8x, were systematically evaluated, and the YOLOv8m model was identified as providing the best balance between detection accuracy and processing speed. The YOLOv8m configuration achieved a processing speed of approximately 14.68 iterations per second, with top-1 and top-5 accuracies of 0.9885 and 0.998, respectively, demonstrating a clear progression from earlier YOLO-based implementations toward more precise and scalable ripeness detection frameworks suitable for practical deployment [11]. Chowdhury *et al.* [108] conducted a comparative evaluation of multiple

YOLO variants for real-time ripeness detection of palm oil FFBs using images collected directly from plantation environments. A custom dataset of 2,000 images acquired from a palm oil plantation in Selangor, Malaysia, was used to train and test YOLOv5, YOLOv6, and YOLOv7 models, including small and medium configurations such as YOLOv5s, YOLOv5m, YOLOv6s, YOLOv6m, and YOLOv7. Experimental results showed that the YOLOv5s model achieved the best performance for real-time deployment, reaching an F1-score of 98.5% while maintaining a processing speed of 35.76 frames per second, demonstrating that lightweight YOLO architectures can provide an effective balance between detection accuracy and inference speed for in-field palm oil FFB ripeness detection. Complementing algorithm-focused studies on CNN-based ripeness detection, recent work has highlighted the importance of dataset quality and representativeness for reliable model development. A comprehensive annotated dataset was introduced based on real operating conditions in palm oil mills, addressing limitations of earlier studies that relied mainly on isolated FFB images with incomplete maturity categorization. The dataset includes videos and images collected directly from mill environments, comprising 45 single-category videos and 56 multi-category videos, recorded using a smartphone at a resolution of 1,280×720 pixels in MP4 format. All data were annotated into 6 maturity categories, namely unripe, under-ripe, ripe, overripe, empty bunches, and abnormal fruit, providing a realistic benchmark for training and evaluating DL models intended for practical mill-side ripeness grading applications [45]. **Table 4** provide the summary of previous published study related to the FFB Localization in palm oil Grading.

**Table 4** Summary of FFB localization.

No	Method	Finding	Ref	Region
1	Applied the YOLOv3 object detection algorithm, trained using labeled images and videos of palm oil bunches within the Darknet framework, and implemented a Python-based interface connected via Tkinter for real-time operation.	The YOLOv3 model was able to detect and differentiate palm oil bunch maturity levels in real time, reaching learning saturation at 6000 training iterations and demonstrating potential for integration with mobile and Internet of Things systems.	[47]	Malaysia
2	Employed a YOLOv10-S object detection model trained for ripeness detection and deployed it within	The YOLOv10-S model achieved a mAP of 0.95, with precision of 0.96 and recall of 0.87,	[112]	Malaysia

No	Method	Finding	Ref	Region
	a web-based interface developed using Python Streamlit.	demonstrating effective performance for web-based oil palm FFB ripeness detection.		
3	Evaluated multiple YOLOv8 architectures using an oil palm fruit ripeness dataset, with a focus on balancing detection accuracy and computational efficiency.	The YOLOv8m model achieved an mAP50-95 of 0.927, demonstrating strong performance and suitability for practical ripeness assessment in the palm oil industry.	[113]	Malaysia
4	The study used a dataset of 180 images with a resolution of 640 × 640 pixels collected from the PTPN IV oil palm plantation in North Sumatra, Indonesia, which were pre-processed and augmented using Roboflow before training and testing YOLOv8 and YOLOv9 models in Google Colab.	Both models successfully detected ripe, rotten, and unripe oil palm fruit, with YOLOv8 achieving an accuracy of 0.984 (98.4%) and YOLOv9 achieving a higher accuracy of 0.99 (99%), indicating improved performance with the newer YOLO version.	[114]	Indonesia
5	Implemented and compared several YOLO-based models, with YOLOv4 proposed as the main model for both 1-class and 4-class ripeness detection.	YOLOv4 achieved the best performance, with an mAP of 98.97%, F1-score of 0.96, and average IoU of 68.80% for 1-class detection, and an mAP of 77.20%, F1-score of 0.7, and average IoU of 51.68% for 4-class detection.	[115]	Malaysia
6	Comparing MobileNetV2 SSD, EfficientDet (Lite0, Lite1, Lite2), and YOLOv5 variants (YOLOv5n, YOLOv5s, YOLOv5m) trained on a dataset with 4 ripeness classes: Ripe, unripe, half-ripe, and over-ripe, using metrics such as mAP, precision, recall, and training time.	Among the evaluated models, YOLOv5m demonstrated the most promising performance, achieving a mAP of 0.842 for oil palm FFBs classification.	[116]	Malaysia
7	Employing an LED-based multispectral imaging system and a 2-class YOLOv4 detection model trained on 2,000 annotated multispectral images of unripe and ripe FFBs, with real-time testing conducted on videos of 30 moving FFBs on a conveyor.	The proposed system achieved an average detection accuracy of 99.66% with a processing speed of 3.32 - 3.62 frames per second, demonstrating strong potential for real-time ripeness assessment in palm oil mills.	[75]	Indonesia
8	The study applied an optimized YOLOv7 model with strategic fine-tuning using a comprehensive dataset and integrated the trained model into a web-based application for real-time assessment.	The optimized YOLOv7 model achieved a classification accuracy of 92.55% and a mAP of 95.08%, demonstrating effective performance for practical deployment in palm oil production systems.	[117]	Thailand
9	The study employed YOLOv6 models, specifically YOLOv6s and YOLOv6m, trained on real plantation images categorized into 4 ripeness classes: Unripe, underripe, ripe, and overripe, and evaluated using precision, recall, F1-score, mAP, training time, and inference speed.	The YOLOv6m model outperformed YOLOv6s, achieving precision of 36.9%, recall of 30%, F1-score of 33.1%, mAP(50) of 36.9%, and mAP(50 - 95) of 16.5% after 100 training epochs, demonstrating its capability for automated multi-class ripeness detection under field conditions.	[118]	Malaysia
10	The study compared image classification using ResNet50 and object detection using YOLOv3, highlighting the effect of localized feature extraction on ripeness classification performance.	YOLOv3 improved classification accuracy by 2% for overripe, 27% for ripe, and 12% for underripe classes, demonstrating the advantage of object-level feature localization over whole-image classification.	[119]	Malaysia

### **Lightweight and edge-deployable CNN models**

The increasing demand for real-time and field-deployable ripeness grading systems has driven significant interest in lightweight and edge-deployable CNN models. In oil palm plantations, ripeness assessment is often conducted under resource-constrained conditions, where high computational power, stable internet connectivity, and cloud-based processing are not always available. As a result,

compact CNN architectures that can operate efficiently on mobile devices, embedded systems, and edge computing platforms have become a critical research direction. Lightweight CNN models are specifically designed to reduce computational complexity, memory usage, and energy consumption while maintaining acceptable classification accuracy [120]. Architectures such as MobileNet, ShuffleNet, EfficientNet, and SqueezeNet have been widely adopted in agricultural

vision tasks due to their use of depthwise separable convolutions, pointwise convolutions, and parameter-efficient scaling strategies [121-123]. These design principles significantly decrease the number of trainable parameters and floating-point operations compared to conventional deep CNNs such as VGG or ResNet, making them more suitable for deployment on smartphones, single-board computers, and low-power edge devices.

In the context of oil palm FFB ripeness grading, several studies have demonstrated that lightweight CNNs can achieve competitive performance when trained on well-curated RGB image datasets. MobileNet-based models, in particular, have shown strong potential due to their balance between inference speed and classification accuracy. When combined with transfer learning from large-scale image datasets, these models can effectively extract discriminative color and texture features associated with different ripeness stages, even under varying illumination and background conditions. Suharjito *et al.* [41] applied ImageNet transfer learning with 3 unfrozen convolutional blocks and a 9-angle crop data augmentation method to 4 lightweight CNN models for mobile-based oil palm FFB ripeness classification. EfficientNetB0 achieved the highest accuracy of 0.898 on Keras and 0.893 after float16 quantization on TensorFlow Lite, with an inference time of 96 ms per image, while MobileNetV1, although less accurate (0.811), provided the fastest inference. The study demonstrated that combining lightweight CNN architectures, 9-angle crop augmentation, and float16 quantization enables accurate and real-time ripeness grading on Android devices using live camera input. A MobileNet-based ripeness classification system was deployed on an Android edge device and evaluated under plantation conditions, achieving an accuracy of 85% [124]. A ResNet50 model achieved the highest adjusted accuracy of 90% with F1-scores above 80% for all ripeness classes, but required a large model size of 405 MB and longer inference times of 2.48 seconds on GPU and 3.27 seconds on CPU. A DenseNet121 model trained from scratch provided a more efficient alternative, achieving 86% accuracy with comparable F1-scores, reduced model size of 100 MB, faster inference times of 1.76 seconds on GPU and 2.56 seconds on CPU, and strong robustness to brightness variations ranging from -70 to +70 [125].

Edge deployability further extends the practical value of lightweight CNNs by enabling on-device inference without reliance on cloud servers. This approach reduces latency, improves data privacy, and enhances system robustness in remote plantation environments where network connectivity is often limited or unstable. In practical implementations, edge computing is commonly realized using low-power embedded platforms such as Raspberry Pi, NVIDIA Jetson Nano, and similar single-board or edge AI devices [126,127]. These platforms provide sufficient computational capability to execute optimized CNN models while maintaining low energy consumption and compact form factors suitable for field deployment. By performing data processing and inference directly on edge devices, image data acquired from cameras can be analyzed locally, allowing real-time ripeness grading without continuous data transmission to centralized servers. This edge-based processing paradigm is particularly advantageous for oil palm plantations, where large-scale image acquisition and variable environmental conditions may impose significant communication and bandwidth constraints. To support efficient execution on such platforms, trained CNN models are typically converted into optimized formats using deployment frameworks such as TensorFlow Lite, ONNX Runtime, and NVIDIA TensorRT [41,76,113]. These frameworks enable model compression and hardware-aware optimization, thereby improving inference speed and reducing memory requirements on edge devices.

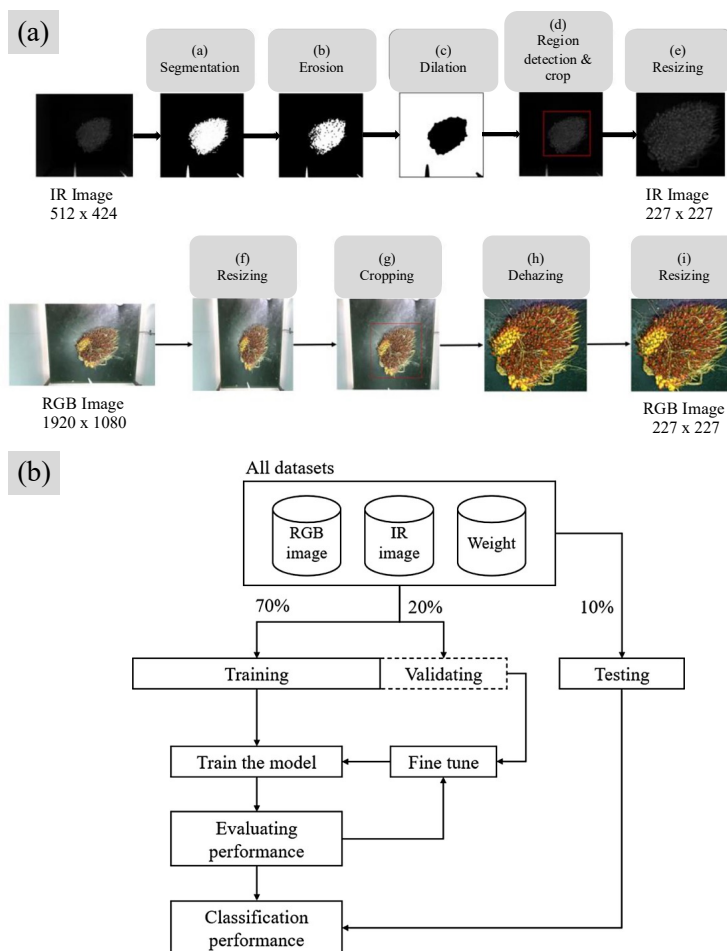
#### ***Multimodal and Hybrid CNN approaches for oil palm assessment***

Recent studies have shown that single-modality image data, particularly RGB images, may be insufficient to fully capture the complex physical changes associated with oil palm FFB ripeness. As a result, multimodal and hybrid CNN approaches have gained increasing attention in oil palm assessment research. These approaches aim to integrate complementary data sources or combine multiple learning architectures to improve classification robustness and generalization performance. Multimodal CNN frameworks typically fuse information from different sensor modalities, such as RGB images, hyperspectral or multispectral data, thermal images, and

depth information. Multimodal fusion can help reduce the weaknesses of individual sensors when assessing oil-palm FFBs [128]. RGB images provide useful color and texture information but may struggle with surface irregularities, dense fruitlet clusters, and natural variation in bunch appearance. Depth sensors give information about bunch structure, such as shape and fruitlet spacing, but they cannot capture color features that are important for judging ripeness. NIR or thermal sensors offer signals related to moisture and early internal changes, although they do not provide detailed visual features. By combining these different types of data, multimodal systems can balance the strengths and weaknesses of each sensor [129]. Depth information can support the model in understanding bunch geometry, while RGB images add important visual signs of maturity. NIR data can also help detect internal changes that are not visible on the surface. In plantations, this fused approach improves consistency when bunches

differ in orientation, fruitlet density, or surface conditions.

Pipitsunthonsan *et al.* [130] proposed a nondestructive oil palm FFB grading method using a multi-input and multi-label CNN that integrates data from an RGB camera, infrared sensor, and load cell to capture colour, shape, and size information as shown in **Figure 8**. The model was trained using 1,575 images from 14 varieties collected at 4 trading sites, with dataset splits of 70% for training, 20% for validation, and 10% for testing. The cross-entropy loss function was applied during training, and Gradient-weighted Class Activation Mapping (Grad-CAM) was used to analyse the influence of image preprocessing on model learning. The proposed approach achieved an accuracy of 90.26%, precision of 89.86%, recall of 89.54%, and an F1-score of 89.68% on the testing dataset. Another study also explored multimodal data for oil palm FFB ripeness assessment by using a dataset that combines videos and still images acquired in palm oil mill environments [45].



**Figure 8** Multimodal CNN for palm oil grading [130]. (a) Image preprocessing; (b) Model development.

Hybrid CNN approaches, on the other hand, focus on architectural integration rather than sensor fusion. A common strategy is the combination of CNNs with traditional ML classifiers, such as SVM or random forests (RF), where CNNs are used for deep feature extraction and conventional classifiers perform the final decision stage [131]. Another important hybrid direction involves the integration of CNNs with temporal or sequential models, such as long short-term memory networks [132]. Although still limited in oil palm research, such approaches are promising for analyzing time-dependent ripening patterns, especially when image data are collected repeatedly from the same plantation blocks or harvesting cycles.

### Current gaps and future directions

Existing research on DL-based ripeness grading of oil palm FFBs continues to face fundamental challenges related to data availability, representativeness, and model generalization. At present, each study uses a different dataset collected under its own conditions, with inconsistent ripeness definitions, annotation formats, image resolutions, and evaluation splits. As a result, commonly reported metrics such as accuracy, F1-score, IoU, or mean average precision cannot be compared directly across publications. A reliable CNN benchmark for oil palm fruit bunch grading requires a fixed maturity scale informed by objective markers such as those referenced by the MPOB [133]. In addition, the dataset should include detailed information about the plantation context, including variety, plant age, and the geographic coordinates of the site to capture topographical conditions. Key imaging parameters, including illumination and sensor geometry, should be documented because variation in these factors is known to affect CNN stability [134]. Annotation of bounding boxes or masks must follow a consistent precision rule to minimize label noise, which has been shown to distort detection results in agricultural datasets. Strengthening dataset design and applying consistent ripeness criteria are important steps for improving the generalization of CNN models beyond controlled settings [135,136]. One possible direction is the development of a shared research platform, proposed here as “PalmNet”, which would function similarly to a curated repository where researchers can contribute datasets, annotation

guidelines, model outputs, and training records under controlled licensing. Such a platform would expand the range of imaging conditions available for model development while maintaining data ownership and quality standards. Providing a common space for structured data and model exchange would support more reliable benchmarking and help reduce dataset driven bias in future studies.

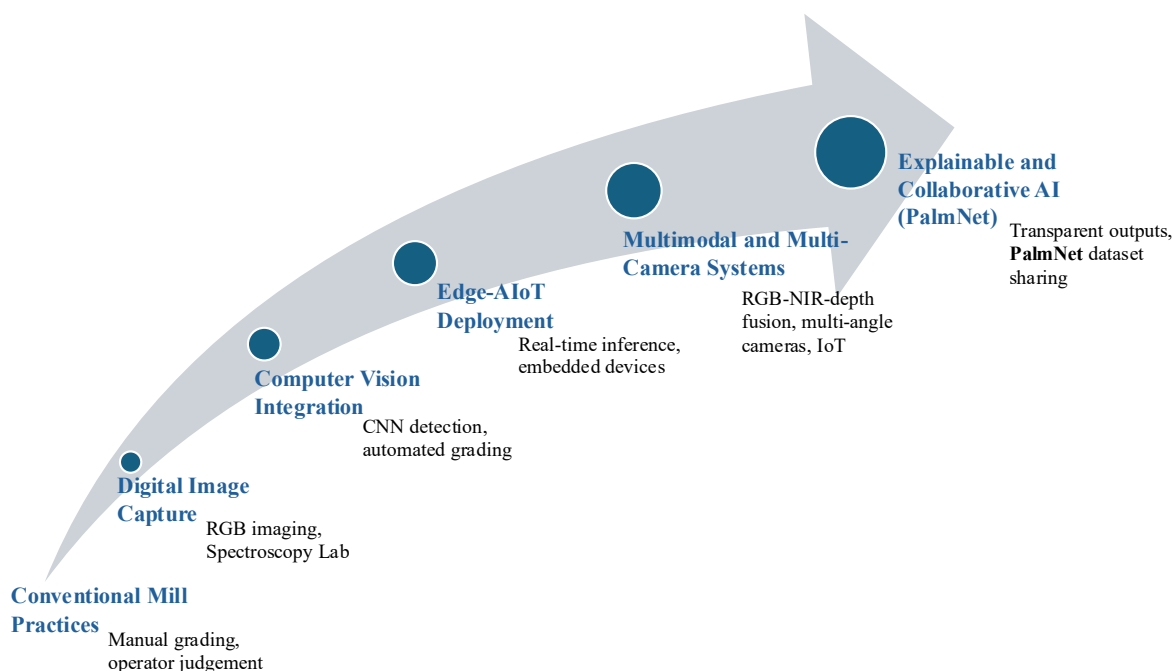
Another prominent research gap lies in computational demands and practical deployability. While advanced CNN architectures can achieve high classification accuracy, they typically require high-performance computing resources that are not readily available in field or mill-side settings. Many studies do not report inference time, memory usage, or energy consumption, even though these factors are critical for real-time applications on smartphones, embedded systems, or edge devices. Although lightweight networks and optimization techniques such as quantization or pruning have been explored, their adoption remains inconsistent, and performance trade-offs are not systematically evaluated. Moreover, DL-based ripeness grading systems are rarely integrated into comprehensive Internet of Things (IoT) infrastructures that support plantation or mill operations. The lack of end-to-end integration, including real-time data transmission, geotagging, storage, and linkage to operational decision-making platforms, limits the usefulness of these systems for plantation offices and quality management workflows. From a quality assessment standpoint, current approaches tend to oversimplify ripeness as a purely visual classification task. Most CNN-based studies focus on external mesocarp characteristics, such as colour and texture, derived primarily from RGB imagery [52]. While these features are informative, they provide only indirect insight into critical quality parameters. Real-time estimation of FFA content, which is strongly correlated with fruit maturity, harvesting delay, and processing conditions, remains largely absent from existing DL frameworks [137,138].

Predicting FFA from RGB images remains technically uncertain because CNNs can only extract features that are visible on the fruit surface, while many chemical processes involved in FFA formation do not produce consistent visual patterns. Although CNNs can

learn latent texture and structural information beyond colour, studies indicate that these RGB-derived features often capture correlations specific to a single dataset rather than biochemical signals that generalize across lighting conditions, cultivars, or harvesting environments [139]. This suggests that RGB-based CNNs are unlikely to provide stable prediction of FFA unless internal changes produce strong and consistent surface effects. Using a near infrared camera offers a more direct pathway, as NIR wavelengths are sensitive to water absorption, and moisture levels are closely linked to the hydrolysis reactions that increase FFA during maturation and early degradation [140,141]. To determine whether CNNs can reliably infer FFA from RGB alone, or whether NIR or RGB-NIR fusion is required, large datasets with paired RGB images, NIR measurements, and biochemical ground truth must be developed and evaluated under a common benchmarking protocol. Ripeness continues to change after harvest, and this temporal factor can lead to differences between the grading recorded at the field collection point and the biochemical quality measured at the mill [142]. Because FFA can rise rapidly during transport and storage, future datasets should document the exact time interval between image capture and biochemical sampling to ensure that model outputs reflect true post-harvest dynamics. Regression-based approaches are more suitable for this purpose, as they can estimate continuous variables such as oil content and FFA levels. A coordinated platform such as the proposed “PalmNet” could support this requirement by enabling researchers to share datasets that include detailed metadata such as plant variety, plant age, imaging conditions, and time-after-harvest together with matched biochemical ground truth. Similarly, kernel-related attributes are rarely addressed, as they are not directly observable from surface imagery. These limitations indicate a need for more comprehensive modelling strategies that combine ripeness grading with quantitative quality estimation.

Future research should advance toward multimodal and multi-task learning frameworks that integrate visual data with additional sensing modalities to better reflect real plantation and mill conditions. Multi-camera setups, including RGB views from different angles, can reduce occlusion and provide more consistent features across harvesting, transport, and

grading. Estimating oil extraction rate (OER) at the individual FFB level will require combining weight measurements, visual indicators, and spectral inputs, with NIR imaging offering promise due to its sensitivity to internal composition and potential links to FFA levels and oil accumulation. Depth cameras can further enhance analysis by generating 3D point clouds that capture volume, fruitlet distribution, and surface structure, which can be fused with RGB and NIR data to improve predictions of mass and maturity-related characteristics [143]. To support these developments, future systems should explore multi-task DL architectures capable of predicting ripeness class, FFA-related indicators, bunch weight, and OER in real time, supported by operational validation and integration with mill information systems. The roadmap in Figure 9 outlines a gradual shift from manual, operator-dependent grading toward more standardised and data-driven mill operations. Introducing RGB image capture provides consistent visual records that form the basis for reproducible datasets and early digital traceability. The adoption of CNN-based computer vision reduces grading subjectivity, though its success depends on careful dataset design and validation under real mill conditions. Deploying models on edge-AI devices enables real-time processing within mill workflows, addressing latency constraints and supporting continuous grading. Multimodal sensing including RGB, NIR, and depth imaging, offers stronger prediction of ripeness, mass, and early biochemical indicators, but uptake may be slowed by human resistance to automation due to concerns about reliability and job displacement. The final stage incorporates explainable and collaborative AI through a PalmNet platform, providing transparent model behaviour and shared benchmarking that can help build trust among plantation and mill personnel. To support this progression, Table 5 summarizes influential studies and highlights the differences in datasets and architectures, helping identify which approaches are most suitable for future industrial deployment.



**Figure 9** Roadmap for industrial deployment of AIoT-based ripeness and quality monitoring in palm-oil mills.

**Table 5** Summary of influential studies on oil-palm FFB detection and ripeness grading.

Dataset Characteristics	Imaging Modality	CNN Architecture	Performance	Limitation	Ref.
47 training videos + 10 test videos	Smartphone-captured outdoor videos	YOLOv4-Tiny-3L	mAP50 90.56%, IoU 58.35%, 105 FPS	Small dataset size	[28]
653 images from an oil-palm mill	Smartphone RGB images (outdoor, natural lighting)	EfficientNetB0	0.898 test accuracy	Small dataset, only single-fruit images (no detection task)	[41]
7,171 images extracted from rotating smartphone videos	Smartphone video (MP4) captured with 360° rotation; natural outdoor lighting	YOLOv8	mAP50 = 0.9949; precision = 0.996, recall = 0.996	only single-fruit images (no detection task)	[113]
1,575 images for 14 varieties oil palm	Multimodal (RGB-IR-Weight)	Modified CNN	accuracy of 90.26%, precision of 89.86%, recall of 89.54%, and an F1-score of 89.68%.	Prototype stage, under controlled lighting	[130]
4,160 RGB images	Smartphone RGB images collected outdoors at mills	YOLOv4-tiny	mAP = 99.70%, Precision 0.96, Recall 1.00, IoU 86.05%.	Dataset limited to midday lighting	[144]

## Conclusions

Palm oil FFB grading is a critical operation within the palm oil production chain, as ripeness strongly influences FFA content, OER efficiency, and overall product quality. Inaccurate grading at the harvesting or mill intake stage can result in elevated FFA levels, suboptimal oil yield, and economic losses, highlighting the need for reliable, objective, and scalable assessment methods. This review shows that research on CNN-based ripeness grading has grown substantially since 2017, following the broader adoption of DL in agricultural engineering and CV. Analysis of Scopus-derived results indicates that Malaysia is the most active contributor in this research area, followed by Indonesia, reflecting strong alignment between scientific activity and the global distribution of palm oil production. The literature also shows a clear transition from conventional image processing toward CNN-based classification and object detection, with 1-stage detectors gaining popularity due to their real-time capabilities. Recent studies further explore practical deployment, including mobile systems, embedded devices, and edge computing solutions that enable on-site grading without relying on high-performance servers.

Looking ahead, the most promising research paths are expected to focus on multimodal sensing integrated with predictive quality monitoring, allowing ripeness assessment to capture both external appearance and internal quality indicators related to OER and FFA formation. Combining RGB imaging with complementary inputs such as weight, NIR response, thermal data, or depth information can improve robustness under field and mill conditions. Integrating these multimodal systems into AIoT-based platforms offers a practical direction for palm-oil management, especially in mill environments where real-time grading, automatic sorting, and continuous material tracking can be coordinated through connected devices. Progress in lightweight yet accurate architectures will also be essential for scalable deployment across plantations of different sizes, including smallholder operations with limited computational resources. A collaborative effort to develop a shared “PalmNet” dataset, consisting of diverse palm oil varieties collected

from multiple regions, would further support model generalization and strengthen future research.

To support more consistent development, the field would benefit from standardized performance metrics that account for biological variability and operational conditions. These metrics may include maturity-class accuracy under different lighting situations, model robustness across cultivars and regions, and throughput measures relevant to real-time processing. In addition, standardizing imaging conditions, for example by using light sensors to automatically adjust the camera and stabilize brightness would improve consistency and reliability during data capture in mills. Establishing shared evaluation criteria and standardized capture protocols will enable fair comparison across studies and guide future progress toward reliable and scalable solutions for ripeness assessment and palm-oil management.

## Acknowledgements

The authors would like to acknowledge the support from: Presidential Scholarship, Chiang Mai University, Thailand and Renewable Energy and Energy Conservation (REEC) Laboratory, Department of Mechanical Engineering, Chiang Mai University, Thailand.

## Declaration of generative AI in scientific writing

The authors acknowledge that generative AI tools (including QuillBot and OpenAI’s ChatGPT) were used exclusively for language refinement and grammatical editing in preparing this manuscript. No AI assistance was involved in generating content or interpreting data. The authors take full responsibility for all interpretations, findings, and conclusions presented in this work.

## CRedit author statement

**Wahyu Nurkholis Hadi Syahputra:** Writing - Original Draft; Conceptualization; Investigation; Visualization. **Chatchawan Chaichana:** Supervision; Conceptualization; Project administration. **Damorn Bundhurat:** Writing - Reviewing and Editing; Supervision; **Patiwet Wuttisarnwattana:** Writing - Reviewing and Editing; Supervision. **Bayu Taruna**

**Widjaja Putra:** Writing - Reviewing and Editing; Supervision; Visualization.

## References

- [1] JP Rajakal, JZH Hwang, MH Hassim, V Andiappan, QT Tan and DKS Ng. Integration and optimisation of palm oil sector with multiple-industries to achieve circular economy. *Sustainable Production and Consumption* 2023; **40**, 318-336.
- [2] R Nabila, W Hidayat, A Haryanto, U Hasanudin, DA Iryani, S Lee, S Kim, S Kim, D Chun, H Choi, H Im, J Lim, K Kim, D Jun, J Moon and J Yoo. Oil palm biomass in indonesia: thermochemical upgrading and its utilization. *Renewable and Sustainable Energy Reviews* 2023; **176**, 113193.
- [3] H Varkkey, A Tyson and SAB Choiruzzad. Palm oil intensification and expansion in indonesia and malaysia: environmental and socio-political factors influencing policy. *Forest Policy and Economics* 2018; **92**, 148-159.
- [4] A Hidayatno, AD Setiawan, A Subroto, H Saheruddin, S Wardono, H Romijn, TN Zahari, I Rahman, BA Jafino, AO Moeis, K Komarudin, AR Fitriani, N Julio and Z Zafira. Exploring the food-versus-fuel debate in indonesia's palm oil industry toward sustainability: A model-based policymaking approach. *Energy Nexus* 2025; **19**, 100511.
- [5] JW Lai, HR Ramli, LI Ismail and WZ Wan Hasan. Oil palm fresh fruit bunch ripeness detection methods: A systematic review. *Agriculture* 2023; **13(1)**, 156.
- [6] M Saufi, M Kassim, W Ishak, W Ismail and A Rahman. Image clustering technique in oil palm Fresh Fruit Bunch (FFB) growth modeling. *Italian Oral Surgery* 2014; **2**, 337-344.
- [7] I Novianty, RG Baskoro, MI Nurulhaq and MA Nanda. Empirical mode decomposition of near-infrared spectroscopy signals for predicting oil content in palm fruits. *Information Processing in Agriculture* 2023; **10(3)**, 289-300.
- [8] JY Goh, Y Md Yunos and MS Mohamed Ali. Fresh fruit bunch ripeness classification methods: A review. *Food and Bioprocess Technology* 2025; **18(1)**, 183-206.
- [9] M Rosbi, Z Omar, U Khairuddin, APPA Majeed and SARSA Bakar. Machine learning for automated oil palm fruit grading: The role of fuzzy C-means segmentation and textural features. *Smart Agricultural Technology* 2024; **9**, 100691.
- [10] Z Omar, APP Abdul Majeed, M Rosbi, SA Ghazalli and H Selamat. Outdoor oil palm fruit ripeness dataset. *Data in Brief* 2024; **55**, 110667.
- [11] J Josdaan, VC Tamsil, J Harefa and K Jingga. Revolutionizing palm oil ripeness classification: Utilizing YOLOv8 for ultra-precise ripeness detection. *Procedia Computer Science* 2024; **245**, 700-709.
- [12] M Nautiyal, S Joshi, I Hussain, H Rawat, A Joshi, A Saini, R Kapoor, H Verma, A Nautiyal, A Chikara, W Ahmad and S Kumar. Revolutionizing agriculture: A comprehensive review on artificial intelligence applications in enhancing properties of agricultural produce. *Food Chemistry: X* 2025; **29**, 102748.
- [13] N Aijaz, H Lan, T Raza, M Yaqub, R Iqbal and MS Pathan. Artificial intelligence in agriculture: Advancing crop productivity and sustainability. *Journal of Agriculture and Food Research* 2025; **20**, 101762.
- [14] MZ Abdullah, LC Guan and BMN Mohd Azemi. Stepwise discriminant analysis for colour grading of oil palm using machine vision system. *Food and Bioprocess Technology* 2001; **79(4)**, 223-231.
- [15] I Attri, LK Awasthi, TP Sharma and P Rathee. A review of deep learning techniques used in agriculture. *Ecological Informatics* 2023; **77**, 102217.
- [16] S Mohapatra, T Swarnkar and J Das. *Deep convolutional neural network in medical image processing*. In: VE Balas, BK Mishra and R Kumar (Eds.). Handbook of deep learning in biomedical engineering. Academic Press, London, 2021, p. 25-60.
- [17] Y Lin, S Xia, L Wang, B Qiao, H Han, L Wang, X He and Y Liu. Multi-task deep convolutional neural network for weed detection and navigation path extraction. *Computers and Electronics in Agriculture* 2025; **229**, 109776.
- [18] N Paul, GC Sunil, D Horvath and X Sun. Deep learning for plant stress detection: A comprehensive review of technologies,

- challenges, and future directions. *Computers and Electronics in Agriculture* 2025; **229**, 109734.
- [19] Y Peng and Y Wang. Cnn and transformer framework for insect pest classification. *Ecological Informatics* 2022; **72**, 101846.
- [20] CC Olisah, B Trehwella, B Li, ML Smith, B Winstone, EC Whitfield, FF Fernández and H Duncalfe. Convolutional neural network ensemble learning for hyperspectral imaging-based blackberry fruit ripeness detection in uncontrolled farm environment. *Engineering Applications of Artificial Intelligence* 2024; **132**, 107945.
- [21] M Tzelepi and A Tefas. Improving the performance of lightweight cnns for binary classification using quadratic mutual information regularization. *Pattern Recognition* 2020; **106**, 107407.
- [22] C Chang, R Parthiban, V Kalavally, YM Hung and X Wang. Unharvested palm fruit bunch ripeness detection with hybrid color correction. *Smart Agricultural Technology* 2024; **9**, 100643.
- [23] K Kipli, S Osman, A Joseph, H Zen, DNSDA Salleh, A Lit and KL Chin. Deep learning applications for oil palm tree detection and counting. *Smart Agricultural Technology* 2023; **5**, 100241.
- [24] R Kurniawan, Samsuryadi, FS Mohamad, HOL Wijaya and B Santoso. Classification of palm oil fruit ripeness based on alexnet deep convolutional neural network. *Sinergi* 2025; **29(1)**, 207-220.
- [25] X Wang, Y Hu, CK Ang, MI Solihin, JJ Tiang and WH Lim. Hyperspectral imaging-based deep learning benchmarks in non-destructive testing of cherry tomatoes. *Applied Food Research* 2025; **5(2)**, 101387.
- [26] MAM Zaki, J Ooi, WPQ Ng, BS How, HL Lam, DC Foo and CH Lim. Impact of industry 4.0 technologies on the oil palm industry: A literature review. *Smart Agricultural Technology* 2025; **10**, 100685.
- [27] J Ng, IY Liao, MF Jelani, ZY Chen, CK Wong and WC Wong. Multiview-based method for high-throughput quality classification of germinated oil palm seeds. *Computers and Electronics in Agriculture* 2024; **218**, 108684.
- [28] FA Junior and Suharjito. Video based oil palm ripeness detection model using deep learning. *Heliyon* 2023; **9(1)**, e13036.
- [29] J Roseleena, J Nursuriati, J Ahmed and CY Low. Assessment of palm oil fresh fruit bunches using photogrammetric grading system. *International Food Research Journal* 2011; **18(3)**, 999-1005.
- [30] MSM Alfatni, S Khairunniza-Bejo, MHB Marhaban, OM Ben Saaed, A Mustapha and AR Shariff. Towards a real-time oil palm fruit maturity system using supervised classifiers based on feature analysis. *Agriculture* 2022; **12(9)**, 1461.
- [31] N Khan, MA Kamaruddin, UU Sheikh, Y Yusup and MP Bakht. Oil palm and machine learning: reviewing one decade of ideas, innovations, applications, and gaps. *Agriculture* 2021; **11(9)**, 832.
- [32] N Donthu, S Kumar, D Mukherjee, N Pandey and WM Lim. How to conduct a bibliometric analysis: an overview and guidelines. *Journal of Business Research* 2021; **133**, 285-296.
- [33] R Wanison, WNH Syahputra, N Kammuang-lue, P Sakulchangsattajatai, C Chaichana, VU Shankar, P Suttakul and Y Mona. Engineering aspects of sodium-ion battery: An alternative energy device for lithium-ion batteries. *Journal of Energy Storage* 2024; **100(A)**, 113497.
- [34] MZ Bin Amiruddin, A Samsudin, A Suhandi, B Coştu and BK Prahani. Scientific mapping and trend of conceptual change: a bibliometric analysis. *Social Sciences & Humanities Open* 2025; **11**, 101208.
- [35] NJ van Eck and L Waltman. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 2010; **84(2)**, 523-538.
- [36] W Aji and K Hawari. A study of deep learning method opportunity on palm oil ffb (fresh fruit bunch) grading methods. *Procedia Computer Science* 2019; **189**, 35-38.
- [37] ZS Ulucak, P Sivashankar, S Huseynov and R Ulucak. Understanding barriers to agricultural technology adoption: evidence from U.S. agribusiness firms. *Technological Forecasting and Social Change* 2026; **224**, 124448.
- [38] T Dibbern, LAS Romani and SMFS Massruhá. Main drivers and barriers to the adoption of digital agriculture technologies. *Smart Agricultural Technology* 2024; **8**, 100459.
- [39] Y Liu, J Xue, D Li, W Zhang, TK Chiew and Z

- Xu. Image recognition based on lightweight convolutional neural network: Recent advances. *Image and Vision Computing* 2024; **146**, 105037.
- [40] N Mamat, MF Othman, R Abdulghafor, AA Alwan and Y Gulzar. Enhancing image annotation technique of fruit classification using a deep learning approach. *Sustainability* 2023; **15(2)**, 901.
- [41] Suharjito, GN Elwirehardja and JS Prayoga. Oil palm fresh fruit bunch ripeness classification on mobile devices using deep learning approaches. *Computers and Electronics in Agriculture* 2021; **188**, 106359.
- [42] Z Ibrahim, N Sabri and D Isa. Palm oil fresh fruit bunch ripeness grading recognition using convolutional neural network. *Journal of Telecommunication, Electronic and Computer Engineering* 2018; **10(3-2)**, 109-113.
- [43] JW Lai, HR Ramli, LI Ismail and WZW Hasan. Real-time detection of ripe oil palm fresh fruit bunch based on YOLOv4. *IEEE Access* 2022; **10**, 95763-95770.
- [44] H Herman, TW Cenggoro, A Susanto and B Pardamean. Deep learning for oil palm fruit ripeness classification with densenet. *International Conference on Information Management and Technology* 2021; **1**, 116-119.
- [45] Suharjito, FA Junior, YP Koeswandy, Debi, PW Nurhayati, M Asrol and Marimin. Annotated datasets of oil palm fruit bunch piles for ripeness grading using deep learning. *Scientific Data* 2023; **10(1)**, 72.
- [46] Suharjito, M Asrol, DN Utama, FA Junior and Marimin. Real-time oil palm fruit grading system using smartphone and modified YOLOv4. *IEEE Access* 2023; **11**, 59758-59773.
- [47] NAMB Selvam, Z Ahmad and IA Mohtar. Real time ripe palm oil bunch detection using YOLOv3 Algorithm. In: Proceedings of the 19<sup>th</sup> Student Conference on Research and Development (SCORED), Kota Kinabalu, Malaysia. 2021, p. 323-328.
- [48] ZY Wong, WJ Chew and SK Phang. Computer vision algorithm development for classification of palm fruit ripeness. *AIP Conference Proceedings* 2020; **2233**, 030012.
- [49] P Rama Rao and G Ramakrishna. Oil palm empty fruit bunch fiber: Surface morphology, treatment, and suitability as reinforcement in cement composites- a state of the art review. *Cleaner Materials* 2022; **6**, 100144.
- [50] LS Woittiez, MT van Wijk, M Slingerland, M van Noordwijk and KE Giller. Yield gaps in oil palm: A quantitative review of contributing factors. *European Journal of Agronomy* 2017; **83**, 57-77.
- [51] FS Ali, R Shamsudin and R Yunus. The effect of storage time of chopped oil palm fruit bunches on the palm oil quality. *Agriculture and Agricultural Science Procedia* 2014; **2**, 165-172.
- [52] M Makky and P Soni. Development of an automatic grading machine for oil palm fresh fruits bunches (FFBS) based on machine vision. *Computers and Electronics in Agriculture* 2013; **93**, 129-139.
- [53] M Rizzo, M Marcuzzo, A Zangari, A Gasparetto and A Albarelli. Fruit ripeness classification: A survey. *Artificial Intelligence in Agriculture* 2023; **7**, 44-57.
- [54] SM Mandal and D Paul. *Spectroscopy: Principle, types and microbiological applications*. In: SM Mandal and DP Paul (Eds.). Automation and basic techniques in medical microbiology. Springer, New York, 2022, p. 49-75.
- [55] MN Hussain, KN Basri, S Arshad, S Mustafa, MFA Khir and J Bakar. Analysis of lard in palm oil using long-wave near-infrared (LW-NIR) spectroscopy and gas chromatography-mass spectroscopy (GC-MS). *Food Analytical Methods* 2023; **16(2)**, 349-355.
- [56] K Srinath, AH Kiranmayee, S Bhanot and PC Panchariya. Detection of palm oil adulteration in sunflower oil using ATR-MIR spectroscopy coupled with chemometric algorithms. *Mapan* 2022; **37(3)**, 483-493.
- [57] S Liang, G Chen, C Ma, J Gu, C Zhu, L Li, H Gao, Z Yang, J Cao and Z Chen. Qualitative identification and adulteration quantification of extra virgin olive oil based on raman spectroscopy combined with multi-task deep learning model. *Food Analytical Methods* 2025; **18(3)**, 385-397.
- [58] M Wu, M Li, B Fan, Y Sun, L Tong, F Wang and L Li. A rapid and low-cost method for detection of nine kinds of vegetable oil adulteration based on 3-d fluorescence spectroscopy. *LWT* 2023; **188**,

- 115419.
- [59] P Phimpisan and C Chungchoo. Real-time oil palm ripeness classification of fresh fruit bunches using fluorescence technology. *Agriculture and Natural Resources* 2023; **57(5)**, 859-868.
- [60] M Makky. A portable low-cost non-destructive ripeness inspection for oil palm FFB. *Agriculture and Agricultural Science Procedia* 2016; **9**, 230-240.
- [61] NK Bharathi, F Fj and S Shoba. Rgb image dataset for okra maturity classification to enhance agricultural quality and market readiness. *Data in Brief* 2025; **62**, 111982.
- [62] S Saifullah, DB Prasetyo, Indahyani, R Dreżewski and FA Dwiyanto. Palm oil maturity classification using k-nearest neighbors based on rgb and lab color extraction. *Procedia Computer Science* 2023; **225**, 3011-3020.
- [63] A Taparugssanagorn, S Siwamogsatham and C Pomalaza-Ráez. A non-destructive oil palm ripeness recognition system using relative entropy. *Computers and Electronics in Agriculture* 2015; **118**, 340-349.
- [64] MSM Alfatni, AR Mohamed Shariff, OM Ben Saaed, AM Albhah and A Mustapha. Colour feature extraction techniques for real time system of oil palm fresh fruit bunch maturity grading. *IOP Conference Series: Earth and Environmental Science* 2020; **540(1)**, 012092.
- [65] J Guo, J Ma, ÁF García-Fernández, Y Zhang and H Liang. A survey on image enhancement for low-light images. *Heliyon* 2023; **9(4)**, e14558.
- [66] BS Luka, BM Yunusa, QM Vihikwagh, KF Kuhwa, TH Oluwasegun, R Ogalagu, TK Yuguda and M Adnoui. Hyperspectral imaging systems for rapid assessment of moisture and chromaticity of foods undergoing drying: Principles, applications, challenges, and future trends. *Computers and Electronics in Agriculture* 2024; **224**, 109101.
- [67] R Budiman, KB Seminar and Sudradjat. The estimation of nutrient content using multispectral image analysis in palm oil (*Elaeis guineensis* Jacq). *IOP Conference Series: Earth and Environmental Science* 2022; **974(1)**, 012062.
- [68] OM Bensaeed, AM Shariff, AB Mahmud, H Shafri and M Alfatni. Oil palm fruit grading using a hyperspectral device and machine learning algorithm. *IOP Conference Series: Earth and Environmental Science* 2014; **20(1)**, 012017.
- [69] C Vaiphasa. Consideration of smoothing techniques for hyperspectral remote sensing. *Isprs Journal of Photogrammetry and Remote Sensing* 2006; **60(2)**, 91-99.
- [70] BG Ram, P Oduor, C Igathinathane, K Howatt and X Sun. A systematic review of hyperspectral imaging in precision agriculture: Analysis of its current state and future prospects. *Computers and Electronics in Agriculture* 2024; **222**, 109037.
- [71] AW Setiawan and OE Prasetya. Palm oil fresh fruit bunch grading system using multispectral image analysis in HSV. In: Proceedings of the IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIOT), Doha, Qatar. 2020, p. 85-88.
- [72] P Junkwon, T Takigawa, H Okamoto, H Hasegawa, M Koike, K Sakai, J Siruntawineti, W Chaeychomsri, A Vanavichit, P Tittinuchanon and B Bahalayodhin. Hyperspectral imaging for nondestructive determination of internal qualities for oil palm (*Elaeis guineensis* Jacq. var. tenera). *Journal of Agricultural Intelligence Research* 2009; **18(3)**, 130-141.
- [73] S Zolfagharnassab, ARBM Shariff, R Ehsani, HZ Jaafar and I Bin Aris. Classification of oil palm fresh fruit bunches based on their maturity using thermal imaging technique. *Agriculture* 2022; **12(11)**, 1-20.
- [74] BI Ismail, MNM Sehmi, H Ahmad, SH Baharom and MF Khalid. Robotic research platform for agricultural environment: Unmanned ground vehicle for oil palm plantation. *Journal of Cases on Information Technology* 2023; **25(1)**, 1-32.
- [75] M Shiddiq, S Saktioto, R Salambue, F Wardana, VV Dasta, IO Harmailil, MF Rabin, N Arpyanti and D Wahyudi. Multispectral imaging and deep learning for oil palm fruit bunch ripeness detection. *Bulletin of Electrical Engineering and Informatics* 2024; **13(6)**, 4168-4181.
- [76] MG Naftali and G Hugo. Palm oil counter: State-of-the-art deep learning models for detection and counting in plantations. *IEEE Access* 2024; **12**, 90395-90417.
- [77] A Septiarini, A Sunyoto, H Hamdani, AA Kasim,

- F Utamingrum and HR Hatta. Machine vision for the maturity classification of oil palm fresh fruit bunches based on color and texture features. *Scientia Horticulturae* 2021; **286**, 110245.
- [78] M Makky and P Soni. Development of an automatic grading machine for oil palm fresh fruits bunches (FFBS) based on machine vision. *Computers and Electronics in Agriculture* 2013; **93**, 129-139.
- [79] WK Fauziah, M Makky, Santosa and D Cherie. Thermal vision of oil palm fruits under difference ripeness quality. *IOP Conference Series: Earth and Environmental Science* 2021; **644(1)**, 012044.
- [80] N Fadilah, J Mohamad-Saleh, ZA Halim, H Ibrahim and SSS Ali. Intelligent color vision system for ripeness classification of oil palm fresh fruit bunch. *Sensors* 2012; **12(10)**, 14179-14195.
- [81] D Cherie, N Fatmawati and M Makky. Non-destructive evaluation of oil palm fresh fruit bunch quality using thermal vision. *IOP Conference Series: Earth and Environmental Science* 2021; **644(1)**, 012024.
- [82] B Büyükarıkan and E Ülker. Using convolutional neural network models illumination estimation according to light colors. *Optik* 2022; **271**, 170058.
- [83] R Zhang, L Du, Q Xiao and J Liu. Comparison of backbones for semantic segmentation network. *Journal of Physics: Conference Series* 2020; **1544(1)**, 012196.
- [84] O Salazar-Campos, J Moran Ruiz, JL Peralta, MR Cieza, BS Medina and J Salazar-Campos. Deep learning approach for automated 'Kent' mango maturity grading in compliance with peruvian standards. *Results in Control and Optimization* 2025; **20**, 100589.
- [85] J Naranjo-Torres, M Mora, R Hernández-García, RJ Barrientos, C Fredes and A Valenzuela. A review of convolutional neural network applied to fruit image processing. *Applied Sciences* 2020; **10(10)**, 3443.
- [86] Y Jiang and C Li. Convolutional neural networks for image-based high-throughput plant phenotyping: a review. *Plant Phenomics* 2020; **2020**, 4152816.
- [87] S Castillo-Girones, S Munera, M Martínez-Sober, J Blasco, S Cubero and J Gómez-Sanchis. Artificial neural networks in agriculture, the core of artificial intelligence: What, when, and why. *Computers and Electronics in Agriculture* 2025; **230**, 109938.
- [88] D Gertsvolf, M Horvat, D Aslam, A Khademi and U Berardi. A u-net convolutional neural network deep learning model application for identification of energy loss in infrared thermographic images. *Applied Energy* 2024; **360**, 122696.
- [89] Z Geradts, N Filius and A Ruifrok. Interpol review of imaging and video 2016 - 2019. *Forensic Science International: Synergy* 2020; **2**, 540-562.
- [90] H Gong, L Liu, H Liang, Y Zhou and L Cong. A state-of-the-art survey of deep learning models for automated pavement crack segmentation. *International Journal of Transportation Science and Technology* 2024; **13**, 44-57.
- [91] R Srivastava, S Kumar and B Kumar. *Classification model of machine learning for medical data analysis*. In: T Goswami and GR Sinha (Eds.). *Statistical modeling in machine learning*. Academic Press, London, 2023, p. 111-132.
- [92] K Maharana, S Mondal and B Nemade. A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings* 2022; **3(1)**, 91-99.
- [93] A Mumuni and F Mumuni. Data augmentation: A comprehensive survey of modern approaches. *Array* 2022; **16**, 100258.
- [94] Suharjito, MG Naftali, G Hugo, MRA Priyadi, M Asrol and DN Utama. Oil palm fruits dataset in plantations for harvest estimation using digital census and smartphone. *Scientific Data* 2025; **12(1)**, 972.
- [95] K Chakraborty, S Bhattacharyya, R Bag and AA Hassanien. *Sentiment analysis on a set of movie reviews using deep learning techniques*. In: N Dey, S Borah and Rosalina (Eds.). *Social network analytics: Computational research methods and techniques*. Academic Press, London, 2019, p. 127-147.
- [96] G Arisandi and N Surantha. *Sleep stage classification using a convolutional neural network based on heart rate variability features*. In: N Surantha (Ed.). *Digital healthcare in Asia and Gulf region for healthy aging and more*

- inclusive societies. Academic Press, London, 2024, p. 115-127.
- [97] HC Chen, AM Widodo, A Wisnujati, M Rahaman, JCW Lin, L Chen and CE Weng. Alexnet convolutional neural network for disease detection and classification of tomato leaf. *Electronics* 2022; **11(6)**, 951.
- [98] WB Zulfikar, AR Atmadja, SLD Agustini, YA Gerhana and DS Maylawati. VGG16 model architecture for assessing cucumber ripeness using image analysis. *In: Proceedings of the 12<sup>th</sup> International Conference on Cyber and IT Service Management, Batam, Indonesia. 2024*, p. 1-6.
- [99] MSM Alfatni, ARM Shariff, MZ Abdullah, MH Marhaban, SB Shafie, MD Bamiruddin and OMB Saaed. Oil palm fresh fruit bunch ripeness classification based on rule-based expert system of roi image processing technique results. *IOP Conference Series: Earth and Environmental Science* 2014; **20(1)**, 012018.
- [100] YA Tan, KW Low, CK Lee and KS Low. Imaging technique for quantification of oil palm fruit ripeness and oil content. *European Journal of Lipid Science and Technology* 2010; **112(8)**, 838-843.
- [101] R Girshick, J Donahue, T Darrell and J Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, United States. 2014*, p. 580-587.
- [102] T Liang, X Chu, Y Liu, Y Wang, Z Tang, W Chu, J Chen and H Ling. Cbnet: A composite backbone network architecture for object detection. *IEEE Transactions on Image Processing* 2022; **31**, 6893-6906.
- [103] MH Junos, AS Mohd Khairuddin, S Thannirmalai and M Dahari. An optimized yolo-based object detection model for crop harvesting system. *Iet Image Processing* 2021; **15(9)**, 2112-2125.
- [104] SY Mohammed. Architecture review: two-stage and one-stage object detection. *Franklin Open* 2025; **12**, 100322.
- [105] R Varghese and M Sambath. A comprehensive review on: Two-stage object detection algorithms. *In: Proceedings of the International Conference on Quantum Technologies, Communications, Computing, Hardware and Embedded Systems Security, Kottayam, India. 2023*, p. 1-7.
- [106] NN Malvade, R Yakkundimath, G Saunshi, MC Elemmi and P Baraki. A comparative analysis of paddy crop biotic stress classification using pre-trained deep neural networks. *Artificial Intelligence in Agriculture* 2022; **6**, 167-175.
- [107] S Koley. Critically reckoning spectrophotometric detection of asymptomatic cyanotoxins and faecal contamination in periurban agrarian ecosystems via convolutional neural networks. *Trends in Sciences* 2024; **21(12)**, 8528.
- [108] AK Chowdhury, WZW Said, AU Rehman, LG Lie, NA Izni and MS Jahan. Comparative study of different YOLO variants for ripeness detection of palm oil fresh fruit bunches. *In: Proceedings of the 6<sup>th</sup> International Conference in Robotics and Manufacturing Automation (ROMA), Selangor, Malaysia. 2025*, p. 18-23.
- [109] H Wibowo, IS Sitanggang, M Mushthofa and HA Adrianto. Large-scale oil palm trees detection from high-resolution remote sensing images using deep learning. *Big Data and Cognitive Computing* 2022; **6(3)**, 89.
- [110] D Sri Rejeki, E Irwansyah and B Golda. Analysis of comparison Faster R-CNN backbone for oil palm tree counting using drone imagery data. *In: Proceedings of the International Conference on Information Technology and Computing (ICITCOM) ICITCOM 2024, Yogyakarta, Indonesia. 2024*, p. 230-235.
- [111] ML Ali and Z Zhang. The yolo framework: A comprehensive review of evolution, applications, and benchmarks in object detection. *Computers* 2024; **13(12)**, 336.
- [112] M Naghipour, LS Ling and T Connie. YOLO-based oil palm FFB ripeness detection. *In: Proceedings of the International Conference on Electrical, Communication and Computer Engineering (ICECCE), Kuala Lumpur, Malaysia. 2024*, p. 1-5.
- [113] TS Gunawan, M Kartiwi, H Mansor and NM Yusoff. Palm fruit ripeness detection and classification using various YOLOv8 models. *In: Proceedings of the 9<sup>th</sup> International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA), Kuala Lumpur,*

- Malaysia. 2023, p. 193-198.
- [114]SJ Purba, WF Tandion and E Irwansyah. Comparison of the latest version of deep learning YOLO model in automatically detecting the ripeness level of oil palm fruit in PTPN IV, North Sumatra, Indonesia. *In: Proceedings of the International Conference on Information Technology and Computing (ICITCOM)*, Yogyakarta, Indonesia. 2024, p. 295-300.
- [115]SNABM Robi, MABM Izhar, MB Sahrim and NB Ahmad. Image detection and classification of oil palm fruit bunches. *In: Proceedings of the 4<sup>th</sup> International Conference on Smart Sensors and Application (ICSSA)*, Kuala Lumpur, Malaysia. 2022, p. 108-113.
- [116]M Yasser, M Ahmed, KD Dambul and KY Choo. Object detection algorithms for ripeness classification of oil palm fresh fruit. *International Journal of Advanced Computer Science and Applications* 2022; **13**, 1326-1335.
- [117]P Chotikawanid, P Saeleung, Y Pianroj, S Jumrat, T Punvichai and J Muangprathub. Optimizing an object detection algorithm for detecting oil palm fruit bunches and their ripeness. *Applied Computational Intelligence and Soft Computing* 2025; **2025(1)**, 6263757.
- [118]AK Chowdhury, WZBW Said and S Saruchi. *Oil palm fresh fruit branch ripeness detection using YOLOv6 algorithm*. *In: R Hamidon, MS Bahari, JM Sah and Z Zainal Abidin (Eds.). Intelligent manufacturing and mechatronics*. Springer, Singapore, 2024.
- [119]N Khamis, H Selamat, S Ghazalli, NIM Saleh and N Yusoff. Comparison of palm oil fresh fruit bunches (FFB) ripeness classification technique using deep learning method. *In: Proceedings of the 13<sup>th</sup> Asian Control Conference (ASCC)*, Jeju, Korea. 2022, p. 64-68.
- [120]H Shen, Z Wang, J Zhang and M Zhang. L-Net: A lightweight convolutional neural network for devices with low computing power. *Information Sciences* 2024; **660**, 120131.
- [121]S Duhan, P Gulia, NS Gill and E Narwal. RTR\_Lite\_MobileNetV2: A lightweight and efficient model for plant disease detection and classification. *Current Plant Biology* 2025; **42**, 100459.
- [122]H Guan, C Fu, G Zhang, K Li, P Wang and Z Zhu. A lightweight model for efficient identification of plant diseases and pests based on deep learning. *Frontiers in Plant Science* 2023; **14**, 1227011.
- [123]S Zhu and H Gao. Mc-shufflenetv2: a lightweight model for maize disease recognition. *Egyptian Informatics Journal* 2024; **27**, 100503.
- [124]I Sonata and Y Arifin. Deep learning approach for palm fruit ripeness classification using MobileNet. *In: Proceedings of the 4<sup>th</sup> International Conference on Creative Communication and Innovative Technology (ICCI)*, Kota Cirebon, Indonesia, 2025, p. 1-7.
- [125]A Accu and A Accuracy. Oil palm fresh fruit bunch ripeness classification by deep learning. *Computers and Electronics in Agriculture* 2023; **121(C)**, 81-104.
- [126]NM Zamri and AK Anuar. Palm fresh fruit bunches (FFBS) colour grading system using raspberry pi. *Journal of Physics: Conference Series* 2023; **4(2)**, 574-581.
- [127]MH Junos, AS Mohd Khairuddin, MS Abu Talip, MI Kairi and YM Siran. Improved hybrid feature extractor in lightweight convolutional neural network for postharvesting technology: Automated oil palm fruit grading. *Neural Computing and Applications* 2024; **36(32)**, 20473-20491.
- [128]S Kalamkar and GM A. Multimodal image fusion: A systematic review. *Decision Analytics Journal* 2023; **9**, 100327.
- [129]M Vahidi, S Shafian and WH Frame. Multi-modal sensing for soil moisture mapping: Integrating drone-based ground penetrating radar and rgb-thermal imaging with deep learning. *Computers and Electronics in Agriculture* 2025; **236**, 110423.
- [130]P Pipitsunthonsan, L Pan, S Peng, T Khaorapapong, S Nakasathien, S Channumsin and M Chongcheawchamnan. Palm bunch grading technique using a multi-input and multi-label convolutional neural network. *Computers and Electronics in Agriculture* 2023; **210**, 107864.
- [131]X Guo, Q Feng and F Guo. CMTNet: A hybrid cnn-transformer network for uav-based hyperspectral crop classification in precision agriculture. *Scientific Reports* 2025; **15(1)**, 12383.
- [132]R Jaiswal and GK Jha. Genetic algorithm

- optimized CNN-LSTM model for forecasting of agriculture commodity prices. *Agricultural Economics Research Review* 2025; **38(1)**, 106-121.
- [133] A Tuerxun, AR Mohamed Shariff, R Janius, Z Abbas and GA Mahdiraji. Oil palm fresh fruit bunches maturity prediction by using optical spectrometer. *IOP Conference Series: Earth and Environmental Science* 2020; **540(1)**, 012085.
- [134] J Nainggolan, DY Niska, F Marpaung, I Taufik and KS S. Palm fruit ripeness detection system using convolutional neural network (CNN) algorithm. *Journal of Physics: Conference Series* 2025; **4(3)**, 1700-1705.
- [135] M Rafieizonooz, HTTL Pham, S Han, J Seo and E Khankhaje. Influence of data source and volume on CNN applications in construction. *Automation in Construction* 2025; **179**, 106476.
- [136] R El yazid, EF Sanaa and B El Habib. Benchmarking normalization methods for a cnn based object detection computer vision model. *Procedia Computer Science* 2025; **265**, 560-565.
- [137] M Basyuni, N Amri, LAP Putri, I Syahputra and D Arifiyanto. Characteristics of fresh fruit bunch yield and the physicochemical qualities of palm oil during storage in North Sumatra, Indonesia. *Indonesian Journal of Chemistry* 2017; **17(2)**, 182-190.
- [138] WNH Syahputra, C Chaichana, D Bundhurat, P Wuttisarnwattana, BTW Putra and Y Wijayanto. Comparative analysis of oil palm fruit free fatty acid estimation using a portable VIS-NIR spectrometer and multispectral imaging. *IOP Conference Series: Earth and Environmental Science* 2026; **1584(1)**, 12033.
- [139] N Liu, M Rogers, H Cui, W Liu, X Li and P Delmas. Deep convolutional neural networks for regular texture recognition. *PeerJ Computer Science* 2022; **8**, e869.
- [140] S Emebu, O Osaikhuiwuomwan, A Mankonen, C Udoye, C Okieimen and D Janáčová. Influence of moisture content, temperature, and time on free fatty acid in stored crude palm oil. *Scientific Reports* 2022; **12(1)**, 1-11.
- [141] ES Mohamed, AM Saleh, AB Belal and A Gad. Application of near-infrared reflectance for quantitative assessment of soil properties. *Egyptian Journal of Remote Sensing and Space Sciences* 2018; **21(1)**, 1-14.
- [142] CH Lim, ZH Cheah, XH Lee, BS How, WPQ Ng, SL Ngan, S Lim and HL Lam. Harvesting and evacuation route optimisation model for fresh fruit bunch in the oil palm plantation site. *Journal of Cleaner Production* 2021; **307**, 127238.
- [143] MD Islam, D Choi, X Zhou and X Wang. Comparative evaluation of depth imaging and 3D point cloud sensing for automatic strawberry runner cutting. *Smart Agricultural Technology* 2026; **13**, 101819.
- [144] E Salim. Hyperparameter optimization of YOLOv4 tiny for palm oil fresh fruit bunches maturity detection using genetics algorithms. *Smart Agricultural Technology* 2023; **6**, 100364.