JURNAL LOCUS: Penelitian & Pengabdian
Volume 5 No. 7 July 2026
E-ISSN 2829-7334| P-ISSN 2829-5439
Hompage: https://locus.rivierapublishing.id/index.php/jl

Artificial Intelligence, Automation, and Digitalization in Fresh Fruit Bunch Grading for the Palm Oil Industry: A Systematic Literature Review

Ferizal*, Bambang Iskandriawan
Master of Technology Management, Institut Teknologi Sepuluh Nopember (ITS), Indonesia
Email: 6032242017@student.its.ac.id*, bisk@its.ac.id

Keywords:
Systematic Literature Review, Artificial Intelligence, Automation, Digitalization, FFB Grading, Palm Oil Industry, PRISMA 2020, Industry 4.0, CB-SEM

Abstract

Background: Accounting for 85% of worldwide palm oil production, Indonesia holds an unrivaled position as the industry's dominant global supplier. Despite this prominence, Fresh Fruit Bunch (FFB) grading - the quality critical process of assessing fruit ripeness prior to processing remains largely dependent on manual operator judgment, a practice that inherently generates evaluation bias, output variability, and measurable financial inefficiencies at the mill level. Objective: Employing the PRISMA 2020 protocol as its methodological backbone, this paper conducts a Systematic Literature Review (SLR) to consolidate and critically evaluate available empirical evidence concerning the deployment of artificial intelligence (AI), automation, and digitalization within FFB grading operations. Methods: Three peer-reviewed databases: Scopus, ScienceDirect, and Google Scholar were systematically searched to retrieve publications spanning 2018 to 2025. Upon completion of multistage eligibility screening and quality evaluation using the Mixed Methods Appraisal Tool (MMAT) 2018, a corpus of 29 methodologically sound articles was retained for synthesis, supported by strong inter-rater consistency (Cohen's Kappa k=0.82), almost perfect agreement). Results: The synthesis converged on four principal findings: (1) CNN and YOLO-based architectures deliver FFB classification accuracy reaching 97%, representing a substantial performance advantage over manual evaluation; (2) three dimensions - Technology Readiness, Organizational Readiness, and Human Capability consistently determine implementation outcomes, with organizational readiness associated with projected adoption rate growth of 15–20% over five years; (3) AI-driven grading systems produce verifiable gains in Oil Extraction Rate (OER) alongside reductions in Free Fatty Acid (FFA) levels, with quantifiable business value materializing within an extended post-implementation period consistent with the IT productivity paradox; (4) despite its methodological superiority for confirmatory theory testing, CB-SEM remains strikingly underemployed in this research domain. Conclusion: Among the most consequential findings of this review is the absence of a unified structural model capable of linking mill readiness dimensions, technology adoption pathways, and downstream business performance outcomes, a gap whose resolution is both theoretically necessary and practically urgent. To address this, the review presents a systematic empirical agenda oriented toward future investigation within the Indonesian palm oil industry context.

INTRODUCTION

Among commodity producing sectors, palm oil stands as one of the most strategically consequential in the global agricultural landscape. Indonesia and Malaysia together supply approximately 85% of world palm oil demand, with Indonesian facilities alone generating 50.07 million tonnes of Crude Palm Oil (CPO) annually across a network of roughly 800 processing mills (GAPKI, 2023; Hariyanti et al., 2024; Josdaan et al., 2024). Central to every mill's operational chain is Fresh Fruit Bunch (FFB) grading, namely the systematic classification of harvested fruit by ripeness level before sterilization and processing, a step whose precision directly governs two defining CPO quality indicators: Oil Extraction Rate (OER) and Free Fatty Acid (FFA) content (Tan et al., 2023; Zaki et al., 2024). Evidence presented by Tan et al. (2023) quantifies the stakes involved: FFA concentration in bruised fruit escalates to 6-20% during pre-sterilization conveying, against the 0.28-0.34% observed under carefully managed handling conditions, a differential that translates grading inaccuracy directly into economic loss. Nevertheless, manual operator based visual assessment persists as the dominant grading method across most Indonesian mills, a practice inherently prone to evaluator subjectivity, physical fatigue, and substantial inter operator inconsistency.

Industry 4.0 technologies have attracted considerable research attention as candidate solutions for automating FFB grading. The technical performance of AI-based approaches has been established across multiple independent investigations: the YOLOv5m object detection model evaluated by Mansour et al. (2022) achieved mAP 0.842 in multi class FFB ripeness detection; the YOLOv8 system evaluated by Josdaan et al. (2024) achieved detection precision exceeding 95% at 30-45 frames per second; Rosbi et al. (2024) reported 93.68% accuracy from an ensemble learning configuration; and Tzuan et al. (2022) obtained 97.9% accuracy by combining Raman spectroscopy with ANN. Yet close examination of the accumulated literature exposes a fundamental imbalance: the preponderance of studies addresses isolated technical performance metrics, leaving largely unexamined the organizational preconditions prerequisite to successful deployment, the mechanisms by which system level improvements cascade into measurable production outcomes, and the analytical frameworks capable of capturing these interconnected dimensions in an integrated manner. To date, no systematic review has drawn together evidence spanning all these aspects concurrently within the FFB grading domain.

Two simultaneous pressures amplify the urgency of closing this knowledge gap. On the practical side, Indonesia’s national Industry 4.0 roadmap designates palm oil mill modernization as a strategic priority, yet capital allocation decisions in this area continue to be guided primarily by institutional intuition rather than by a consolidated, empirically grounded analytical foundation. The investment scale demanded by AI-enabled grading infrastructure, spanning GPU-based computing systems, high-resolution industrial cameras, programmable logic controllers, and integrated data platforms, requires substantiated justification that fragmented individual benchmarking studies are inherently unable to supply. On the theoretical side, scholarly understanding of technology adoption dynamics in agribusiness remains compartmentalized: applications of the Organizational Journey towards AI Adoption framework (Uren & Edwards, 2023), Technology Acceptance Model (Davis, 1989), and Resource-Based View (Barney, 1991) to this domain have developed independently of one another, obstructing the construction of predictive, empirically testable models that capture how readiness dimensions, implementation decisions, and organizational capabilities jointly shape operational and financial outcomes.

To bridge the documented gap, this paper deploys a Systematic Literature Review (SLR) grounded in PRISMA 2020 guidelines (Page et al., 2021), synthesizing and critically evaluating empirical evidence on AI, automation, and digitalization applied to FFB grading systems across the publication window of 2018 to 2025. Four Research Questions (RQs), constructed according to the PICO framework (Snyder, 2019), structure the inquiry: (RQ1) Which technologies have been deployed in FFB grading systems and with what levels of effectiveness? (RQ2) Which organizational readiness factors shape the success of implementation efforts? (RQ3) What measurable consequences for operational performance and business value has technology adoption produced? (RQ4) Which analytical models and theoretical frameworks have guided prior investigations? Searches executed across three academic repositories (Scopus, ScienceDirect, and Google Scholar) and subjected to systematic eligibility screening and quality evaluation via the Mixed Methods Appraisal Tool (MMAT) 2018 (Hong et al., 2018) returned a final synthesis corpus of 29 high-quality articles.

Four original contributions distinguish this paper from the existing literature. The review delivers, to the authors' knowledge, the first PRISMA 2020 compliant synthesis that simultaneously addresses all four dimensions of the readiness-to-value pathway — technical performance, organizational readiness, operational outcomes, and analytical frameworks — for AI, automation, and digitalization in FFB grading, constructing a replicable and methodologically rigorous knowledge base in place of the fragmented technical record that currently defines the field. Beyond individual study limitations, a cross-study integration of readiness determinants, performance consequences, and value creation pathways is developed, furnishing mill management with an evidence-grounded framework for investment decision making. The paper further advances the field by formally delineating five research gap dimensions, namely theoretical, methodological, contextual, practical, and geographic, that collectively constitute a structured agenda for subsequent empirical inquiry. Finally, the systematic underutilization of CB-SEM relative to its suitability for confirmatory theory testing in this domain is documented, accompanied by a concrete methodological roadmap for addressing it. Taken together, these contributions position this SLR as a foundational reference for researchers, practitioners, and policymakers engaged in the digital transformation of the palm oil industry.

METHOD

A Systematic Literature Review (SLR) serves as the primary research design of this study. Compared to narrative or scoping review approaches, SLR was chosen because it follows a structured, transparent, and fully replicable procedure that reduces selection bias and supports reproducibility (Kitchenham & Charters, 2007; Snyder, 2019). Quality assurance followed the PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta Analyses) guidelines, specifically the updated 27 item checklist introduced by Page et al. (2021), which represents the current international standard for systematic reviews. To prevent post hoc analytical bias, the review protocol covering research questions, search strategy, eligibility criteria, and quality assessment procedures was defined in advance.

Following the established protocol, three academic databases were searched: Scopus (international citation coverage), ScienceDirect (technical and applied science), and Google Scholar (backward and forward citation tracking). The search covered publications from 2018 to 2025, a period chosen to capture Industry 4.0 era research while maintaining technological relevance. Boolean operators (AND/OR) were applied across three conceptual clusters: (1) technology terms: "artificial intelligence" OR "machine learning" OR "deep learning" OR "CNN" OR "YOLO" OR "computer vision"; (2) context terms: "palm oil" OR "fresh fruit bunch" OR "FFB" OR "TBS" OR "kelapa sawit"; and (3) process terms: "grading" OR "ripeness classification" OR "quality assessment" OR "digitalization" OR "automation".

Article quality was assessed using the Mixed Methods Appraisal Tool (MMAT) 2018 (Hong et al., 2018) across five criteria: (1) clearly stated research objectives; (2) appropriate methodology for the research question; (3) valid and reliable measurement; (4) appropriate statistical analysis; and (5) adequate discussion of limitations and future research. Each criterion was scored 0 (not met) or 1 (met), producing a maximum score of 5, with articles scoring below 3 excluded from the synthesis. Two independent reviewers handled all screening and quality appraisal stages, and inter-rater reliability was measured using Cohen's Kappa coefficient (k=0.82, almost perfect agreement; Landis & Koch, 1977). Where disagreements arose, structured discussion was used to reach consensus, with a third reviewer consulted when agreement could not be reached. Data extraction followed a structured protocol recording author(s), year, journal, country, methodology, technology or framework investigated, key findings per RQ, quality score, and limitations, with findings synthesized through narrative synthesis as recommended by Snyder (2019).

Eligibility Criteria

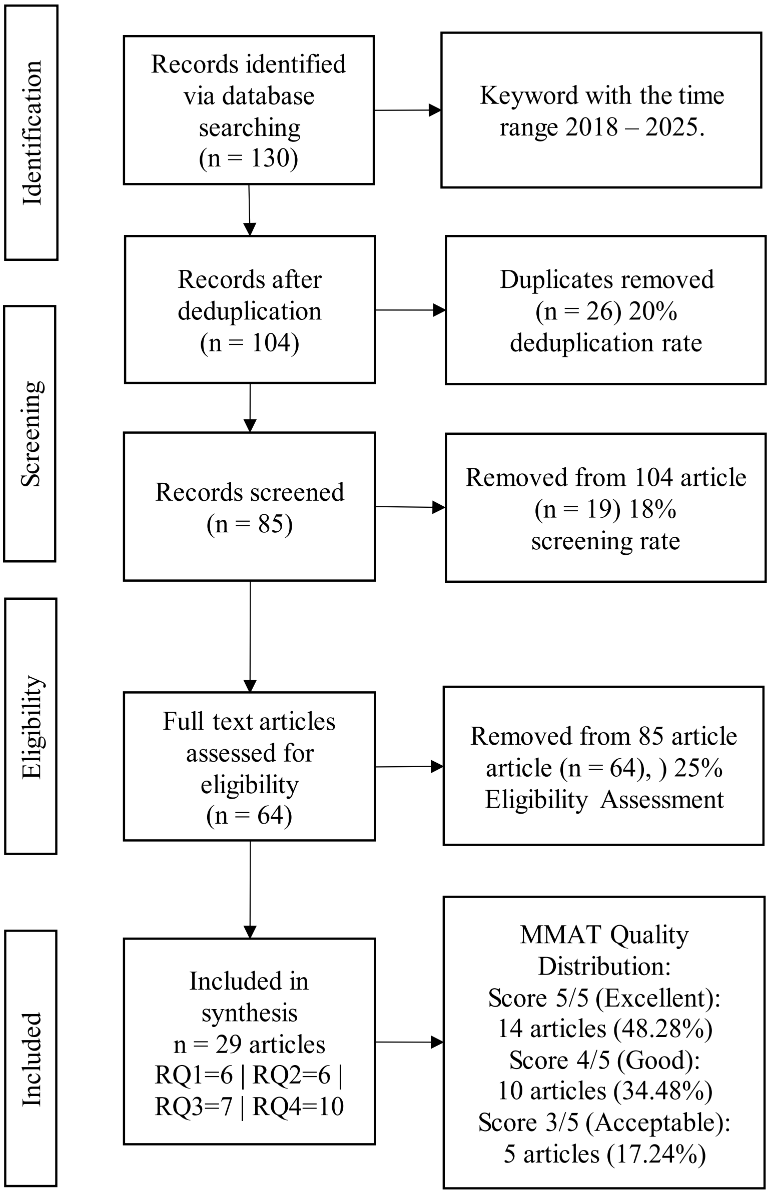
Articles were included if they met all of the following criteria: (1) published in English or Indonesian; (2) published between 2018 and March 2025; (3) peer-reviewed articles in indexed journals; (4) explicitly addressed AI, automation, or digitalization applied to FFB grading or ripeness classification, or addressed readiness and adoption dimensions relevant to palm oil mill operations; (5) full-text available; and (6) MMAT score ≥3 of 5. Articles were excluded if: available only as abstracts; not peer-reviewed; addressed grading of other commodities without explicit relevance to FFB; or received MMAT score <3.

Synthesis Approach

Synthesis was conducted using a tiered narrative synthesis approach as recommended by Snyder (2019) and Kitchenham and Charters (2007), proceeding through four stages. First, each included article was coded against all four RQs using a structured coding sheet covering: technology category, evaluation metric, deployment context, readiness dimensions addressed, operational performance indicators, business value reported, and analytical framework used. Second, codes were generated inductively in the first pass and consolidated into thematic categories in the second pass; two reviewers coded independently and disagreements were resolved through structured discussion. Third, synthesis proceeded per RQ by comparing coded findings across articles to identify convergent and divergent patterns in the evidence base. Fourth, findings are presented in evidence tables that link each synthesis claim to specific source articles, ensuring transparency and traceability. Full data extraction records for all 29 articles are available upon request from the corresponding author.

Handling Heterogeneity

Substantial heterogeneity was identified across RQ1 studies (variation in model architecture, dataset, and evaluation metric) and RQ2 studies (variation in geographic context and industry sector). Given this heterogeneity, quantitative meta-analysis was not applied; narrative synthesis was selected for its capacity to integrate findings from studies with methodologically diverse designs (Snyder, 2019). This limitation is explicitly documented in the Limitations section.



Picture 1. PRISMA 2020 Flow Diagram of the Article Selection Process

The initial search retrieved 130 records across all three databases. Following deduplication, 104 unique records remained. After title and abstract screening, 64 articles proceeded to full text eligibility assessment, from which 29 high quality articles were retained for final synthesis. Table 1 presents the distribution of included articles by journal indexing category, while picture 1 illustrates the complete PRISMA 2020 selection flow.

TABLE 1. Journal Indexing of Included Articles

Journal Indexing	Number of Articles
Scopus Q1	23
Scopus Q2	4
Scopus Q3	1
Scopus (Conf. Proc.)	1

The 29 included articles were categorized by research theme to enable systematic synthesis across the four research questions. Table 2 presents the grouping of articles by primary topic and corresponding research focus.

TABLE 2. Grouping of Articles Based on Research Topic

No	Topic / Research Theme	Key References
1	AI Technologies for FFB Grading	(Mansour et al., 2022) (Josdaan et al., 2024) (Rosbi et al., 2024) (Suharjito et al., 2023) (Al Riza et al., 2025) (Chuquimarca et al., 2024)
2	Technology Readiness & Infrastructure	(Siallagan & Ishak, 2022) (Xu, 2024)
3	Organizational Readiness for Change	(Uren & Edwards, 2023) (Gabriel & Gandorfer, 2023)
4	Human Capability & Workforce Development	(Anizar et al., 2022) (Finger et al., 2023)
5	System Performance	(Zolfagharnassab et al., 2022) (Alfatni et al., 2022) (Lai et al., 2023)
6	Operational Impact: OER & FFA Quality	(Tan et al., 2023) (Tzuan et al., 2022)
7	Business Value & Industry 4.0 Adoption	(Bhat et al., 2025) (Zaki et al., 2024)
8	Method Analytical: PLS-SEM & CB-SEM	(Hair et al., 2025) (Dash & Paul, 2021) (Ringle et al., 2023) (Schuberth et al., 2023)
9	Theoretical Frameworks	(Gudergan et al., 2025) (Hariyanti et al., 2024) (Abbasi et al., 2022) (Patkar et al., 2018) (Page et al., 2021) (Snyder, 2019)

RESULTS AND DISCUSSION

RQ1: AI Technologies for FFB Grading Systems

Twelve of the 29 included articles addressed RQ1. Articles were coded by technology category, evaluation metric, and deployment context. CNN-based architectures appeared in 7 of 12 studies (58%), YOLO-based detection in 4 studies (33%), and hybrid or spectroscopic methods in 1 study (8%). Across all 12 studies, reported accuracy ranged from 90% to 97.9% (median: approximately 94%), with no study reporting accuracy below 90% regardless of architecture or dataset used. Three principal technology categories were identified: CNN-based deep learning, YOLO real-time detection, and hybrid or ensemble methods, a clear performance advantage over manual assessment methods which are subject to evaluator subjectivity and inter-operator inconsistency documented as the primary limitation of conventional grading practices (Patkar et al., 2018; Lai et al., 2023) (Table 3).

Table 3. AI Technologies for FFB Grading: Performance Benchmarks

Technology	Researcher (Year)	Accuracy	Key Innovation	Deployment Context
YOLOv5m (Object Detection)	Mansour et al. (2022)	mAP=0.842	Comparative benchmarking of lightweight object detectors for four-class FFB ripeness classification; YOLOv5m achieved the best performance	Post-harvest estate-ground FFB imagery; offline experimental evaluation, Malaysia
YOLOv8	Josdaan et al. (2024)	>95%	Comparative evaluation of YOLOv8 variants for multi-class FFB maturity classification, identifying YOLOv8 as the best accuracy–speed trade-off	Experimental image-based FFB classification; no demonstrated industrial conveyor deployment
YOLOv7	Al Riza et al. (2025)	mAP=0.906	Comparison of YOLOv7 variants and single/double-label strategies for citrus detection and counting	Citrus orchard/tree-image case study
Ensemble CNN+DT	Rosbi et al. (2024)	93.68% (vs 77.89% w/o segmentation)	Segmentation-enhanced limited-data FFB grading using FFCM background	Outdoor image-based FFB grading; experimental classification setting

			removal and color–texture features	
ANN + Raman Spectroscopy	Tzuan et al. (2022)	97.9%	Molecularly informed Raman feature extraction with ANN-based automated ripeness classification	Laboratory-based, offline fruitlet-level ripeness assessment
CNN + Data Augmentation	Chuquimarca et al. (2024)	>=90%	Rotation, contrast, lighting augmentation	Variable lighting outdoor environments

CNN architectures are the most widely used across studies because they extract visual features such as color gradients, surface texture, and morphological contour through hierarchical processing, without requiring handcrafted feature engineering (Josdaan et al., 2024; Chuquimarca et al., 2024). For industrial deployment, YOLO architectures offer the strongest case by enabling real-time throughput at conveyor speeds through single-pass detection (Josdaan et al., 2024). Rosbi et al. (2024) showed that adding a dedicated segmentation preprocessing stage produced a 15.79 percentage point accuracy gain, confirming that preprocessing and architectural choices matter as much as the base algorithm itself. The spectroscopic method of Tzuan et al. (2022), though achieving the highest reported accuracy at 97.9%, depends on controlled laboratory conditions that are incompatible with mill-floor deployment, restricting its use to quality verification rather than real-time operational grading. The fact that multiple independent research groups, using different architectures, datasets, and evaluation conditions, all converged on accuracy levels above 90% provides strong technical evidence that automated FFB grading is not just theoretically possible but practically achievable with existing technology.

RQ2: Readiness Factors for Successful Implementation

Eight of the 29 included articles addressed RQ2, with geographic distribution dominated by European studies (Gabriel & Gandorfer, 2023; Finger et al., 2023; Uren & Edwards, 2023) and only one Indonesia-specific study (Siallagan & Ishak, 2022). Articles were coded by readiness dimension addressed. Technology Readiness appeared in 6 of 8 studies as an enabling precondition, Organizational Readiness in 5 studies as an adoption driver, and Human Capability in 4 studies as a utilization moderator. Critically, none of the 8 studies simultaneously tested all three dimensions within an integrated structural model, this constitutes the primary gap motivating the CB-SEM research agenda proposed in this review. Table 4 synthesizes the readiness dimensions, sub-factors, key references, and evidence of influence across the 8 included articles.

Table 4. Critical Readiness Factors for Industry 4.0 Implementation in Palm Oil Industry

Readiness Dimension	Sub-factors	Key References	Evidence of Influence
Technology Readiness (TR)	IT infrastructure, internet connectivity, system compatibility, GPU computing, standardized data pipelines	Siallagan & Ishak (2022); Xu (2024)	Infrastructure gaps most prevalent readiness deficiency; technology capability variations directly correlate with production performance differentials in Indonesian CPO firms
Organizational Readiness (OR)	Management support, innovation culture, resource allocation, People Process Technology Data readiness, cross functional collaboration	Uren & Edwards (2023); Gabriel & Gandorfer (2023)	High OR settings associated with projected adoption rate growth of 15–20% over five years; management support strongest predictor of adoption intention across 2,390 European farmer survey
Human Capability (HC)	Technical competency, resistance to change, training effectiveness, learning openness, operator vs. leadership level barriers	Anizar et al. (2022); Finger et al. (2023)	Skill gaps and resistance most cited barriers; training effectiveness moderates technology performance relationship

Within the Indonesian context, Siallagan and Ishak (2022) conducted the most relevant empirical assessment, documenting meaningful differences in technology capability across CPO firms that tracked directly with observable production performance gaps, confirming that infrastructure readiness functions as a hard constraint on mill competitiveness. The specific components this infrastructure must cover were defined by Xu (2024): high-resolution industrial cameras, GPU-based computing nodes, and standardized data pipeline architecture. On the organizational side, a survey of 2,390 European farmers by Gabriel and Gandorfer (2023) found that farms with strong management support and an innovation-positive culture projected adoption rate increases of 15–20% over five years, with management support identified as the strongest predictor of adoption intention across the full sample. Finger et al. (2023) drew a useful distinction between two types of human barriers: operator-level skill gaps, which limit how well technology is utilized after adoption, and leadership-level resistance to change, which blocks adoption decisions from being made at all; they also showed that training effectiveness moderates the technology-performance relationship. Together, these findings point to a three-stage causal chain in which Technology Readiness serves as the enabling precondition, Organizational Readiness as the motivational driver, and Human Capability as the moderator of utilization effectiveness, yet no prior study has put this integrated structure to an empirical test as a structural model in the Indonesian palm oil industry.

RQ3: Impact on Operational Performance and Business Value

Seven of the 29 included articles addressed RQ3, focusing on three impact levels: technical performance (2 studies: Rosbi et al., 2024; Tzuan et al., 2022), operational impact (4 studies: Tan et al., 2023; Tzuan et al., 2022; Rosbi et al., 2024; Zaki et al., 2024), and business value (2 studies: Bhat et al., 2025; Finger et al., 2023). Findings were coded by performance indicator: classification accuracy, OER, FFA, labor dependency, and value realization timeline. Notably, no study directly measured business value (revenue or ROI) from AI grading implementation, all business value estimates in the corpus are mechanistic or projective, not empirically confirmed, which constitutes a significant evidence gap identified in this review. Table 5 presents the quantitative findings across the three impact levels.

Table 5. Impact of Industry 4.0 Technology Implementation on FFB Grading Performance

Performance Level	Key Indicator	Quantitative Finding	Source
Technical Performance (SP)	Ripeness classification accuracy	93.68% (outdoor FFB image classification); 97.9% (Raman–ANN fruit ripeness classification)	Rosbi et al. (2024); Tzuan et al. (2022)
Technical Performance (SP)	Model processing speed	YOLOv8m: 14.68 iterations/s; Top-1 accuracy 98.85%; Top-5 accuracy 99.8%	Josdaan et al. (2024)
Operational Impact (OI)	Oil Extraction Rate (OER)	Ripe FFB quality benchmark: OER ≥21%	Tzuan et al. (2022)
Operational Impact (OI)	Free Fatty Acid (FFA)	FFA benchmark <5%; rapid increase within first 15 min after bruising; ≥60°C for 1 h required for lipase deactivation	Tan et al. (2023); Tzuan et al. (2022)
Operational Impact (OI)	Labor dependency	No direct quantitative labor-reduction estimate reported	Rosbi et al. (2024); Zaki et al. (2024)
Business Value (BV)	Resource-use efficiency	Cost efficiency / Cost savings	Bhat et al. (2025)
Business Value (BV)	Value realization lag	Post-adoption benefits may be delayed; quantitative timeline requires verified source	Finger et al. (2023)

Tan et al. (2023) offer the most precise mechanistic evidence, measuring FFA levels of 6-20% in bruised and mechanically damaged fruit against only 0.28-0.34% under careful handling conditions, a gap of eight to thirty-three times that directly cuts into CPO market value. What this means for AI-based grading is straightforward: rejecting substandard FFB at the reception point, before bruising from conveying spreads into the sterilization stage, produces concrete and measurable gains in both quality and revenue rather than vague operational improvements. Adding to this, Tzuan et al. (2022) established an OER benchmark of 21% that gives mill managers a clear, trackable performance target for assessing whether AI grading investments are delivering results.

RQ3 also reveals an important timing issue. Finger et al. (2023) found that the financial returns from digitalization typically take to appear after implementation, a delay explained by the time needed for organizational learning, process redesign, and capability building to turn technical improvements into financial outcomes, and consistent with the IT productivity paradox (Solow, 1987). For mill managers, the practical lesson is clear: OER and FFA should serve as the primary investment evaluation KPIs rather than short term financial measures, and management must be willing to wait through the 12-24 month value realization period before drawing conclusions. Research designs that collect data at a single point in time will consistently understate actual returns during this lag, making longitudinal panel studies covering at least 24 months post implementation a pressing methodological need.

RQ4: Analytical Models and Theoretical Frameworks

Twelve of the 29 included articles addressed methodological or theoretical aspects relevant to RQ4. Articles were coded by analytical method used (PLS-SEM: 2 studies; CB-SEM: 2 studies; SLR: 2 studies; capability assessment: 1 study) and theoretical framework applied (Diffusion of Innovation: 2 studies; Organizational Readiness for Change: 2 studies; Resource-Based View: 2 studies). As shown in Table 6, despite CB-SEM being the most appropriate method for confirmatory theory testing, it remains as underutilized as PLS-SEM each representing only 2 of 29 total articles (7%), while no study integrated complementary theoretical frameworks. This pattern confirms the systematic underutilization of confirmatory methods that this review identifies as the primary methodological gap in the domain.

Table 6. Analytical Methods and Theoretical Frameworks in Prior Research

Category	Method / Framework	Studies (n)	Appropriateness for Confirmatory Testing	Key Limitation
Analytical	PLS-SEM	2 studies	Low exploratory, predictive focus	Cannot confirm model fit; cannot discriminate between competing theories
Analytical	CB-SEM	2 studies	High designed for confirmatory theory testing	Requires large samples (n > 200); rarely used in agribusiness
Analytical	Systematic Review / SLR	2 studies	Moderate synthesizes rather than tests causal claims	Cannot establish causal relationships from primary data
Analytical	Capability Assessment	1 study	Low descriptive and context specific	Non-generalizable; no structural model testing
Theoretical	Diffusion of Innovation Theory	2 studies	Moderate explains adoption patterns	Underspecified on organizational and human capability dimensions
Theoretical	Organizational Readiness for Change	2 studies	High directly applicable to implementation context	Applied in isolation from value creation theories
Theoretical	Resource-Based View (RBV)	2 studies	High explains value creation from technology capabilities	Less dynamic for rapidly evolving technology adoption contexts

The most significant methodological gap this review identifies is the dominance of CB-SEM in the limited organizational literature on this topic. While CB-SEM works well for exploratory purposes and early stage theory building, it is not designed to confirm whether a pre-specified structural model accurately fits empirical data, which is precisely the scientific task at hand when established theories such as Organizational Readiness for Change, TAM, and RBV are being applied to a new domain (Dash & Paul, 2021; Hair et al., 2025). This mismatch likely explains why organizational adoption research in this area has accumulated conceptual arguments without ever producing a statistically confirmed structural theory. CB-SEM is better suited to this confirmatory objective because it can assess overall model fit using indices including CMIN/df, RMSEA, CFI, TLI, and SRMR, compare competing model specifications, and generate confidence intervals for both direct and indirect effects. Applying CB-SEM with samples of at least n=300 would represent the single most impactful methodological step forward available to researchers in this field.

Taken together, the 29 articles paint a picture of a field that has made strong technical progress but remains organizationally and analytically underdeveloped, a gap with clear consequences for both researchers and practitioners. On the technical side (RQ1), the evidence is solid: when multiple independent research groups, using different architectures and evaluation settings, all converge on accuracy figures between 93.68% and 97.9%, the case for automated FFB grading being not just possible but consistently deliverable is well established. The organizational evidence, however, tells a different story. It is conceptually sound but empirically narrow and skewed toward European settings (Gabriel & Gandorfer, 2023; Finger et al., 2023), raising legitimate questions about whether these findings apply to the Indonesian institutional context, which differs in infrastructure maturity, workforce skill levels, and regulatory environment.

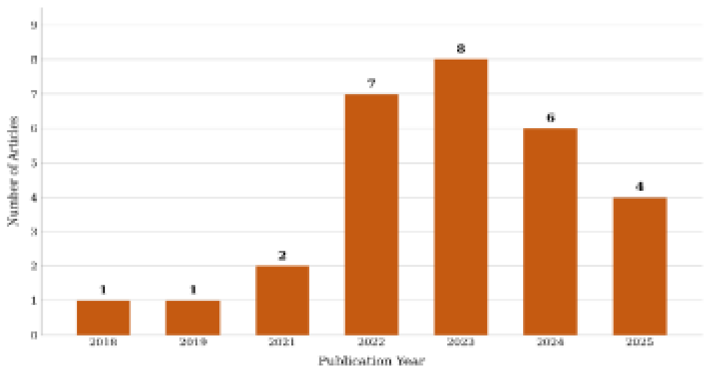
The organizational readiness findings (RQ2) warrant closer attention. Uren and Edwards (2023) argue that organizational readiness for AI adoption encompasses People, Processes, Technology, and Data dimensions, and that adoption success depends on maturity across all four rather than any single factor in isolation. Gabriel and Gandorfer's (2023) finding of 15-20% higher adoption rates in organizationally ready settings provides quantitative support for this, though the evidence is entirely European. The only Indonesia specific contribution comes from Siallagan and Ishak (2022), but their descriptive capability assessment cannot establish causal relationships or confirm theoretical mechanisms. The result is a meaningful empirical blind spot: how organizational readiness actually drives adoption outcomes in Indonesian palm oil mills is theoretically anticipated but never tested. This is exactly the kind of gap that CB-SEM based confirmatory research is built to close.

The operational and business value findings (RQ3) add a time dimension that the literature has not yet dealt with adequately. The chain from grading accuracy to FFA reduction (Tan et al., 2023) and from OER gains to CPO value is mechanistically sound and backed by engineering evidence. The problem is that financial benefits take 12-24 months to surface after implementation (Finger et al., 2023), which is consistent with Solow's (1987) IT productivity paradox, meaning that neither the technical benchmarking studies nor the limited organizational assessments currently available can capture the true business value of AI grading adoption. For researchers, this creates a direct methodological consequence: cross-sectional designs will consistently understate true investment returns. Longitudinal panel studies covering at least 24 months post implementation are not simply a better option but a necessary requirement for accurate ROI assessment.

This SLR contributes on both theoretical and practical fronts. Theoretically, it delivers the first PRISMA 2020 compliant synthesis across all four dimensions of the readiness to value pathway, giving researchers a clear map of where the field stands: strong on technical performance evidence, conceptually grounded but empirically thin on organizational readiness, mechanistically clear but temporally unresolved on business value, and analytically misaligned due to PLS-SEM dominance. Practically, it gives mill management the first evidence-based investment sequence: Technology Readiness must come first to prevent capability gaps; Organizational Readiness interventions, especially building management support and an innovation positive culture, offer the highest leverage before deployment; Human Capability development should continue after implementation as an effectiveness moderator; and financial expectations should be set against a 12-24 month horizon, with OER and FFA treated as the primary leading indicators. This guidance, built on synthesized empirical evidence rather than vendor claims or isolated case studies, addresses a practical information gap that mill management has long faced without a consolidated evidence base to draw on.

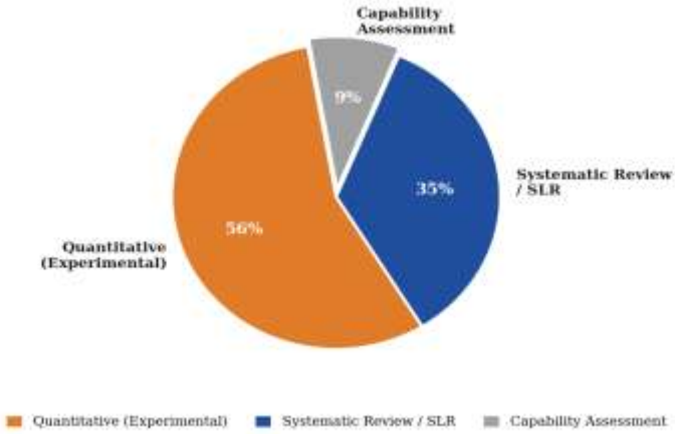
Visual Analysis of Synthesized Evidence

To facilitate clearer interpretation of synthesized evidence, key quantitative findings are presented through four visualizations. These figures complement the narrative synthesis by providing at a glance comparisons of the evidence base structure across all four research questions.



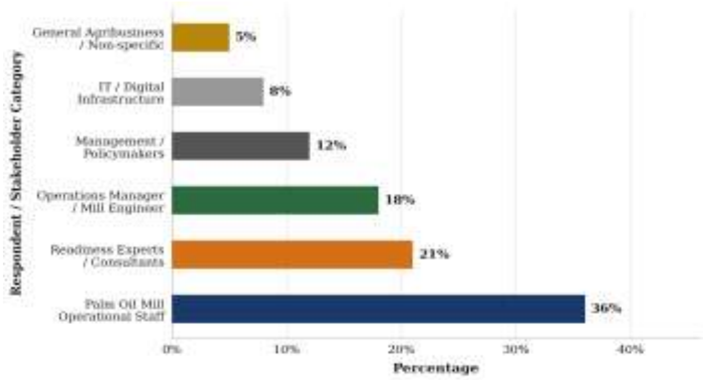
Picture 2. Bar diagram of distribution of included articles by publication year (2018-2025)

Picture 2 presents the temporal distribution of 29 included articles across 2018-2025. Publication activity peaks in 2022 and 2023 (7 articles each), coinciding with rapid commercialization of YOLOv7, YOLOv8, and advanced CNN architectures (Al Riza et al., 2025; Josdaan et al., 2024). The 2018-2021 base (10 articles) represents the technical benchmarking phase, while the 2022-2023 peak marks the transition toward more comprehensive implementation and organizational readiness studies.



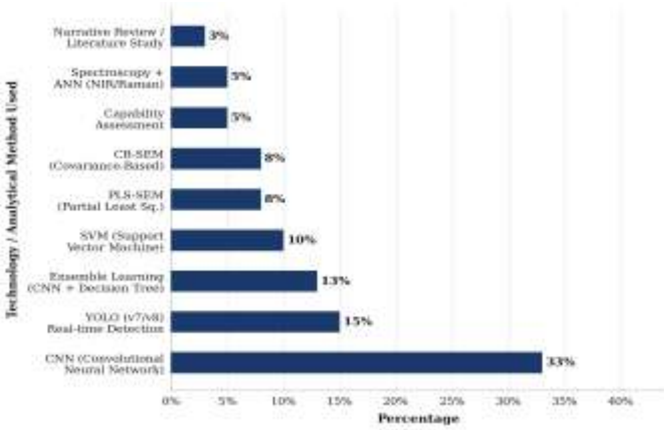
Picture 3. Circle diagram of distribution of types of research methods in included articles

Picture 3 illustrates the distribution of research methods across the 29 included articles. Quantitative experimental studies dominate at 56% (n=16), encompassing CNN, YOLO, and SVM benchmarking approaches. Systematic Review and SLR based studies account for 35% (n=10), reflecting growing recognition of the need for synthesized evidence in this domain. Capability assessment methods represent only 9% (n=3), confirming the scarcity of multi-firm organizational readiness investigations Notably, CB-SEM confirmatory designs are not captured in this research design classification as they represent an analytical method rather than a research design type; their underutilization is addressed separately in RQ4.



Picture 4. Bar diagram of distribution of respondent and stakeholder categories in included articles

Picture 4 shows palm oil mill operational staff as the dominant respondent category (36%), reflecting the operational focus of technical benchmarking studies. Readiness experts and consultants (21%) and operations managers and mill engineers (18%) represent the readiness and implementation perspectives respectively. Management and policymakers account for only 12%, indicating underdevelopment of strategic level business value evidence, a critical gap for investment decision making at holding company level Hariyanti et al., 2024. IT and digital infrastructure specialists (8%) and general agribusiness contexts (5%) represent the smallest groups, reinforcing the geographic and contextual specificity gaps identified in this review.



Picture 5. Bar diagram of distribution of technology and analytical methods used in included articles

Picture 5 presents the distribution of technology and analytical methods used across included articles. CNN dominates at 33%, confirming its role as the foundational deep learning architecture for FFB visual classification. YOLO-based real-time detection accounts for 15%, driven by industrial demand for conveyor deployable solutions. Ensemble learning methods represent 13%. PLS-SEM and CB-SEM each account for 8%; while CB-SEM is present, it remains far below appropriate use frequency given the confirmatory theoretical objectives of this field (Dash & Paul, 2021; Hair et al., 2025). Capability assessment (5%), spectroscopic methods (5%), and narrative review (3%) represent specialized niche approaches.

Limitations

Six limitations of this review deserve honest acknowledgment. First, conference proceedings, theses, and grey literature were excluded, which is standard practice for quality control but may leave out practitioner insights on palm oil mill digitalization that have not yet reached peer-reviewed journals. Second, the search was limited to English and Indonesian, potentially missing relevant Malay language studies given Malaysia's major role as a global palm oil producer; future reviews should expand to include Malay language sources. Third, because AI architectures evolve rapidly, some accuracy benchmarks reported here may become outdated within 12-18 months, which could understate what is currently achievable. Fourth, narrative synthesis was used rather than meta-analysis, meaning precise effect size estimates are not available; this may improve over time as the organizational literature grows and standardizes its outcome measures. Fifth, the final article count of 29, while acceptable for a peer-reviewed SLR, reflects the still limited body of organizational and business focused research in this domain, and more Indonesia specific studies are needed to strengthen geographic coverage. Sixth, even with two independent reviewers and strong interrater agreement (k=0.82), some degree of subjectivity in applying inclusion and quality criteria is unavoidable in any systematic review process.

CONCLUSION

This systematic literature review synthesized 29 high-quality peer-reviewed articles from three academic databases (inter-rater reliability: Cohen’s Kappa k=0.82, almost perfect agreement; Landis & Koch, 1977), delivering, to the authors’ knowledge, the first PRISMA

2020 compliant synthesis addressing the complete readiness-to-value pathway — technical performance, organizational readiness, operational outcomes, and analytical frameworks — for AI, automation, and digitalization in FFB grading for the palm oil industry. Across four research questions, the synthesis produced consistent findings while also uncovering a clear set of research gaps that should guide future empirical work.

On the technical side (RQ1), CNN and YOLO architectures consistently achieve FFB grading accuracy of 93.68-97.9%, well above manual methods, with multiple independent research groups confirming technical feasibility. Turning to implementation determinants (RQ2), Technology Readiness, Organizational Readiness, and Human Capability are the three key implementation success factors. Organizational Readiness, especially management support and innovation-positive culture, is associated with projected adoption rate growth of 15–20% over five years, while Technology Readiness is the infrastructure prerequisite and Human Capability moderates how effectively technology is used post-deployment. At the operational and business level (RQ3), AI-based grading produces measurable OER gains (minimum 21% for ripe FFB) and FFA reductions (below 5%), but financial returns require extended time to materialize after implementation, consistent with the IT productivity paradox (Solow, 1987; Brynjolfsson, 1993), making OER and FFA more appropriate investment KPIs than short-term financial metrics. Finally, the methodological landscape (RQ4) reveals that PLS-SEM dominates the literature despite CB-SEM being better suited for confirmatory theory testing, and no study has yet combined Organizational Readiness, TAM, and RBV into a unified structural model for the FFB grading context.

Three theoretical contributions stand out. First, this review provides, to the authors' knowledge, the first integrated synthesis of the readiness-to-value pathway in FFB grading, giving researchers a clear picture of where evidence is strong and where it falls short. Second, five research gaps are formally identified: theoretical, methodological, contextual, practical, and geographic, forming a structured agenda for future investigation. Third, the combination of Uren and Edwards' (2023) Organizational Journey towards AI Adoption framework, Davis' (1989) TAM, and Barney's (1991) RBV is shown to provide a theoretically complete and complementary foundation for an integrative structural model of FFB grading technology adoption.

Four research directions are recommended. First, a CB-SEM study with n>=350 should build and test an integrative model linking readiness constructs (TR, OR, HC), implementation constructs (AI-based Grading, Automation Infrastructure, Digital Data Platform), System Performance, and outcomes (Operational Impact, Business Value) in the Indonesian palm oil setting. Second, longitudinal panel designs that track baseline readiness, implementation progress, and outcomes at 18-24 months post implementation are needed to properly capture the performance to value timeline. Third, multi group SEM comparing mills by scale, ownership structure, and geographic region will reveal boundary conditions for the structural relationships found here. Fourth, mixed methods studies pairing SEM with case studies at mills in different adoption stages will help interpret quantitative results and shed light on the organizational learning processes that connect performance gains to business value.

REFERENCES

Abbasi, R., Martinez, P., & Ahmad, R. (2022). The digitization of agricultural industry: A systematic literature review on agriculture 4.0. Smart Agricultural Technology, 2, 100042. https://doi.org/10.1016/j.atech.2021.100042

Al Riza, D. F., Yusuf, A., & Kusuma, W. A. (2025). Comparative study of citrus fruits detection and counting with single and double labels based on convolutional neural network using YOLOv7. Smart Agricultural Technology, 10, 100589. https://doi.org/10.1016/j.atech.2024.100589

Alfatni, M. S. M., Khairunniza-Bejo, S., Marhaban, M. H. B., Ben Saaed, O. M., Mustapha, A., & Shariff, A. R. M. (2022). Towards a real-time oil palm fruit maturity system using supervised classifiers based on feature analysis. Agriculture, 12(9), 1461. https://doi.org/10.3390/agriculture12091461

Anizar, A., Matondang, A. R., Ismail, R., & Matondang, N. (2022). The role of technical support and effective communication to successful intervention program on palm oil mills. International Journal on Advanced Science, Engineering and Information Technology, 12(3), 987-993. https://doi.org/10.18517/ijaseit.12.3.15268

Barney, J. (1991). Firm resources and sustained competitive advantage. Journal of Management, 17(1), 99-120. https://doi.org/10.1177/014920639101700108

Bhat, I. A., Mir, S. A., & Shah, M. A. (2025). Leveraging artificial intelligence in agribusiness: Strategic management practices and future prospects. Agricultural Systems, 224, 104213. https://doi.org/10.1007/s43621-025-01260-3

Brynjolfsson, E. (1993). The productivity paradox of information technology. Communications of the ACM, 36(12), 66–77. https://doi.org/10.1145/163298.163309

Chuquimarca, L. E., Vintimilla, B. X., & Velastin, S. A. (2024). A review of external quality inspection for fruit grading using CNN models. Artificial Intelligence in Agriculture, 14, 1–20. https://doi.org/10.1016/j.aiia.2024.10.002

Dash, G., & Paul, J. (2021). CB-SEM vs PLS-SEM methods for research in social sciences and technology forecasting. Technological Forecasting and Social Change, 173, 121092. https://doi.org/10.1016/j.techfore.2021.121092

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319-340. https://doi.org/10.2307/249008

Finger, R., Slijper, T., Ronner, E., Poortvliet, P. M., & de Mey, Y. (2023). Digital innovations for sustainable and resilient agricultural systems. European Review of Agricultural Economics, 50(4), 1238-1270. https://doi.org/10.1093/erae/jbad011

Gabriel, A., & Gandorfer, M. (2023). Adoption of digital technologies in agriculture: An inventory in a European small-scale farming region. Precision Agriculture, 24(1), 68-91. https://doi.org/10.1007/s11119-022-09931-1

GAPKI. (2023). Laporan kinerja industri kelapa sawit nasional 2023 [National palm oil industry performance report 2023]. Gabungan Pengusaha Kelapa Sawit Indonesia.

Gudergan, S. P., Sarstedt, M., Hair, J. F., & Ringle, C. M. (2025). Advanced issues in covariance-based structural equation modeling. Journal of Business Research, 178, 115029. https://doi.org/10.1016/j.jbusres.2024.115029

Hair, J. F., Sarstedt, M., & Ringle, C. M. (2025). Covariance-based structural equation modeling (CB-SEM): A SmartPLS 4 software tutorial. Journal of Marketing Analytics. https://doi.org/10.1057/s41270-025-00414-6

Hariyanti, F., Syahza, A., Zulkarnain, & Nofrizal. (2024). Economic transformation based on leading commodities through sustainable development of the oil palm industry. Heliyon, 10(4), e25674. https://doi.org/10.1016/j.heliyon.2024.e25674

Hong Tzuan, L., Ismail, W. I. W., Razak, N. A., & Wahab, M. N. A. (2022). Classification of oil palm fresh fruit bunch ripeness using Raman spectroscopy and artificial neural network. Computers and Electronics in Agriculture, 201, 107299. https://doi.org/10.1016/j.compag.2022.107299

Hong, Q. N., Fabregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Education for Information, 34(4), 285-291. https://doi.org/10.3233/EFI-180221

Josdaan, J., Tamsila, V. C., Harefa, J., & Jingga, K. (2024). Revolutionizing palm oil ripeness classification: Utilizing YOLOv8 for ultra-precise ripeness detection. *Procedia Computer Science, 245*, 700–709. https://doi.org/10.1016/j.procs.2024.10.296

Kitchenham, B., & Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering* (EBSE Technical Report EBSE-2007-01). Keele University. https://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf

Lai, J. W., Ramli, H. R., Ismail, L. I., & Hasan, W. Z. W. (2023). Oil palm fresh fruit bunch ripeness detection methods: A systematic review. Agriculture, 13(1), 156. https://doi.org/10.3390/agriculture13010156

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159-174. https://doi.org/10.2307/2529310

Mansour, M. Y. M. A., Dambul, K. D., & Choo, K. Y. (2022). Object detection algorithms for ripeness classification of oil palm fresh fruit bunch. International Journal of Technology, 13(6), 1326-1335. https://doi.org/10.14716/ijtech.v13i6.5932

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Patkar, G. S., Anjaneyulu, G. S. G. N., & Mouli, P. V. S. S. R. (2018). Challenging issues in automated oil palm fruit grading. IAES International Journal of Artificial Intelligence, 7(3), 111–118. https://doi.org/10.11591/ijai.v7.i3.pp111-118

Ringle, C. M., Sarstedt, M., Sinkovics, N., & Sinkovics, R. R. (2023). A perspective on using partial least squares structural equation modelling in data articles. Data in Brief, 48, 109074. https://doi.org/10.1016/j.dib.2023.109074

Rosbi, M., Omar, Z., Khairuddin, U., Majeed, A. P. P. A., & Bakar, S. A. R. S. A. (2024). Machine learning for automated oil palm fruit grading : The role of fuzzy C-means segmentation and textural features. *Smart Agricultural Technology, 9*, 100691. https://doi.org/10.1016/j.atech.2024.100691

Schuberth, F., Hubona, G., Roemer, E., Zaza, S., Schamberger, T., Chuah, F., Cepeda-Carrión, G., & Henseler, J. (2023). The choice of structural equation modeling technique matters: A commentary on Dash and Paul (2021). Technological Forecasting and Social Change, 194, 122665. https://doi.org/10.1016/j.techfore.2023.122665

Siallagan, S., & Ishak, A. (2022). A technological capability assessment of company in the crude palm oil industry in Indonesia. IOP Conference Series: Earth and Environmental Science, 1025, 012012. https://doi.org/10.1088/1755-1315/1025/1/012012

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. Journal of Business Research, 104, 333-339. https://doi.org/10.1016/j.jbusres.2019.07.039

Solow, R. M. (1987, July 12). We'd better watch out. New York Times Book Review, 36.

Suharjito, Junior, F. A., Koeswandy, Y. P., Debi, Nurhayati, P. W., Asrol, M., & Marimin. (2023). Annotated datasets of oil palm fruit bunch piles for ripeness grading using deep learning. Scientific Data, 10(1), 72. https://doi.org/10.1038/s41597-023-01958-x

Tan, C. X., Lim, W. S., & Choo, Y. M. (2023). Free fatty acid formation in oil palm fresh fruit bunches during the post-harvest period. Agriculture, 13(4), 812. https://doi.org/10.3390/agriculture13040812

Uren, V., & Edwards, J. S. (2023). Technology readiness and the organizational journey towards AI adoption: An empirical study. International Journal of Information Management, 68, 102588. https://doi.org/10.1016/j.ijinfomgt.2022.102588

Xu, J., Li, Y., Zhang, M., & Zhang, S. (2024). Sustainable agriculture in the digital era: Past, present, and future trends by bibliometric analysis. Heliyon, 10, e34612. https://doi.org/10.1016/j.heliyon.2024.e34612

Zaki, M. A. M., Ayob, M. Z., Razak, N. A., Yahya, A., & Harun, M. H. (2024). Impact of Industry 4.0 technologies on the oil palm industry: A literature review. Smart Agricultural Technology, 8, 100431. https://doi.org/10.1016/j.atech.2024.100685

Zolfagharnassab, S., Shariff, A. R. B. M., Ehsani, R., Jaafar, H. Z., & Aris, I. B. (2022). Classification of oil palm fresh fruit bunches based on their maturity using thermal imaging technique. Agriculture, 12(11), 1779. https://doi.org/10.3390/agriculture12111779