ISSN 2723-6471

A Diagnostic Framework for Staged AI Adoption in Batik Motif Recognition: Integrating CNN Evidence and Implementation Readiness

Irwan Sembiring[1,*], Paminto Agung Christianto[2,*], Eko Budi Susanto[3], Suharyadi[4], Cheryl Louisa Loedwyca[5]

[1,3,4,5]*Faculty of Information Technology, Satya Wacana Christian University, Dr. O. Notohamidjojo St. No. 1–10, Salatiga City 50715, Indonesia*

[2,3]*Faculty of Information Technology, Institut Widya Pratama, Patriot Street No. 25, Pekalongan City 51146, Indonesia*

(Received: February 1, 2026; Revised: March 25, 2026; Accepted: June 10, 2026; Available online: July 12, 2026)

Abstract

This study proposes a diagnostic dual-layer decision-support framework for staged artificial intelligence adoption in batik motif recognition. The objective is to examine whether technical evidence from convolutional neural network classification and perceived implementation-side readiness can be jointly interpreted to prevent premature deployment in cultural-heritage recognition. The contribution of the study is not a new classifier architecture, but an operational diagnostic logic that treats model performance, class-level instability, readiness perception, governance, security, and feedback mechanisms as complementary but non-substitutable evidence. Methodologically, the technical layer evaluated three transfer-learning baselines, namely VGGNet-16, ResNet50, and MobileNetV2, using 983 batik images across 20 motif classes. The implementation layer assessed perceived readiness among 173 information technology practitioners using the Technology-Organization-Environment-Human plus Feedback dimensions. The integration layer then mapped technical-readiness evidence and readiness perception into explicit staged-adoption decisions rather than averaging them as interchangeable indicators. The analysis used performance summaries, readiness profiles, decision matrices, security checklists, learning curves, and confusion-matrix diagnostics to connect empirical observations with staged adoption recommendations. The best-performing baseline was ResNet50, with 45% accuracy and a macro F1-score of 0.40, showing low technical readiness and substantial motif-specific instability. In contrast, the readiness survey indicated high perceived implementation-side readiness, with an average agreement score of 78.7%. This mismatch reveals a readiness asymmetry: implementation support may exist even when the recognition model remains technically immature. The findings imply that batik-recognition systems should prioritize dataset expansion, expert label validation, model refinement, moderated feedback, security governance, and controlled pilot testing before operational deployment. The framework provides a transparent basis for risk-aware, staged adoption decisions in artificial-intelligence-assisted cultural heritage preservation.

Keywords: Artificial Intelligence Adoption, Batik Recognition, Organizational Readiness, Technical Readiness, Decision-Support Framework, Cultural Heritage

1. Introduction

Batik is an Indonesian living cultural heritage whose motifs encode regional identity, symbolic meaning, and historical narratives. Artificial intelligence (AI)-based image recognition, particularly convolutional neural networks (CNNs), can support motif documentation, retrieval, and cultural knowledge management. However, prior recognition studies commonly evaluate success through accuracy, precision, recall, and F1-score alone [1], [2], [3], [4], [5], [6]. These metrics are necessary but insufficient for adoption decisions because a technically promising model may fail when governance, infrastructure, security, and human capability are weak. Conversely, organizations may be enthusiastic about AI even when a model is not technically reliable.

This study addresses this socio-technical gap by proposing a diagnostic dual-layer framework. The technical layer evaluates convolutional-neural-network-based batik motif classification as evidence of model and dataset readiness, whereas the implementation layer evaluates perceived implementation-side readiness using the Technology-Organization-Environment-Human plus Feedback (TOEH+Feedback) framework. Recent artificial-intelligence-readiness studies emphasize that adoption capacity depends on technology readiness, organizational capability, human

***Corresponding author: Irwan Sembiring (irwan@uksw.edu), Paminto Agung Christianto (p_a_chr@stmik-wp.ac.id)**
DOI: https://doi.org/10.47738/jads.v7i3.1427

competence, and contextual conditions rather than model performance alone [7], [8], [9]. Accordingly, the novelty of this work lies in the operational integration of technical evidence and readiness perception to support staged adoption decisions, not in proposing a new convolutional neural network architecture.

This positioning is important because prior studies generally treat artificial intelligence development either as a technical classification problem or as an organizational adoption problem. In contrast, the proposed framework interprets classification outputs, readiness indicators, security-governance requirements, and feedback mechanisms as interdependent inputs for adoption decision-making. The framework follows a non-substitutability principle: high predictive performance does not automatically justify deployment without governance and human capability, and high implementation enthusiasm cannot compensate for an unreliable model. In the batik-heritage domain, this distinction is critical because incorrect motif recognition can distort cultural interpretation, reduce user trust, and weaken the credibility of artificial-intelligence-assisted preservation initiatives.

The study is guided by three research questions (RQs): RQ1, how do baseline CNN architectures perform in classifying batik motifs under limited dataset conditions? RQ2, what is the level of perceived implementation-side readiness among information-technology practitioners for supporting AI-based batik recognition? RQ3, how can technical and organizational evidence be integrated into a practical adoption decision framework? The objective is not to claim deployment-ready recognition performance, but to show how technical limitations and perceived readiness can be jointly interpreted to prevent premature deployment.

2. Literature Review

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

AI-based image recognition has become central to digital cultural heritage because it converts visual archives into searchable and interpretable knowledge assets. Prior work in cultural heritage and traditional textile recognition shows that deep learning, transfer learning, feature fusion, and augmentation can improve classification performance, but results depend strongly on dataset scale, curation, labeling quality, and inter-class separability [13], [14], [15], [16], [17]. Batik recognition is especially difficult because motifs may share geometric structures, repetitive ornaments, color composition, and contextual meaning.

Technical readiness in this study is broader than model accuracy. CNN architectures such as VGGNet, ResNet, and MobileNet are useful baselines because they represent different trade-offs between feature depth, residual learning, and computational efficiency [3],[17]. Under limited data, low macro F1-score or unstable class-level performance should be interpreted as evidence of dataset and deployment risk rather than only algorithmic failure [18],[19]. AI adoption is also an organizational phenomenon. Readiness studies emphasize technological fit, infrastructure, management support, policy pressure, human competence, digital capability, and governance capability [7], [8], [9], [20], [21], [22], [23], [24]. The TOEH model captures technical, organizational, environmental, and human dimensions; this study extends it with Feedback because recognition systems require human-in-the-loop correction, model monitoring, and controlled learning cycles [30], [31], [32]. Security is treated as a governance requirement across data collection, training, inference, feedback, and model updating because vulnerabilities such as data poisoning, adversarial manipulation, privacy exposure, and model extraction may undermine trust [10], [11], [12], [25]-[29].

Security and governance further extend the meaning of readiness. Artificial-intelligence-based image-recognition services operate across a life cycle that includes data collection, data labeling, model training, inference, feedback, and model updating. Each stage introduces different risks, including unverified image provenance, label noise, data poisoning, adversarial manipulation, privacy exposure, model extraction, and uncontrolled model drift [10], [11], [12], [25], [26], [27], [28], [29]. In a cultural-heritage setting, these risks are not only technical; they may also affect cultural credibility because an unreliable or weakly governed system can produce misleading motif interpretations. Therefore, security is interpreted as a readiness requirement rather than a post-deployment add-on.

ISSN 2723-6471

Feedback is equally important because AI recognition systems are rarely static. Human-in-the-loop mechanisms allow users, curators, or domain experts to correct uncertain predictions, validate labels, and improve the system over time [30], [31], [32]. However, feedback can also introduce bias or malicious corrections when it is not moderated. For this reason, the Feedback dimension in TOEH+Feedback is not treated merely as a satisfaction channel, but as a controlled learning mechanism that supports auditability, model monitoring, and responsible continuous improvement. The research gap is that technical recognition studies rarely incorporate organizational readiness, while adoption studies often lack domain-specific technical evidence. This study bridges the gap by treating CNN performance and TOEH+Feedback readiness as complementary but non-substitutable evidence for AI adoption decisions in cultural heritage contexts.

3. Methodology

This study employed a diagnostic mixed-method design consisting of a technical layer, an implementation-readiness layer, and an integration layer, as shown in figure 1. The technical layer addressed RQ1 through convolutional-neural-network-based batik image classification. The readiness layer addressed RQ2 through a perception survey among information-technology practitioners. The integration layer addressed RQ3 by converting both forms of evidence into staged adoption decisions. Operationally, integration was conducted in three steps: first, model-level and class-level metrics were interpreted as technical-readiness evidence; second, survey agreement scores were interpreted as perceived implementation-side readiness; and third, both layers were mapped into a decision matrix in which technical and implementation readiness were treated as non-substitutable conditions.

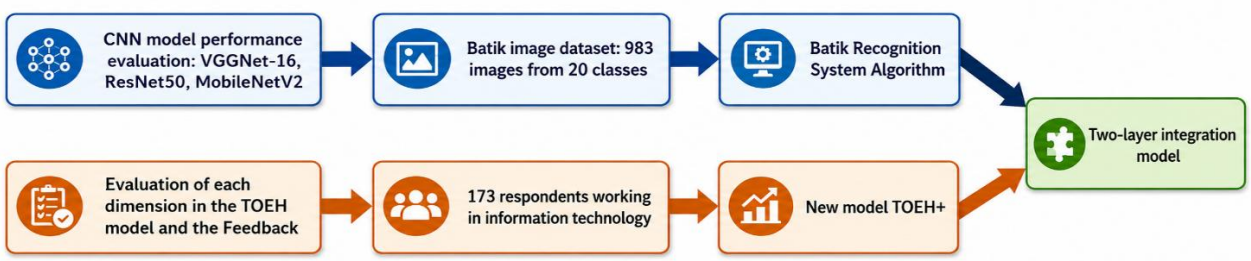

Figure 1. Research design of the dual-layer framework

3.1. Technical Layer: CNN-Based Batik Image Classification

The dataset consisted of 983 batik images distributed across 20 motif classes. Each class contained approximately 45–50 images, making the dataset relatively balanced but still limited for deep learning. Images were collected from publicly accessible sources, curated by removing duplicates and low-quality samples, resized to 224 × 224 pixels, normalized, and split into 80% training and 20% testing data. Data augmentation included rotation, width and height shifts, shear, zoom, and horizontal flipping [17]. A summary of the technical dataset and model configuration is shown in table 1.

Table 1. Summary of technical dataset and model configuration

Aspect	Configuration
Dataset	983 batik images across 20 motif classes
Class distribution	Approximately 45–50 images per class
Preprocessing	Resize to 224 x 224 pixels and normalization
Split	80% training and 20% testing
Augmentation	Rotation, shifts, shear, zoom, and horizontal flip
Models	VGGNet-16, ResNet50, and MobileNetV2
Training setup	Adam optimizer, learning rate 0.001, categorical cross-entropy, 50 epochs, batch size 32

Evaluation metrics	Accuracy, macro precision, macro recall, and macro F1-score

VGGNet-16, ResNet50, and MobileNetV2 were selected as representative baselines rather than state-of-the-art competitors. VGGNet-16 represents a simple sequential architecture, ResNet50 represents deeper residual learning, and MobileNetV2 represents lightweight deployment-oriented design. Their outputs were interpreted as technical-readiness signals, not as evidence of deployment-ready recognition.

For reproducibility, the transfer-learning configuration was explicitly treated as a frozen-feature baseline. Each architecture used ImageNet-pretrained weights; the convolutional backbone was kept frozen during the baseline comparison, while a task-specific classification head was trained for the 20 batik motif classes using categorical cross-entropy. Full unfreezing and staged fine-tuning were not performed in this exploratory experiment because the dataset contained only approximately 45-50 images per class. Consequently, the results should be interpreted as baseline diagnostic evidence rather than as the maximum attainable performance of each architecture.

3.2. Implementation-Readiness Layer: TOEH+Feedback Readiness Assessment

The implementation-readiness layer used the TOEH+Feedback model. The questionnaire contained 14 Likert-scale items and was completed by 173 respondents working in information-technology (IT) roles, including developers, consultants, and data-processing staff (Complete data is shown in table 2). Therefore, the survey represents perceived implementation-side readiness among potential technical implementers rather than direct institutional readiness among batik micro, small, and medium enterprises (MSMEs), museums, curators, artisans, cultural experts, or heritage organizations. This scope limitation was made explicit to avoid overgeneralizing the readiness findings.

Table 2. TOEH+Feedback measurement structure

Dimension	Items	Main focus
Technology	5	Perceived benefit, compatibility, and system security
Organization	3	Management support and infrastructure
Environment	2	Policy support and competitive pressure
Human	2	Knowledge and competence
Feedback	2	Learning and system improvement mechanisms

Reliability was evaluated using Cronbach's Alpha, whereas item validity was assessed using significance testing. Because the study is exploratory and diagnostic, these tests were used for descriptive readiness profiling rather than causal inference. Readiness was summarized as the proportion of respondents selecting agree or strongly agree.

For operational transparency, the questionnaire was interpreted as a descriptive profiling instrument rather than a causal adoption model. The Technology dimension covered perceived benefit, compatibility, technical suitability, and system security. The Organization dimension captured management support, infrastructure availability, and implementation support. The Environment dimension reflected policy support and external pressure, while the Human dimension focused on knowledge and competence. The Feedback dimension assessed the availability of learning and system-improvement mechanisms. Because several dimensions were represented by only two or three items, the results were used to support exploratory readiness diagnosis rather than definitive construct validation.

3.3. Integration and Readiness Scoring

To reduce subjectivity, technical and implementation-side readiness were operationalized through explicit diagnostic thresholds before interpretation. Technical readiness was categorized as low when accuracy or macro F1-score was below 0.60, moderate when both indicators ranged from 0.60 to 0.80, and high when both exceeded 0.80 with stable class-level performance. Implementation-side readiness was categorized as low when the mean agreement score was below 60%, moderate when it ranged from 60% to 75%, and high when it exceeded 75% (Complete data is shown in table 3). These cutoffs are exploratory diagnostic boundaries derived for transparent decision staging, not universal deployment standards or validated psychometric thresholds. Therefore, any low technical-readiness result overrides deployment enthusiasm and leads to a non-deployment recommendation.

ISSN 2723-6471

Table 3. Operational readiness scoring criteria

Layer	Indicator	Low	Moderate	High
Technical	Accuracy	< 0.60	0.60–0.80	> 0.80
Technical	Macro F1-score	< 0.60	0.60–0.80	> 0.80
Technical	Class-level stability	Many classes below 0.30 F1	Mixed stability	Stable across most classes
Implementation-side	Mean agreement score	< 60%	60%–75%	> 75%

Security was incorporated as a readiness consideration, not as an empirical adversarial-robustness, poisoning-resilience, privacy, or penetration test. Minimum governance controls include dataset provenance, expert label validation, version control, confidence monitoring, moderated feedback, audit logging, and controlled model updating. Integration was performed by mapping technical and perceived implementation-side readiness into a decision matrix. A descriptive readiness-gap proxy was calculated only to make the asymmetry explicit; it was not used as a compensatory score that could convert a technically immature model into a deployable system.

4. Results and Discussion

Figure 2 presents the integrated adoption-readiness framework used to connect the technical classification experiment with the implementation-readiness assessment. The figure should be read as a staged diagnostic workflow rather than as a conventional model-training pipeline. Its purpose is to show how evidence from batik motif classification, readiness perception, governance preparation, and feedback control is converted into an adoption recommendation.

The technical side of the framework begins with image preprocessing, transfer-learning baseline evaluation, learning-curve inspection, and confusion-matrix diagnostics. These outputs are not interpreted only as predictive-performance indicators; they are converted into technical-readiness evidence. Low accuracy, low macro F1-score, and unstable class-level performance indicate that the recognition system remains at the exploratory-prototype stage and requires dataset expansion, expert label validation, and model refinement.

The implementation-readiness side captures perceived readiness among IT practitioners through the TOEH+Feedback dimensions. The integration layer then applies a non-compensatory decision rule: strong perceived implementation support may justify preparation and controlled piloting, but it cannot override low technical reliability. Therefore, figure 2 operationalizes the central argument of this study by separating technical maturity, implementation capacity, and governance controls before any staged AI adoption decision is made.

ISSN 2723-6471

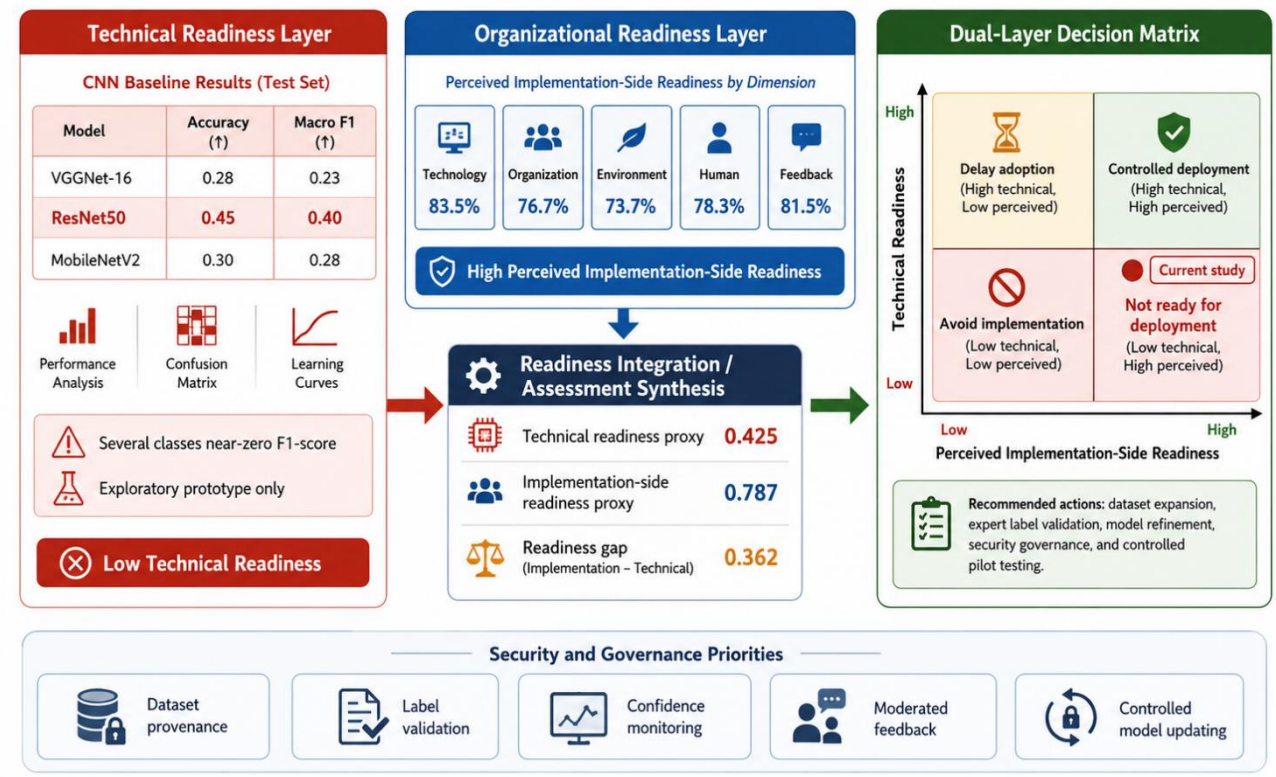

Figure 2. Integrated Framework for AI Adoption Readiness in Batik Heritage Recognition

4.1. Technical Readiness Results

The technical layer showed low readiness. ResNet50 performed best, but its accuracy of 0.45 and macro F1-score of 0.40 remain below the exploratory readiness threshold. VGGNet-16 and MobileNetV2 achieved lower aggregate performance. Class-level results showed several zero or near-zero F1-scores, indicating motif-specific instability and risk of unreliable cultural interpretation, complete data is shown in table 4.

A class-level inspection provides additional diagnostic evidence. Although some classes obtained relatively stronger F1-scores under ResNet50, such as Ciamis, Dicelup, Sogan, Bunga, and Sidomukti, several other classes showed zero or near-zero performance. Priangan, Pekalongan, Lasem, Gentongan, Garutan, and Megamendung were particularly problematic in the class-level analysis. This instability indicates that the model does not merely suffer from low aggregate accuracy; it also fails unevenly across motif categories whose visual semantics may overlap or whose training samples may be insufficiently representative. These low-performing classes should therefore become priority targets for expert label audit, additional image collection, and motif-group analysis before any deployment-oriented experiment is attempted.

Table 4. Overall CNN performance and diagnostic interpretation

Model/Aspect	Accuracy	Macro precision	Macro recall	Macro F1-score	Readiness implication
VGGNet-16	0.28	0.22	0.26	0.23	Low technical readiness
ResNet50	0.45	0.40	0.45	0.40	Best-performing baseline, but not deployment-ready
MobileNetV2	0.30	0.34	0.28	0.28	Low technical readiness
Class instability	Several classes near zero F1-score	-	-	-	High motif-specific risk
Deployment status	Below threshold	-	-	-	Exploratory prototype only

The confusion matrix and learning curves indicate that the model does not fail uniformly (as shown in figure 3); rather, failure is concentrated in motif categories with overlapping visual semantics and limited sample diversity. The

ISSN 2723-6471

diagnostic value of the confusion matrix is therefore not to claim acceptable recognition accuracy, but to identify which motif groups require additional curation and expert validation. Because the current analysis did not compute a ranked pairwise confusion table, the pair-level interpretation remains limited; future experiments should report the most frequent misclassified motif pairs, normalized confusion rates, and visual examples of ambiguous motifs.

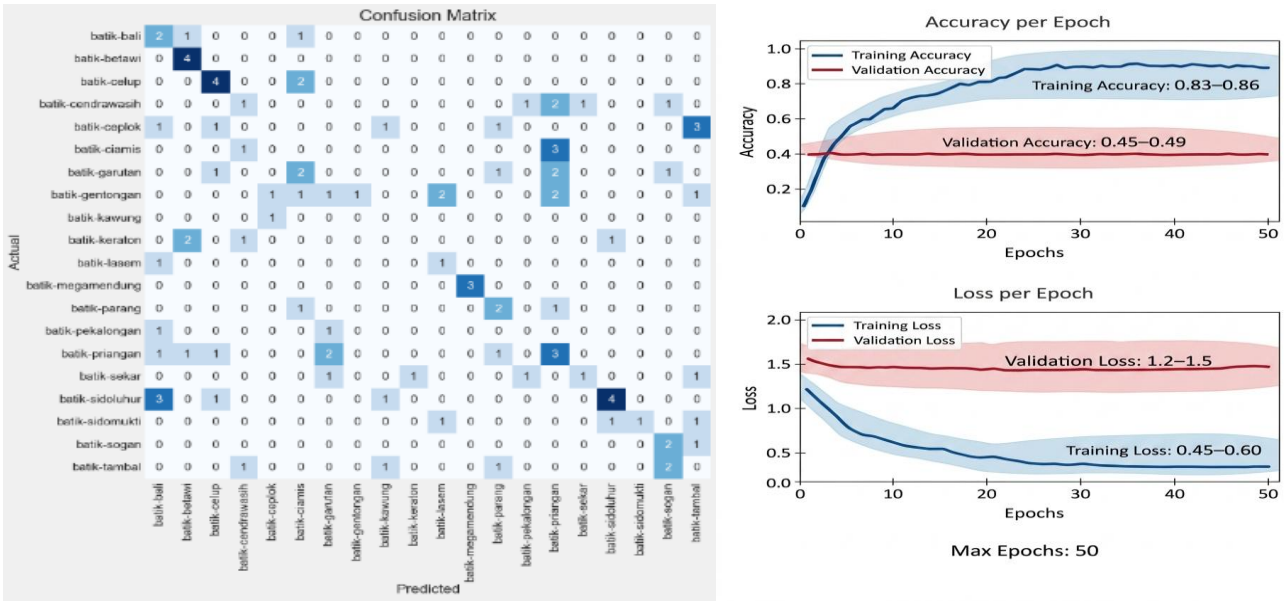

Figure 3. Confusion matrix and training-validation accuracy and loss curves of the ResNet50 model

The learning-curve pattern also supports this interpretation. Gradual convergence during training suggests that the model learned some discriminative features, but the gap and fluctuation between training and validation performance indicate limited generalization. This behavior is consistent with a small 20-class dataset where augmentation reduces, but does not eliminate, overfitting risk. Therefore, the observed results should be interpreted as evidence that dataset scale, label consistency, class separability, transfer-learning strategy, and validation design must be improved before stronger deployment claims can be made.

Compared with prior batik image-classification studies, the current work evaluates a more complex 20-class setting and therefore should not be compared with 3-6 class studies as a direct benchmark competition, as shown in table 5. The lower accuracy should not be treated as a successful recognition outcome or as evidence that class complexity alone explains the performance gap. Instead, the comparison contextualizes why dataset expansion, motif-label validation, fine-tuning, external validation, and more robust experimental design are required before deployment-oriented claims can be supported.

Table 5. Contextual comparison with prior batik image-classification studies

Study	Dataset size	Classes	Best accuracy	Main focus
Rasyidi and Bariyah [33]	994	6	94%	CNN/DenseNet batik pattern recognition
Meranggi et al. [34]	621	5	88.88% ± 0.88	CNN classification with data improvement
Rasyidi et al. [35]	120	3	87.61%	Batik-making method identification
Current study	983	20	45%	Deployment-readiness diagnostics

4.2. Perceived Implementation-Side Readiness

The Technology-Organization-Environment-Human plus Feedback results indicate high perceived implementation-side readiness among information-technology practitioners, with mean agreement scores above 75% in most dimensions. Technology and Feedback showed the strongest readiness, whereas Environment was lower but still moderate-to-high. These results suggest implementation support capacity among technical actors, but they should not

be generalized as direct readiness of batik micro, small, and medium enterprises, museums, curators, artisans, cultural experts, or cultural heritage organizations.

The reliability results also require cautious interpretation. Technology showed strong internal consistency, while Organization was acceptable. However, Environment, Human, and Feedback produced marginal Cronbach's Alpha values between 0.630 and 0.691, as shown in table 6. These values may be acceptable for an exploratory diagnostic study, but they indicate that the constructs should be refined before the instrument is used for confirmatory measurement or policy-level readiness assessment. Future work should expand the number of items per dimension, involve domain stakeholders, and test convergent and discriminant validity.

Table 6. TOEH+Feedback readiness results

Dimension	Items	Reliability/validity evidence	Agreement summary	Interpretation
Technology	5	Alpha = 0.910; $p < 0.05$	82.6%–85.0%; mean 83.5%	High perceived readiness
Organization	3	Alpha = 0.750; $p < 0.05$	67.7%–82.1%; mean 76.7%	High; infrastructure-related item was weakest
Environment	2	Alpha = 0.646; $p < 0.05$	66.5%–80.9%; mean 73.7%	Moderate-to-high; marginal reliability
Human	2	Alpha = 0.630; $p < 0.05$	76.3%–80.3%; mean 78.3%	High; marginal reliability
Feedback	2	Alpha = 0.691; $p < 0.05$	78.0%–85.0%; mean 81.5%	High; marginal reliability

4.3. Integrated Readiness Assessment

The integrated assessment places the study in the low technical readiness–high perceived implementation-side readiness quadrant. The technical readiness proxy, calculated as the average of ResNet50 accuracy and macro F1-score, was 0.425. The organizational readiness proxy, calculated as the average TOEH+Feedback agreement score, was 0.787. The resulting readiness gap of 0.362 indicates a substantial asymmetry: implementation-side enthusiasm and support exceed current model maturity; complete data is shown in table 7.

To make this synthesis transparent, the technical readiness proxy was computed on a normalized 0-1 scale as (Accuracy_ResNet50 + MacroF1_ResNet50) / 2 = (0.45 + 0.40) / 2 = 0.425. The implementation-side readiness proxy was computed as the average mean agreement across the Technology, Organization, Environment, Human, and Feedback dimensions. The readiness gap was then calculated as Implementation-side readiness proxy - Technical readiness proxy = 0.787 - 0.425 = 0.362. This proxy is descriptive and auditable, but it is not a validated multi-criteria decision-analysis model. More importantly, it is non-compensatory: a high implementation-side score cannot offset low technical reliability.

The integrated interpretation follows a non-substitutability principle. High perceived implementation-side readiness may justify continued investment, infrastructure preparation, user training, and pilot planning, but it cannot justify operational deployment when the recognition model remains unstable, as shown in figure 2 and table 8. Conversely, a stronger model would still require governance, human capability, security monitoring (details are shown in table 9), and moderated feedback before it could be responsibly deployed in cultural-heritage settings.

Table 7. Integrated readiness classification and adoption decision

Evidence	Observed value	Category	Decision implication
ResNet50 accuracy	0.45	Low technical readiness	Not ready for operational deployment
ResNet50 macro F1-score	0.40	Low technical readiness	High classification risk
Class-level stability	Several classes near-zero F1-score	Low technical readiness	Requires dataset and label improvement

 ISSN 2723-6471

Mean TOEH+Feedback readiness	0.787	High perceived readiness	Implementation-side support exists
Readiness gap	0.362	Asymmetry	Develop, validate, and pilot before deployment
Scoring interpretation	Normalized descriptive proxy	Exploratory	Requires validation through weighting, expert judgment, or multi-criteria decision analysis

Table 7 summarizes the evidence-level classification used to determine the current adoption position. The table makes explicit that the readiness gap is a descriptive indicator of asymmetry, not a compensatory score. Because the technical evidence remains low, the decision logic proceeds to a non-deployment recommendation despite high perceived implementation-side readiness.

Table 8 translates the dual-layer evidence into actionable adoption alternatives. In the current case, the relevant quadrant is low technical readiness and high perceived implementation-side readiness; therefore, the recommended action is to improve the dataset, model, security preparation, and validation design before deployment. The governance controls required to support such staged improvement are specified separately in table 9.

Table 8. Dual-layer decision matrix

Technical readiness	Perceived implementation-side readiness	Decision	Recommended action
High	High	Controlled deployment	Pilot with governance and monitoring
Low	High	Not ready for deployment	Improve dataset, model, security, and validation
High	Low	Delay adoption	Strengthen organizational capability
Low	Low	Avoid implementation	Redesign system and readiness program

Table 9 complements the decision matrix by identifying minimum security-readiness controls across the AI lifecycle. These controls are not presented as results of empirical penetration or adversarial testing; rather, they define governance checkpoints that should be satisfied before moving from an exploratory prototype to a controlled pilot.

Table 9. Security readiness checklist for staged deployment

AI lifecycle stage	Potential risk	Readiness indicator	Recommended control
Data collection	Unverified sources	Dataset provenance recorded	Source logging and metadata validation
Data labeling	Incorrect motif label	Expert review available	Label validation protocol
Training data	Data poisoning	Dataset integrity maintained	Version control and anomaly checking
Inference	Misleading prediction	Low-confidence output detected	Confidence threshold and human review
Feedback	Malicious or biased feedback	Feedback traceability	Moderated feedback and audit log
Model update	Uncontrolled model drift	Update history recorded	Controlled retraining protocol

The main implication is that organizational or implementation-side readiness should be interpreted as capacity for investment and improvement, not as justification for immediate deployment. In cultural heritage contexts, premature deployment may reduce trust and distort cultural interpretation. The framework therefore supports staged adoption: dataset governance and expert validation first, model refinement and security preparation second, and controlled pilot implementation only after technical reliability improves.

4.4. Discussion and Threats to Validity

The findings demonstrate that low-performing artificial-intelligence models can provide scientific value only when the failure patterns are analyzed transparently and the deployment claims are restricted accordingly [18], [19]. In this study, the low ResNet50 performance is not presented as a successful classifier; it is treated as evidence that the current dataset, labels, transfer-learning configuration, and validation design are insufficient for operational batik motif recognition. This distinction prevents diagnostic interpretation from becoming a justification for methodological weakness.

The diagnostic contribution is therefore bounded. The framework helps identify the mismatch between technical readiness and perceived implementation-side readiness, but it does not replace stronger empirical validation. A future operational version should include repeated train-test splits or cross-validation, confidence intervals, statistical comparison among architectures, external test data, and class-pair confusion analysis. These additions are necessary because small multi-class datasets are highly sensitive to random data partitioning and motif-label ambiguity.

The practical implication is that dataset development should be treated as a governance activity, not only as a preprocessing task [14], [17]. Before deployment, image sources should be documented, duplicate and low-quality samples should be removed systematically, labels should be validated by batik or cultural-domain experts, and class definitions should be reviewed to reduce semantic overlap. Model improvement should also involve more robust validation strategies, including repeated experiments, external test sets, and possibly ensemble or fine-tuned architectures. These steps are prerequisites for moving from an exploratory prototype toward a controlled pilot.

The security-readiness checklist complements this staged path because prior work shows that machine-learning systems are exposed to adversarial manipulation, data poisoning, privacy leakage, and feedback-related risks [10], [11], [12], [25], [26], [27], [28], [29]. Dataset provenance, version control, confidence thresholding, human review, moderated feedback, and controlled retraining are necessary because an AI-assisted recognition service is not only a classifier but also a socio-technical system that interacts with users, data, and institutional workflows [30], [31], [32]. Without these controls, feedback loops may amplify bias, low-confidence outputs may be accepted as authoritative, and model updates may introduce drift. Security and governance therefore become conditions for trustworthy adoption rather than optional implementation details.

The study also reframes adoption as a non-substitutable dual-layer problem. High perceived readiness among IT practitioners indicates that implementation-side actors may support AI initiatives, but it does not prove institutional readiness among batik MSMEs, museums, or cultural heritage agencies. Conversely, a technically promising model would still require governance, feedback control, human capability, and security monitoring. This dual interpretation is the core contribution of the paper.

Several threats to validity should be considered. First, the technical evaluation is constrained by dataset size and motif diversity. Although the dataset is relatively balanced, approximately 45-50 images per class remain limited for deep-learning-based 20-class recognition, especially when motifs share repetitive ornaments, similar geometric structures, and contextual visual meaning. Second, the study used a single experimental configuration and did not include repeated cross-validation, confidence intervals, statistical testing between architectures, or external validation using an independent dataset. The reported differences among VGGNet-16, ResNet50, and MobileNetV2 should therefore be interpreted as exploratory diagnostic evidence rather than definitive architecture ranking. Third, the evaluation did not test real operational factors such as image acquisition quality, lighting variation, mobile capture conditions, user interaction, system latency, or institutional workflow integration. Fourth, the readiness survey relied on self-reported perceptions from information-technology practitioners. These respondents are relevant as potential implementers, but they do not directly represent batik micro, small, and medium enterprises, museums, curators, artisans, or cultural heritage agencies. Fifth, the readiness-gap calculation is a transparent proxy rather than a validated quantitative decision-analytics model. Finally, security readiness was addressed conceptually through governance controls; no empirical penetration testing, adversarial evaluation, poisoning simulation, or privacy-risk assessment was conducted.

5. Conclusion

This study proposed a diagnostic dual-layer framework for evaluating staged artificial-intelligence adoption readiness in batik heritage recognition. The technical layer showed that ResNet50 achieved the best baseline performance, with 45% accuracy and a macro F1-score of 0.40, but this result remains insufficient for operational deployment. Class-level instability further confirms that the current system should be treated as an exploratory prototype rather than a deployable recognition service. The implementation-readiness layer showed high perceived implementation-side readiness among IT practitioners. However, this readiness should not be interpreted as direct institutional readiness among batik MSMEs or cultural heritage organizations. The integrated assessment revealed a readiness gap of 0.362, indicating that implementation support is stronger than technical maturity.

The study contributes a practical, non-compensatory decision-support logic for avoiding premature artificial-intelligence deployment. Under the proposed decision matrix, the current condition is low technical readiness and high perceived implementation-side readiness; therefore, deployment is not recommended. Priority should be given to dataset expansion, expert label validation, transfer-learning refinement, repeated validation, security governance, moderated feedback, and controlled pilot testing before real-world adoption.

Future work should expand this framework in three directions. Technically, future studies should enlarge the batik dataset, strengthen expert label validation, evaluate external test sets, apply repeated cross-validation, and compare stronger fine-tuned or ensemble architectures. Organizationally, future work should involve batik micro, small, and medium enterprises, museums, cultural heritage agencies, artisans, curators, and cultural experts to validate whether perceived implementation-side readiness translates into institutional readiness. Methodologically, the diagnostic framework can be operationalized through weighted multi-criteria decision analysis, risk-based deployment thresholds, explainable artificial intelligence, and longitudinal pilot studies so that staged adoption decisions become more precise and auditable.

6. Declarations

6.1. Author Contributions

Conceptualization: I.S., P.A.C., and E.B.S.; Methodology: P.A.C.; Software: E.B.S.; Validation: I.S., P.A.C., and E.B.S.; Formal Analysis: I.S., P.A.C., and E.B.S.; Investigation: I.S.; Resources: P.A.C.; Data Curation: P.A.C. and E.B.S.; Writing Original Draft Preparation: I.S., P.A.C., E.B.S., S., and CLL; Writing Review and Editing: I.S., P.A.C., E.B.S., S., and CLL; Visualization: I.S., and CLL; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received financial support for the research, writing, and/or publication of this article from Satya Wacana Christian University, under research contract no. 108/SPK-PF/RIK/09/2024.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] E. Cantemir and O. Kandemir, "Use of artificial neural networks in architecture: determining the architectural style of a building with a convolutional neural networks," *Neural Computing and Applications*, vol. 36, no. 11, pp. 6195–6207, Jan. 2024, doi: 10.1007/S00521-023-09395-Y.

[2] F. A. Mohammed, K. K. Tune, B. G. Assefa, M. Jett, and S. Muhie, "Medical Image Classifications Using Convolutional Neural Networks: A Survey of Current Methods and Statistical Modeling of the Literature," *Machine Learning and Knowledge Extraction*, vol. 6, no. 1, pp. 699–735, Mar. 2024, doi: 10.3390/make6010033.

[3] L. Chen, S. Li, Q. Bai, J. Yang, S. Jiang, and Y. Miao, "Review of image classification algorithms based on convolutional neural networks," *Remote Sensing*, vol. 13, no. 22, Art. no. 4712, pp. 1-51, 2021, doi: 10.3390/rs13224712.

[4] W. Hu, X. Li, C. Li, R. Li, T. Jiang, H. Sun, X. Huang, M. Grzegorzek, and X. Li, "A state-of-the-art survey of artificial neural networks for whole-slide image analysis: From popular convolutional neural networks to potential visual transformers," *Computers in Biology and Medicine*, vol. 161, no. July, Art. no. 107034, pp. 1-25, 2023, doi: 10.1016/j.compbiomed.2023.107034.

[5] E. A. Nabila, C. A. Sari, E. H. Rachmawanto, and M. Doheir, "A Good Performance of Convolutional Neural Network Based on AlexNet in Domestic Indonesian Car Types Classification," *Advance Sustainable Science, Engineering and Technology*, vol. 5, no. 3, pp. 0230302(01)–0230302(08), Oct. 2023, doi: 10.26877/ASSET.V5I3.16854.

[6] N. Al Said, D. Gura, and D. Karlov, "Efficiency of Smart AI-Based Voice Apps and Virtual Services Operating With Chatbots," *MENDEL*, vol. 28, no. 2, pp. 9–16, Dec. 2022, doi: 10.13164/MENDEL.2022.2.009.

[7] K. Jamil, W. Zhang, A. Anwar, and S. Mustafa, "Exploring the Influence of AI Adoption and Technological Readiness on Sustainable Performance in Pakistani Export Sector Manufacturing Small and Medium-Sized Enterprises," *Sustainability*, vol. 17, no. 8, pp. 1–28, Apr. 2025, doi: 10.3390/SU17083599.

[8] H. Felemban, M. Sohail, and K. Ruikar, "Exploring the Readiness of Organisations to Adopt Artificial Intelligence," *Buildings*, vol. 14, no. 8, pp. 1–21, Aug. 2024, doi: 10.3390/BUILDINGS14082460.

[9] J. Jöhnk, M. Weißert, and K. Wyrtki, "Ready or not, AI comes: An interview study of organizational AI readiness factors," *Business & Information Systems Engineering*, vol. 63, no. December, pp. 5–20, 2021, doi: 10.1007/s12599-020-00676-7.

[10] Y. Hu *et al.*, "Artificial Intelligence Security: Threats and Countermeasures," *ACM Computing Surveys*, vol. 55, no. 1, pp. 1–36, Jan. 2023, doi: 10.1145/3487890.

[11] B. Biggio and F. Roli, "Wild patterns: Ten years after the rise of adversarial machine learning," *Pattern Recognition*, vol. 84, no. December, pp. 317–331, Dec. 2018, doi: 10.1016/J.PATCOG.2018.07.023.

[12] J. Wang, C. Wang, Q. Lin, C. Luo, C. Wu, and J. Li, "Adversarial attacks and defenses in deep learning for image recognition: A survey," *Neurocomputing*, vol. 514, no. December, pp. 162–181, Dec. 2022, doi: 10.1016/J.NEUCOM.2022.09.004.

[13] S. Varshney, S. Singh, C. V. Lakshmi, and C. Patvardhan, "Content-based image retrieval of Indian traditional textile motifs using deep feature fusion," *Scientific Reports*, vol. 14, no. 1, pp. 1–17, May 2024, doi: 10.1038/s41598-024-56465-9.

[14] M. Mishra and P. B. Lourenço, "Artificial intelligence-assisted visual inspection for cultural heritage: State-of-the-art review," *Journal of Cultural Heritage*, vol. 66, no. March, pp. 536–550, Mar.–Apr. 2024, doi: 10.1016/j.culher.2024.01.005.

[15] A. L. Carvalho Ottoni and L. T. Cordeiro Ottoni, "A deep learning approach for cultural heritage building classification using transfer learning and data augmentation," *Journal of Cultural Heritage*, vol. 74, no. July, pp. 214–224, Jul.–Aug. 2025, doi: 10.1016/j.culher.2025.06.010.

[16] J. Zhu and C. Zhu, "Research on the innovative application of Shen Embroidery cultural heritage based on convolutional neural network," *Scientific Reports*, vol. 14, no. 1, pp. 1–11, Apr. 2024, doi: 10.1038/s41598-024-60121-7.

[17] M. Xu, S. Yoon, A. Fuentes, and D. S. Park, "A Comprehensive Survey of Image Augmentation Techniques for Deep Learning," *Pattern Recognition*, vol. 137, no. May, Art. no. 109347, pp. 1-12, May 2023, doi: 10.1016/J.PATCOG.2023.109347.

[18] G. Naidu, T. Zuva, and E. M. Sibanda, "A review of evaluation metrics in machine learning algorithms," in *Artificial Intelligence Application in Networks and Systems*, Lecture Notes in Networks and Systems, vol. 724, R. Silhavy and P. Silhavy, Eds. Cham, Switzerland: Springer, vol. 2023, no. July, pp. 15–25, 2023, doi: 10.1007/978-3-031-35314-7_2.

[19] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 6, pp. 1–13, Jan. 2020, doi: 10.1186/S12864-019-6413-7.

[20] V. Uren and J. S. Edwards, "Technology readiness and the organizational journey towards AI adoption: An empirical study," *International Journal of Information Management*, vol. 68, no. February, Art. no. 102588, pp. 1-12, Feb. 2023, doi: 10.1016/J.IJINFOMGT.2022.102588.

[21] A. N. Tehrani, S. Ray, S. K. Roy, R. L. Gruner, and F. P. Appio, "Decoding AI readiness: An in-depth analysis of key dimensions in multinational corporations," *Technovation*, Art. no. 102948, pp. 1-13, Mar. 2024, doi: 10.1016/J.TECHNOVATION.2023.102948.

[22] M. F. Arroyabe, C. F. A. Arranz, I. Fernandez De Arroyabe, and J. C. Fernandez de Arroyabe, "Analyzing AI adoption in European SMEs: A study of digital capabilities, innovation, and external environment," *Technology in Society*, vol. 79, Art. no. 102733, pp. 1-14, Dec. 2024, doi: 10.1016/J.TECHSOC.2024.102733.

[23] S. G. Ayinaddis, "Artificial intelligence adoption dynamics and knowledge in SMEs and large firms: A systematic review and bibliometric analysis," *Journal of Innovation & Knowledge*, vol. 10, no. 3, pp. 1–15, May 2025, doi: 10.1016/J.JIK.2025.100682.

[24] E. Ode, I. F. Awolowo, R. Nana, and F. S. Olawoyin, "Social capital and artificial intelligence readiness: The mediating role of cyber resilience and value construction of SMEs in resource-constrained environments," *Information Systems Frontiers*, vol. 2025, no. April, pp. 1-23, 2025, doi: 10.1007/s10796-025-10608-z.

[25] A. E. Cinà *et al.*, "Wild Patterns Reloaded: A Survey of Machine Learning Security against Training Data Poisoning," *ACM Computing Surveys*, vol. 55, no. 13s, pp. 1–39, Dec. 2023, doi: 10.1145/3585385.

[26] M. Rigaki and S. Garcia, "A Survey of Privacy Attacks in Machine Learning," *ACM Computing Surveys*, vol. 56, no. 4, pp. 1–34, Apr. 2024, doi: 10.1145/3624010.

[27] S. Z. El Mestari, G. Lenzini, and H. Demirci, "Preserving data privacy in machine learning systems," *Computers & Security*, vol. 137, Art. no. 103605, pp. 1–22, Feb. 2024, doi: 10.1016/J.COSE.2023.103605.

[28] F. A. Yerlikaya and Ş. Bahtiyar, "Data poisoning attacks against machine learning algorithms," *Expert Systems with Applications*, vol. 208, no. December, Art. no. 118101, pp. 1-22, Dec. 2022, doi: 10.1016/j.eswa.2022.118101.

[29] A. Paracha, J. Arshad, M. B. Farah, and K. Ismail, "Machine learning security and privacy: a review of threats and countermeasures," *EURASIP Journal on Information Security*, vol. 2024, no. 1, Art. no. 10, pp. 1-25, Apr. 2024, doi: 10.1186/S13635-024-00158-3.

[30] E. Mosqueira-Rey, E. Hernández-Pereira, D. Alonso-Ríos, J. Bobes-Bascarán, and Á. Fernández-Leal, "Human-in-the-loop machine learning: a state of the art," *Artificial Intelligence Review*, vol. 56, no. 4, pp. 3005–3054, Aug. 2022, doi: 10.1007/S10462-022-10246-W.

[31] J. Meyer, "Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions," *Computers and Education: Artificial Intelligence*, vol. 6, no. June, Art. no. 100199, pp. 1-11, Jun. 2024, doi: 10.1016/J.CAEAI.2023.100199.

[32] M. Glickman and T. Sharot, "How human–AI feedback loops alter human perceptual, emotional and social judgements," *Nature Human Behaviour*, vol. 9, no. 2, pp. 345–359, Dec. 2024, doi: 10.1038/s41562-024-02077-2.

[33] M. A. Rasyidi and T. Bariyah, "Batik pattern recognition using convolutional neural network," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 4, pp. 1430–1437, Aug. 2020, doi: 10.11591/eei.v9i4.2385.

[34] D. G. Meranggi, N. Yudistira, and Y. A. Sari, "Batik Classification Using Convolutional Neural Network with Data Improvements," *JOIV: International Journal on Informatics Visualization*, vol. 6, no. 1, pp. 6–11, Mar. 2022, doi: 10.30630/joiv.6.1.716.

[35] M. A. Rasyidi, R. Handayani, and F. Aziz, "Identification of batik making method from images using convolutional neural network with limited amount of data," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 3, pp. 1300–1307, Jun. 2021, doi: 10.11591/eei.v10i3.3035.