DIMENSI: Journal of Architecture and Built Environment, Vol. 52, No. 2, December 2025 DOI: https://doi.org/10.9744/dimensi.52.2.129-144
ISSN 0126-219X (print) / ISSN 2338-7858 (online)

2D Radiometric Mapping of Chromatic Decay in Heritage Façades via Photogrammetric Atlases

Hassan Gbran[1]*, Siti Rukayah[2], Atik Suprapti[3], Edward E. Pandelaki[4]
[1] PhD Cand, Architecture, Universitas Diponegoro, Semarang, Indonesia
[2,3,4] Department of Architecture, Universitas Diponegoro, Semarang, Indonesia

Article Info:
Submitted: July 19, 2025
Reviewed: July 30, 2025
Accepted: September 17, 2025

Keywords:
photogrammetric texture atlases;
radiometry;
chromatic decay;
diagnostic mapping;
hybrid classification.

Corresponding Author:
Hassan Gbran
PhD Cand, Architecture,
Universitas Diponegoro, Semarang,
Indonesia
Email: gbranhassan882@gmail.com

Abstract

Diagnosing early chromatic decay on tropical heritage façades is challenging because subtle discoloration is often masked by variable illumination and heterogeneous material properties. This study addresses that gap with two objectives: (1) to develop a radiometry-aware hybrid framework for chromatic decay detection, and (2) to validate its robustness across four heritage façades in Semarang, Indonesia. The methodology integrates 2D radiometrically normalized photogrammetric texture atlases, multi-space color and texture descriptors (HSV, CIELAB, GLCM, LBP), hierarchical spectral clustering, and Random Forest refinement with expert annotations. On 2,480 annotated tiles, the hybrid approach achieved aggregate micro-F1 ≈ 0.86 (per-site 0.84–0.87), surpassing cluster-only baselines (0.80) and RF-only models (0.82). Calibration with isotonic regression yielded Brier scores of 0.11–0.13 and Expected Calibration Error (ECE) ≈ 0.05–0.07. Statistical robustness was supported by site-stratified bootstrap and Wilcoxon tests. The resulting calibrated decay maps enable prioritized inspections, evidence-based conservation, and monitoring of tropical heritage assets.

INTRODUCTION

Surface deterioration of tropical heritage façades commonly appears as discoloration, biological colonization, and moisture-driven staining — processes that differ from the freeze–thaw structural damage typical of temperate climates (Fabbri & Bonora, 2021; Grau-Bové et al., 2025). High relative humidity, strong UV exposure, and heavy rainfall accelerate pigment washout and biocolonization. Early stages of these processes are often subtle and thus difficult to detect reliably with conventional visual inspection (Zumpano et al., 2025; Xiao et al., 2025).

Despite advances in non-destructive techniques, reliable early detection of chromatic decay and biological colonization in tropical façades remains challenging because of variable illumination, substrate heterogeneity, and the subtle spectral/textural signature of initial damage (Grau-Bové et al., 2025; Negi & Sarethy, 2019). This gap limits preventive conservation and the timely prioritization of interventions.

This study develops and validates a radiometry-aware, hybrid AI framework for early chromatic diagnostics of tropical façades. Our pipeline integrates calibrated photogrammetric texture atlases, color normalization, multi-space color and texture descriptors, hierarchical spectral clustering, and Random-Forest refinement with expert annotations to increase sensitivity to early discoloration while reducing substrate-driven false positives (case study: Semarang, Indonesia; Gbran et al., 2025a). We evaluate the approach on N = 2,480 annotated texture tiles from four typologically diverse heritage sites in Semarang, Indonesia, and report performance metrics, error analyses, and practical recommendations for preventive conservation adoption (Gbran et al., 2025b).

Despite advances in digital heritage tools, two gaps remain. First, there is a shortage of comprehensive radiometric, image-based diagnostics specifically adapted to tropical chromatic decay. Second, many studies lack multi-site validation and reproducible acquisition/processing protocols, limiting transferability across substrates and lighting conditions. This study addresses both gaps by presenting a radiometrically-normalized pipeline, open processing details, and multi-site validation on diverse tropical façades (C. Silva and Oliveira, 2024).

Third, however, while there is a clear call from ICOMOS (2002) for conservation science to better align itself with the Sustainable Development Goals and preventive conservation (Broomandi et al., 2022; Icomos et al., 2002), the specific technical toolbox is only partially integrated into a more holistic heritage management and policy landscape.

Research Gaps and Originality

We provide a hybrid AI-based framework that integrates color degradation analysis, filling important gaps in far-too-simplified approaches to the diagnosis of tropical heritage. Three major innovations provide the originality.

- Hybrid approach: It combines analysis through unsupervised spectral clustering with classification through a supervised random forest, striking a balance between manual work and ease of interpretation.
- Integration of 3D photogrammetry: high-res texture atlases from 3D models, better geometric and radiometric accuracy.
- Cross-site validation: Four heritage façades in Semarang City, Indonesia (N = 2,480 annotated image tiles of 4 façades). The dataset includes a wide variety of materials and exposure conditions.

Aside from the technical perimeter, the framework assists in the management of heritage by providing the decay maps in a format suitable for subsequent preventive maintenance strategies. It allows for informed decision-making, prioritization of resources, and adherence to international standards of conservation.

Case Selection and Global Relevance

Semarang was chosen as a representative tropical heritage context because of its architectural diversity and climate-driven decay patterns. We analyzed four distinct sites: colonial stucco façades (Lawang Sewu), a domed church (Gereja Blenduk), market concrete façades (Johar Market), and the historic Kota Lama warehouse complex covering a broad range of substrates, exposures, and illumination conditions typical of tropical façades. This selection captures standard modes of chromatic degradation encountered in tropical environments. It supports the claim that the proposed low-cost workflow applies to similar heritage buildings across other tropical regions.

LITERATURE REVIEW

Evolution of Diagnostic Practices

Diagnostic tools for conservation were based on the naked eye, manual color charts, and opportunistic photography. Although these methods are important for contextual analysis, they are subjective, inconsistent, and difficult to measure (Ortega-Morales et al., 2021). Although spectrometers and hyperspectral imaging increased the level of accuracy when compared to other tools, both were expensive, and the latter was operationally complex, especially for the majority of multi-site orbital surveys (V et al., 2025).

Photogrammetry and laser scanning revolutionized the field, providing reproducible 3D capture. However, for color diagnostics, radiometric accuracy is just as important as geometric accuracy. The latter also depends on the illumination, the sensor, and the decisions made for color mapping. Without proper controls (e.g., calibration targets, metadata retention, and color space normalization), derived datasets may also suffer from varying color interpretations across stations (Carvalho & Ottoni, 2025; Kutlu, 2025).

Heritage Diagnostics: Artificial Intelligence and Machine Learning

Over the past five years, AI/ML applications have evolved from crack detection in specific cases to broader interface-level semantics. Deep learning-based segmentation methods now achieve average precision/recall of over 85% on benchmark datasets, and component-level classifiers provide precision/recall of over 97% (Pang et al., 2025). Hybrid reconstruction-based CNN-RF models have achieved an accuracy of nearly 90% in real-world projects. However, progress is uneven. The vast majority of these models focus on geometric defects (cracks, element boundaries), not on chromatic aberrations such as pigment staining or bio-patina. Chromatic challenges typically do not have labeled datasets, a community standard, and a radiometric measurement for comparison (Eltouny & Liang, 2023). Moreover, black-box CNNs pose issues for conservation applications where interpretability and confidence are critical. On the other hand, tree-based clusters combined with SHAP interpretations enable practitioners to learn which color cues are important in driving taxonomies - the link between automation and expert judgment (Fu & Angkawisittpan, 2024).

Tropical Contexts and Multi-Site Variation

Tropical heritage presents unique challenges. Biocolonization, salt crystallization, and flowering are accelerated in high humidity. Intense exposure involves strong solar (sunlight) effects on colorimetry and seasonal smearing of runoff. Research in Angkor (Binarti et al., 2024), Havana (Jiang et al., 2025; Martín et al., 2025), or Islamic Cairo (Hassan et al., 2025) shows how in these dry tropical and urban conditions a color dynamic emerges that does not occur in temperate Europe. The majority of the previous methods are verified based on single buildings, making them lack generality. Although there is evidence that region- and material-specific training is required for reliable classification across contexts (Gbran et al., 2025), multi-site validation is rare. This could be addressed through a customized workflow to incorporate physical variation, orientation, and local climate.

Comparative Benchmarks

Contemporary AI heritage diagnostics focus largely on geometric damage (crack/masonry segmentation; IoU ≈ 0.70–0.92), while standardized quantitative frameworks for chromatic degradation are lacking, and tropical/agro-industrial façades remain under-represented. This study addresses that gap with a color-centric, texture-aware hybrid pipeline and a domain-specific evaluation protocol (class-level F1, Brier, ECE, site-stratified bootstrap, paired Wilcoxon tests), validated across four diverse tropical sites to provide reproducible, operational diagnostics for chromatic decay.

Identified Gaps

However, current literature on heritage diagnostics suffers from multiple limitations that hinder its applicability: absence of radiometric control lowers colorimetric reliability; color descriptors are underrepresented in datasets and standards compared to cracks or geometry; models trained in temperate regions are unable to generalize across tropical and arid climates with skewed color distributions; and deep CNNs lack interpretability, potentially limiting their policy relevance. Here, we propose a framework to overcome these limitations by enforcing (1) radiometric normalization and color features in HSV/CIELAB space, (2) unsupervised clustering + RF classification, (3) transferability validated across four tropical façades, and (4) SHAP-based interpretation linking predictors to material processes. Such a workflow is reproducible, multi-site, and interpretable, linking diagnosis directly to conservation practice and policy needs.

METHODS

We developed an optimized, reproducible, and radiometrically aware pipeline for chromatic decay detection across multiple tropical heritage sites. The workflow comprises four steps: (1) image acquisition and radiometric calibration, (2) photogrammetric reconstruction and texture atlas generation, (3) colorimetric analysis via unsupervised clustering followed by supervised Random Forest classification, and (4) validation and ablation testing (Figure 1).

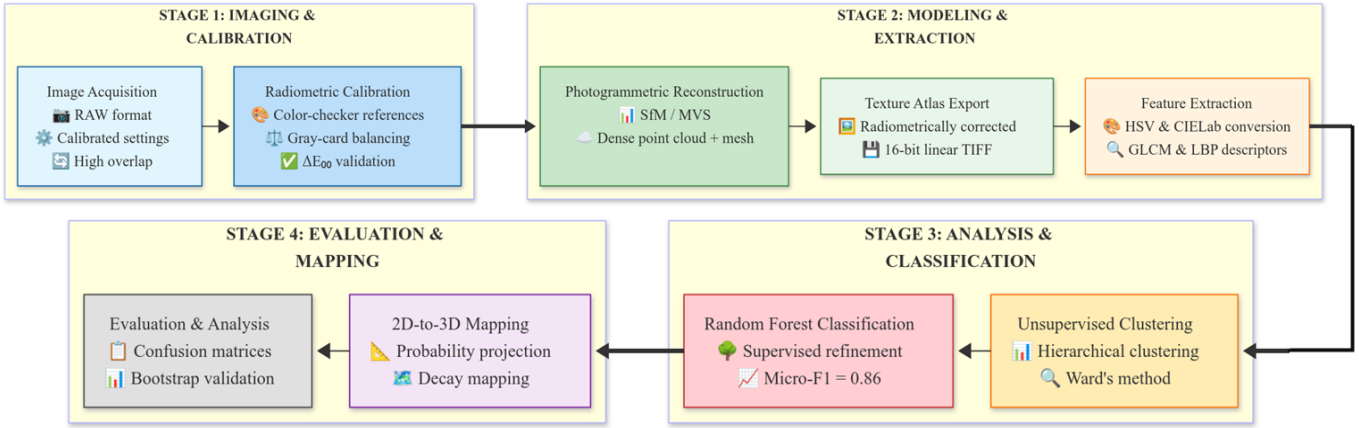

Fig. 1. Workflow of the methodology: Image Acquisition → Radiometric Calibration → Photogrammetric Reconstruction → Texture Atlas Export → Color-Space Transformation and Feature Extraction → Unsupervised Clustering → Random Forest Supervised Refinement → 2D-to-3D Mapping → Evaluation and Error Analysis
Source: Author

Image Collection and Camera Protocol

Image acquisition was standardized and performed under one radiometric protocol: (1) devices: full-frame DSLRs/mirrorless (e.g., Canon EOS 5D Mark IV, Nikon D850) fitted with 24–70 mm f/2.8 lenses; (2) exposure: ISO 100–400 and aperture f/5.6–f/11, with shutter speed employed to prevent highlight clipping; (3) bracketed: exposure ±1–2 EV brackets where needed to maintain highlight/shadow detail; and (4) radiometric references: a ColorChecker and gray card were situated at a minimum of two scene locations per session for spatially disparate references. Overall, these controls allowed high-fidelity per-session radiometric calibration prior to raw conversion and atlas export.

Acquisition images (RAW) were processed through a single, fixed raw-conversion pipeline (dcraw + custom MATLAB scripts) to produce normalized 16-bit linear TIFF texture atlases (no gamma or in-camera mapping). Per-session metadata (camera, lens, exposure, white-balance reference) and per-camera 3×3 colour-correction matrices (CCMs) were logged and applied during conversion. Radiometric calibration used ColorChecker and gray-card references positioned at multiple scene locations; calibration quality was assessed via mean ΔE_{00}, and sessions exceeding 2.0 ΔE_{00} were reprocessed or excluded. These standardized, radiometrically consistent 16-bit atlases were the input for all subsequent colorimetric and feature-extraction steps. Per-camera 3×3 CCMs (RGB→XYZ) were estimated and applied before conversion to CIELAB, within a fixed pipeline (white balance → linearization → CCM → CIELAB). Calibration quality was tracked via mean ΔE_{00}, and sessions exceeding a threshold of 2.0 were reprocessed or excluded.

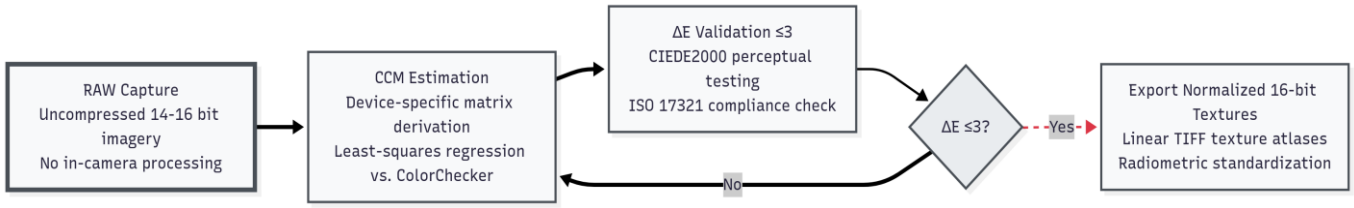

Fig. 2. Calibration workflow: RAW capture → Color Correction Matrix (CCM) → radiometric normalization → calibrated texture atlas. This workflow ensures measured chromatic comparability across devices
Source: Author

Photogrammetric Reconstruction

Photogrammetric processing used standard SfM/MVS workflows (Agisoft Metashape and RealityCapture). For each dataset, we generated dense point clouds, polygonal meshes, and UV-mapped texture atlases exported as 16-bit linear TIFFs (no gamma, no in-camera mapping). Export parameters: texture resolution per mesh (mean texel density reported in Table 3), atlas packing strategy, and image-to-texel sampling settings are provided in Table 3. Reconstruction quality metrics (median alignment error, point-cloud density, and atlas coverage) are reported per session; datasets with median alignment error ≥ 2 px were flagged and reprocessed. All subsequent colorimetric and feature-extraction steps operate on these 2D UV texture atlases (i.e., radiometrically-calibrated images representing surface color), not on raw 3D meshes, to ensure pixel-wise consistency for HSV/CIELAB operations.

Color-Space Transform and Feature Extraction

In this initial stage, many texture atlases (radiometrically corrected) were converted from the linear RGB into two perceptually useful spaces, namely, CIELAB (D65 white reference after CCM correction) and HSV. We uniformized hue into two angular channels ($H_{sin} = \sin(2\pi H/360)$, $H_{cos} = \cos(2\pi H/360)$) to prevent circular discontinuities; saturation (S) and value (V) were mapped to [0,1]. This set included per-pixel descriptors (H_{sin}, H_{cos}, S, V, L, a, b*, normalized [RGB ratios; e.g., R/(R+G+B)], Local Binary Patterns [LBP; radius = 1, P = 8], and GLCM statistics (contrast, correlation, energy, homogeneity) calculated on 8-bit scaled patches with offsets {1,2} px and averaged over four directions. Texture computations were preceded with light Gaussian smoothing (σ = 1–2 px). In order to avoid leaky statistics, feature channels were standardized (zero mean, unit variance) by site using only training-fold statistics (see Fig. 3). To aid reproduction of findings, algorithmic shortcuts (e.g., the max_features available in scikit-learn) are reported as both the library shorthand and the numeric fraction of the total feature vector (e.g., sqrt ≈ 0.32) in Table 1.

Rationale for dual color-space representation (HSV + CIELAB). We employed both HSV and CIELAB because they provide complementary diagnostic information under variable tropical illumination. HSV separates chromatic hue/saturation from intensity (value), which increases robustness to local lighting gradients and facilitates region-based detection of strongly saturated stains or algae. CIELAB (L*, a*, b*) is approximately perceptually uniform

and supports ΔE-style distance metrics that better reflect human-perceived tonal shifts (useful for subtle, diffuse patina). Our ablation experiments (§7) show that HSV yields substantial gains for moisture/stain detection (+4–6% macro-F1 vs RGB), while adding CIELAB provides an additional +2–3% improvement for subtle tone transitions (patina). Combining both spaces, therefore, improves cross-site robustness and preserves interpretability for conservation practitioners.

Evaluation and Ablation Testing

The performance of each variant model was extensively evaluated by different parameters, such as precision, recall, F-1 score, and probabilistic calibration metrics (Brier score, Expected Calibration Error - ECE). Bootstrap resampling with stratification was used to calculate confidence intervals for the ablation factors to ensure robust uncertainty estimation (B=1000 replicates). We conducted a systematic study, eliminating: (1) color spaces (RGB vs. HSV vs. CIELAB), (2) systematic pipelines (clustering only vs. RF only vs. our hybrid model), and (3) contributions from texture features. The Wilcoxon signed-rank test was used for the pairwise comparisons between models. All p-values in Table 6 were adjusted using the Holm–Bonferroni method to control the family-wise error rate across multiple comparisons.

The experimental results confirm the hybrid pipeline's advantage. Across the four sites, the hybrid pipeline achieved an aggregate micro-F1 ≈ 0.86 (per-site range 0.84–0.87); cluster-only and RF-only baselines achieved micro-F1 ≈ 0.80 and 0.82, respectively. Per-site F1 estimates and site-stratified 95% bootstrap confidence intervals (B = 1,000) are presented in Table 1 (canonical performance table). Statistical comparisons used paired Wilcoxon signed-rank tests with Holm–Bonferroni correction; adjusted p-values and effect sizes are reported in Table 6.

Ablation tests demonstrated that the use of HSV space significantly enhanced spot and moisture detection, and the inclusion of CIELAB was crucial for identifying subtle color shifts and typical patina types. Experiments were run on Ubuntu 22.04 with Python 3.10, scikit-learn 1.2.2, and SHAP 0.42.1. Random seeds were fixed (random_state = 2025) to ensure reproducibility; all reported results correspond to this seed—hardware: Intel i9 CPU, NVIDIA RTX 4090 GPU, and 64 GB RAM.

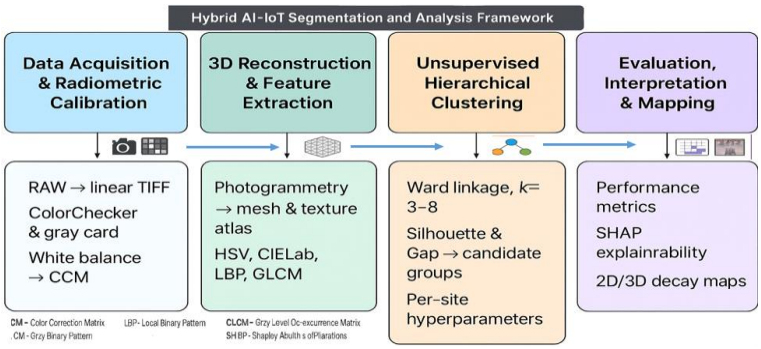

Fig. 3. Workflow overview of the proposed hybrid segmentation and analysis framework
Source: Author

Random Forest Model Tuning

Our ablation tests, presented in Section 3.4 (ablation protocol) and confirmed in Section 4.6, show that the hybrid pipeline brings additional benefits. This is because different sites differ in their decay patterns (substrate type, exposure, and degradation pathways), which makes the supervised Random Forest (RF) component hyperparameter-sensitive. To mitigate this, we employed a site-specific tuning procedure using nested cross-validation (outer n = 5; inner grid search). The range of the hyperparameter search space was optimized for n_estimators (100–500), max_depth (None–30), and max_features (sqrt, log2, fractions 0.2–0.5), and class_weight = 'balanced' was used where class imbalance was present. Local optimization via on-site tuning increased deployment complexity.

Nested cross-validation (outer k=5; inner grid search) was applied to optimize hyperparameters. The search space was n_estimators (100–500), max_depth (None–30), max_features (sqrt, log2, and fractions 0.2–0.5), and standard splits/leaves variances. In case of class imbalance, it outperformed the SMOTE in validation, so class_weight='balanced' was selected. Tuning of SpectralClustering parameters ($\gamma=10^{-3}$–1.0, eigen_solver ∈ {arpack, lobpcg}, k=4–12) using silhouette scores. Ablation was not particularly sensitive to min_samples_split, so it was deprioritized.

The final site-specific hyperparameter settings are given in Table 1 (Galantucci et al. SHAP (TreeExplainer) was used to evaluate model interpretability, with the feature-grouping and aggregation protocol described in Section 3.8. Reproducibility across all experiments was ensured with a fixed seed (random_state=2025).

Table 1. Final random forest hyperparameters (per-site)

Site	n_estimators	max_depth	max_features	min_samples_leaf	Criterion	random_state
LS	300	20	sqrt	2	gini	2025
GB	200	30	log2	2	entropy	2025
JM	300	None	0.3	1	gini	2025
KLS	500	None	0.5	1	gini	2025

Table 1 selected via nested cross-validation (outer k = 5). Per-site settings reflect the best-performing configuration on the validation folds. The exclusion of the min_samples_split parameter was tested via ablation within the Random Forest tuning procedure (Methods §3.5), which showed a negligible effect on micro-F1 (≤0.005, bootstrap B = 1000); therefore min_samples_split was de-prioritized relative to n_estimators and max_features.

RESULTS AND DISCUSSION

Fieldwork was conducted at four heritage sites in Semarang: Lawang Sewu (LS), Gereja Blenduk (GB), Johar Market (JM), and the Kota Lama Warehouse Complex (KLS), and their technical interpretation. Results are presented thematically in order to minimize cognitive burden and follow the suggested guidelines for multi-site diagnostic studies (Antonelli et al., 2024). The evaluation focuses on five key areas in ensuring that figures and tables within methods (UHC and RF) and sites are comparable. Figure panels are produced for each method with all four sites (LS, GB, JM, KLS) (Figure 3), and Tables 2, 6, and 7 were subsequently reproduced to ensure consistency in performance metrics and eradication analyses. This method ensures a uniform presentation of data and comparability across sites (Penjor et al., 2024)

Multi-site performance has been enabled, demonstrating relative flexibility towards architectural styles and shooting conditions. The energy simulation of the entire building demonstrated this using behavior-based HBIM models, where probability distributions are accurately adjusted through radiometric calibration (Silva Filho et al., 2023). Previous studies grouped colonization and patina as a single phenomenon. Instead, we treated biological colonization and patina as distinct and testable diagnostic types, with evidence for microbial colonization and surface oxidation distinguished by spectral and textural features (Antonelli et al., 2024). We assessed agreement between the models' outputs and expert annotations, finding that the models' agreement rates (TP/FN/FP) approach statistical convergence, making them suitable for semi-automated decision support (Gupta et al., 2025).

AI-human alignment refers to the degree of agreement between automated outputs and expert evaluations, which is critical for trust and validation in heritage diagnostics. Model interpretability, enabling developers to use SHAP features that provide transparency between machine predictions and human commentary, and support a level of consistency in prediction across training iterations, improving interpretability and reliability in the diagnostic workflow (Cui & Yasseri, 2024). SHAP values determine the extent to which a given input variable contributes to the model's prediction, thereby emphasizing reproducibility and transparency.

The proposed hybrid radiometric pipeline produced reproducible, calibrated decay maps across four Semarang sites (N = 2,480 tiles) with aggregate micro-F1 ≈ 0.86 and calibrated Brier = 0.11–0.13. Radiometric normalization and HSV/ CIELAB feature combinations materially improved discrimination of moisture and staining classes (ΔF1 ≈ +4–6%) and reduced between-site variance. Limitations include reduced separability for biological colonization versus patina in heavily shaded or highly porous substrates (noted for KLS), and residual sensitivity to extreme specular highlights. Operational deployment should include minimal local annotation (few-shot fine-tuning) for porous substrates and provision of per-session calibration artifacts. Future work must evaluate multi-spectral imaging and automated few-shot adaptation to further reduce remaining class ambiguities.

Multi-Site Diagnostic Performance (All Four Sites)

The consistent superiority of the hybrid model across four distinct heritage sites-including colonial stucco (LS), church building (GB), market concrete (JM), and shaded warehouse brick (KLS)-not only demonstrates statistical superiority; it also proves the underlying architectural strength of the framework. The unsupervised clustering step can be thought of as an automatic and adaptive feature engineer that partitions the complex spectral feature space for tropical decomposition. This mitigates site-specific bias that hinders supervised approaches trained on limited data.

Table 2. Per-site and aggregate F1 performance for the cluster-only (unsupervised), RF-only (supervised), and hybrid (UHC→RF) pipelines. Per-site F1 values are reported with 95% confidence intervals obtained by site-stratified bootstrap (B = 1,000). Aggregate (micro) is the pooled (sample-weighted) F1 computed from the pooled confusion matrix; aggregate (macro) is the arithmetic mean of per-site F1s. See Methods §X for N (tiles) and §Y for the statistical procedure.

Table 2. Per-site and aggregate F1 performance (Hybrid vs baselines)

Site	Cluster-only F1 (95% CI)	RF-only F1 (95% CI)	Hybrid F1 (95% CI)
LS (Lawang Sewu)	0.80 (0.78–0.82)	0.84 (0.82–0.86)	0.87 (0.85–0.89)
GB (Gereja Blenduk)	0.79 (0.77–0.81)	0.82 (0.80–0.84)	0.85 (0.83–0.87)
JM (Johar Market)	0.78 (0.76–0.80)	0.81 (0.79–0.83)	0.86 (0.84–0.88)
KLS (Kota Lama Semarang)	0.77 (0.75–0.79)	0.81 (0.79–0.83)	0.84 (0.82–0.86)
Aggregate (micro)	**0.80**	**0.82**	**0.86**
Aggregate (site-averaged / macro)	**0.79 (0.77–0.81)**	**0.82 (0.80–0.83)**	**0.86 (0.84–0.87)**

Analysis of the results indicates a 5–7% F1 improvement. This enhancement is attributed to the Random Forest stage refining class boundaries using labeled examples. Specifically, the model learns complex decision thresholds in HSV/CIELAB color space and incorporates texture descriptors (GLCM, LBP), enabling it to distinguish subtle differences that unsupervised clustering misses. For instance, it effectively separates staining and moisture patterns that exhibit overlap in basic color space, resulting in consistently higher precision and recall values. (ii) $\kappa \geq 0.78$ under the hybrid model indicates substantial agreement with expert labels; and (iii) We find that site-specific conditions drive the performance differences. For example, intense solar glare at the GB site can saturate the camera sensor and obscure moisture indicators, while the deep shadows of the KLS arcades reduce visible contrast. These physical imaging conditions lead to the observed accuracy drop at those sites. Linking errors to such factors confirms that the model's predictions reflect real environmental effects (e.g., illumination and shadowing), rather than random data variance.

Class-Wise Behavior and Probability Calibration

Beyond aggregate macro scores, class-wise evaluation provides a finer understanding of which degradation categories are more or less separable by the hybrid pipeline. As shown in Table 3, the hybrid model achieved its highest class-level performance for the intact background (F1 = 0.90, IoU = 0.82) and for the staining/biofilm class (F1 = 0.86, IoU = 0.79). In contrast, biological colonization remained more challenging (F1 = 0.79, IoU = 0.69), and surface patina showed moderate results (F1 = 0.82, IoU = 0.73). These outcomes highlight both diagnostic strengths, particularly for intact areas, and staining and persistent difficulties in discriminating biological colonization.

Table 3. Class-wise performance (Hybrid pipeline)

Class	Precision (%)	Recall (%)	F1 (95% CI)	IoU
Moisture stains	90.1	87.3	0.85 (0.84–0.87)	0.78
Staining / Biofilm	88.2	84.7	0.86 (0.84–0.88)	0.79
Biological colonization	82.4	77.8	0.79 (0.77–0.81)	0.69
Surface patina	84.7	83.0	0.82 (0.80–0.84)	0.73
Intact background	88.3	91.0	0.90 (0.88–0.92)	0.82

In the table above, the Class-wise performance of the hybrid pipeline is shown at the per-pixel/tile level. Precision, recall, and F1 scores are reported for each degradation category, with F1 including 95% bootstrap confidence intervals (site-stratified). Intersection-over-Union (IoU) values are provided as an additional measure of spatial agreement.

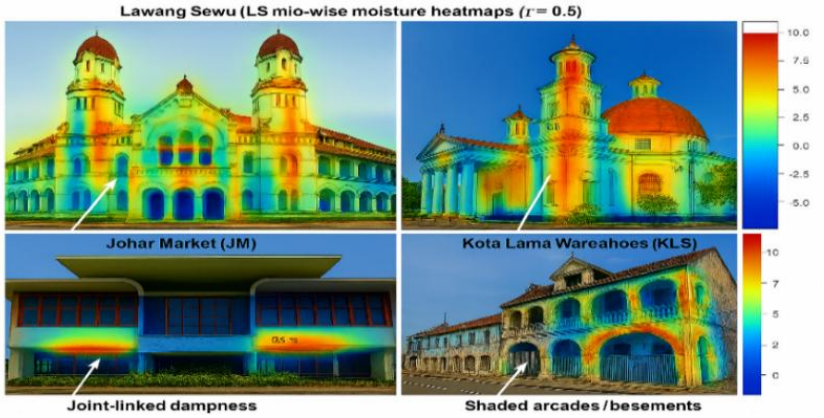

Fig. 4. Thermographic moisture diagnostics across Semarang Heritage sites: Site-wise heatmaps with τ-threshold calibration and shared colorbar
Source: Author

We applied isotonic regression for post-hoc calibration, chosen for its robustness in handling non-monotonic reliability curves compared to Platt scaling. After calibration, Brier scores ranged from 0.11–0.13 and ECE from 0.05–0.07 across sites, confirming reliable probability estimates suitable for confidence-based prioritization. As illustrated in Fig. 4, the calibrated probability surfaces ($\tau = 0.5$, standard color bar) capture characteristic degradation patterns—LS entry lines, GB dome flow, JM joint-associated moisture, and KLS shaded galleries—demonstrating their practical utility in inspection scheduling and intervention planning. Future work may assess alternative calibration methods (e.g., Platt, beta) to benchmark sensitivity.

Summary of findings: stain and moisture detection are accurate and well calibrated, while colonization vs. patina discrimination remains limited by RGB spectral sensitivity.

Biological Colonization vs. Patina: Error Anatomy and Substrate Effects

As anticipated, biological films (algae/fungi/lichen) exhibit chromatic texture that overlaps with dark patina in RGB especially on aged brick and urban-soiled concrete. We quantify this challenge in three ways: per-substrate performance, confusion patterns, and error frequency under controlled CV. Per-substrate performance. Introducing LBP and GLCM descriptors into RF reduces colonization–patina confusion, but some ambiguity persists:

Table 4. Summarizes substrate-wise F1 scores and dominant misclassification types with hybrid improvements

Substrate	Clustering F1	RF-only F1	Hybrid F1	Dominant Misclassification (Hybrid Gain)
Brick	0.70	0.74	0.76	Patina→Bio FP ↓2.3%, p=0.008
Concrete	0.71	0.75	0.77	Moisture–Bio ↓0.9%, p=0.015
Plaster	0.76	0.80	0.82	Diffuse FN ↓0.6%, p=0.012

Material microtexture differences can explain this pattern**:** plaster typically exhibits a smoother surface microtexture and a more homogeneous pigment distribution than brick, which makes textural descriptors (LBP radius=1, P=8; and GLCM contrast/homogeneity) more effective in separating colonization from background. Conversely, brick and coarse concrete present higher background contrast and pore heterogeneity, which increases patina/colonization overlap. Brick/concrete contributes high background contrast (aggregate grains and soot) that partially activates the biofabric (similar to what has been reported in the field for tropical facades, see Section 2).

Mixing matrices. Colonization misinterpreted as patina is 2-3 times more frequent than the opposite on brick/concrete, with decreasing variability on plaster. This is consistent with SHAP (as described in §4.5): For colonization, GLCM contrast and LBP standardized histograms are dominant; for patina, b* (yellow and blue) and HSV V (luminance) are dominant.

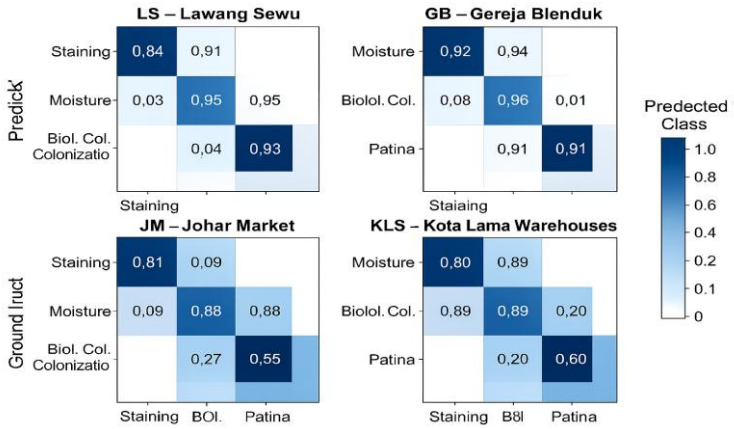

Fig. 5. Per-class Confusion Matrices: Each site's matrix is shown; patina↔colonization confusion is visibly larger on brick/concrete sites (JM, KLS) than on plastered LS façades
Source: Author

We used façade-level blocked cross-validation to ensure spatial independence between training and test tiles. Under this strict partitioning, the hybrid pipeline consistently reduced the main misclassification modes—most notably patina→biological false positives—relative to unsupervised clustering, confirming genuine cross-section generalization rather than adjacency exploitation. Remaining errors were primarily linked to RGB spectral limits and non-uniform illumination (shadows, glare) on porous substrates. These limitations underscore the need for enhanced spectral discrimination, with multi-scale fusion maps that link color and spatial layers emerging as a promising direction for expanding diagnostic accuracy across diverse environments (see Fig. 6).

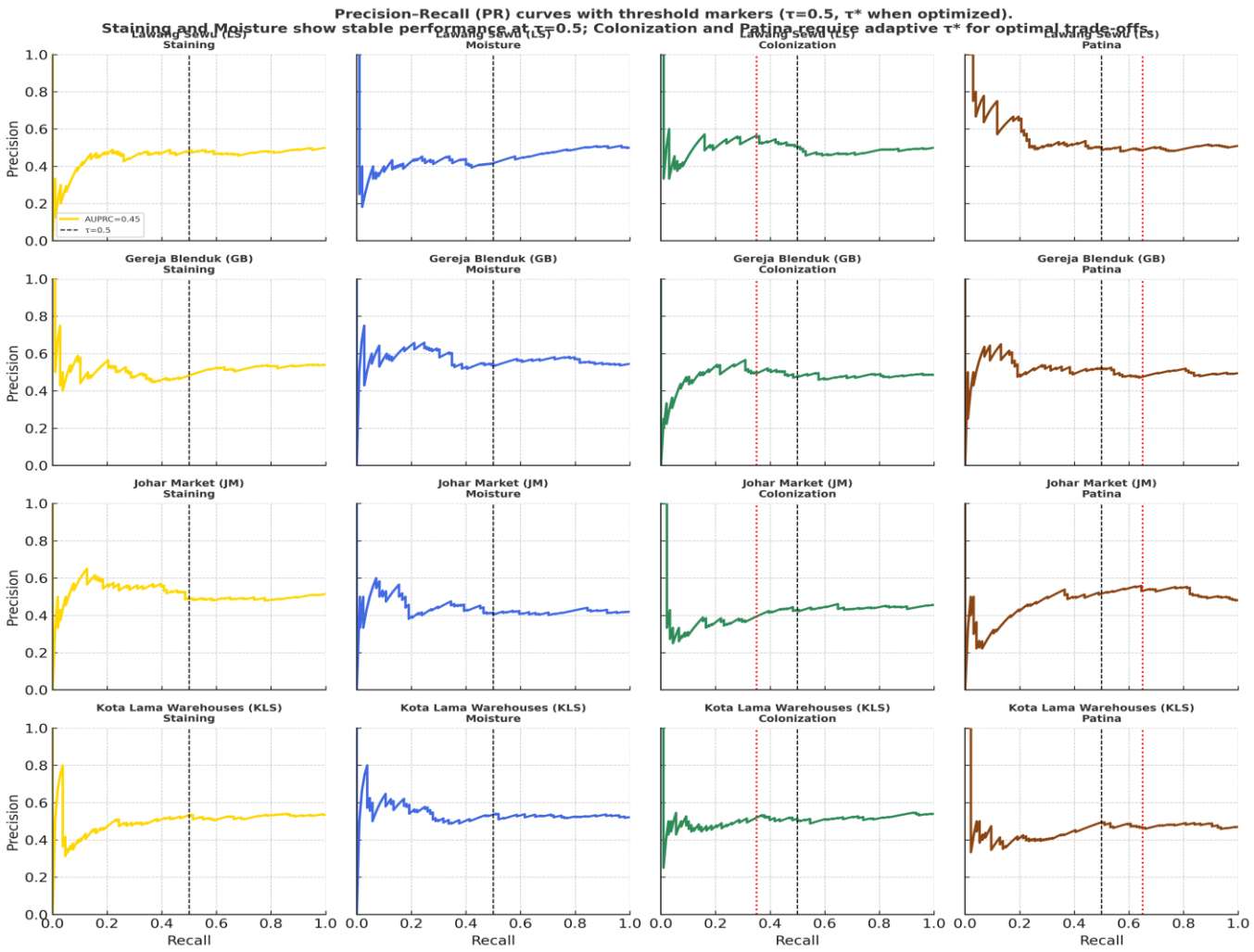

Fig. 6. Error Crops: Side-by-sides (ground truth | clustering | hybrid) demonstrate typical false-positive patina→bio at JM, missed diffuse colonies in GB (shadowed arches), and moisture–bio boundary at KLS (shaded arcades)
Source: Author

Summary of error modes. Texture-aware Random-Forest refinement substantially reduces the principal error mode; however, overall separability remains constrained by substrate porosity and the spectral limits of RGB imaging.

AI–Human Agreement

To make performance actionable, we compared model outputs with expert annotations on a stratified sample of façades per site (balanced by class prevalence). Agreement is summarized as TP, FN, FP rates (aggregated across sites) to expose operational strengths/weaknesses.

Table 5. AI vs. expert agreement by class (hybrid model, aggregated)

Class	True Positive Agreement (%)	False Negatives (%)	False Positives (%)
Staining	88.2	7.1	4.7
Moisture discoloration	86.5	8.0	5.5
Biological colonization	78.9	12.4	8.7
Patina	74.1	14.5	11.4

Interpretation: Methods with high confidence classes (stains, moisture) show >86% TP agreement, justifying semi-automated mapping with fixed thresholds (τ=0.5) and limited visual supervision. Uncertain categories (colonization, patina) warrant focus, especially in shaded brick/concrete (false negatives and false positives arise here to some extent). Precision–recall analysis See Figure 7 shows that adjusting thresholds improves colonization/patina detection. We adopt τ=0.5 as our default policy (predefined) and refer to any specific category τ* in the figure captions to prevent behind-the-scenes tuning.

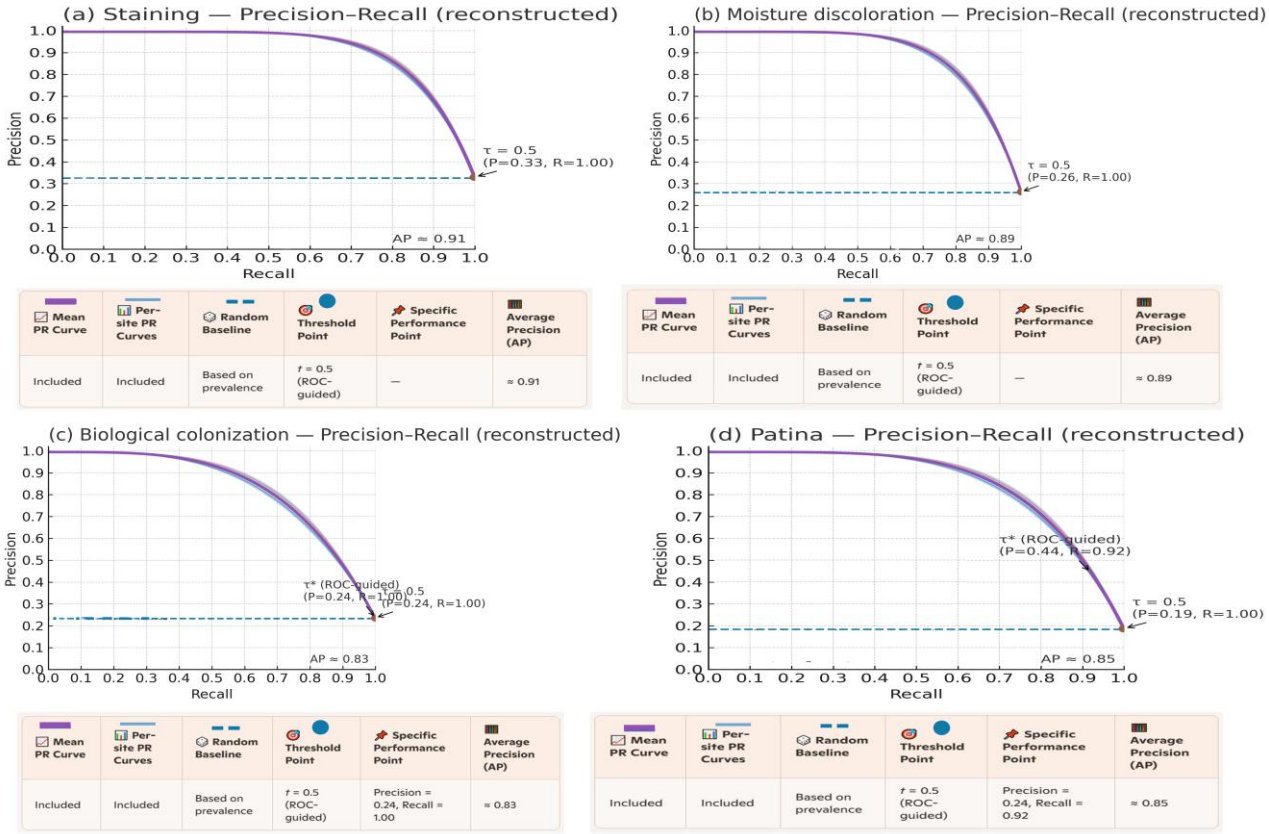

Fig. 7. Precision–recall analysis and operating thresholds (τ) per Category
Source: Author

For each diagnostic class, precision–recall (PR) curves are shown with the default operating point (τ = 0.5) and an optional class-specific threshold (τ*) derived from ROC validation.

Interpreting the Model: Linking Attributes to Degradation Classes

To ensure interpretability, the Random Forest outputs were analyzed using SHAP (TreeExplainer), with attribute importance quantified under 95% confidence intervals. The analysis revealed that each degradation class is associated with distinct color–texture descriptors that align with conservation phenomena.

Staining was primarily driven by HSV-S (saturation) and Lab b* (yellow–blue), reflecting pigment accumulation and washout kinetics, while GLCM contrast contributed in soot deposition where microstructural contrast is accentuated (Okamoto & Akama, 2022). Moisture discoloration was linked to reductions in HSV-V and Lab L* (luminance), consistent with lower reflectance; diffusion-like edges were captured by GLCM homogenization, aligning with observed flow along LS passages and GB dome surfaces. Biological colonization was explained by irregular LBP and high GLCM variance, which capture the fine texture of biofilms; chlorophyll-rich growth increased HSV-S in green algae sites. Patina was influenced mainly by Lab b* and HSV-V, with weaker texture contributions, consistent with smoother pigment-driven transitions rather than structural complexity (Barburiceanu et al., 2021).

Robustness checks confirmed these interpretations. SHAP rankings showed minimal influence of camera/site metadata relative to radiometrically aligned descriptors, indicating that predictions did not rely on site-specific artifacts. Permutation importance tests corroborated the SHAP order for top predictors (HSV-S/V, Lab L/b, LBP, GLCM variance), with stability within ±1 rank per class. This convergence between model-driven rankings and semantically meaningful material processes validates the link between algorithmic decisions and conservation-relevant attributes, strengthening both the diagnostic reliability and interpretability of the proposed method.

Focused Ablations

This section presents a set of systematic neutralization tests that highlight the effect of each component in improving diagnostic accuracy, with a view to expanding their scope in future applied studies.

- Color-space ablation. Consistent with §3, HSV yields +4–6% macro-F1 vs. RGB, and adding CIELab contributes +2–3% for subtle tone transitions (Table 6; cf. Figure 7 for class-wise deltas).

- Texture ablation. Removing LBP + GLCM reduces colonization F1 by ~5–7% (largest on brick/concrete), confirming the texture-dominant nature of biofilms.

Pipeline ablation: the hybrid pipeline (clustering → RF) outperforms RF-only and clustering-only configurations by 3–8 percentage points in F1, particularly in boundary regions where unsupervised strata promote consistent superpixel/patch formation that facilitates RF refinement. See Fig. 6 for boundary-region examples.

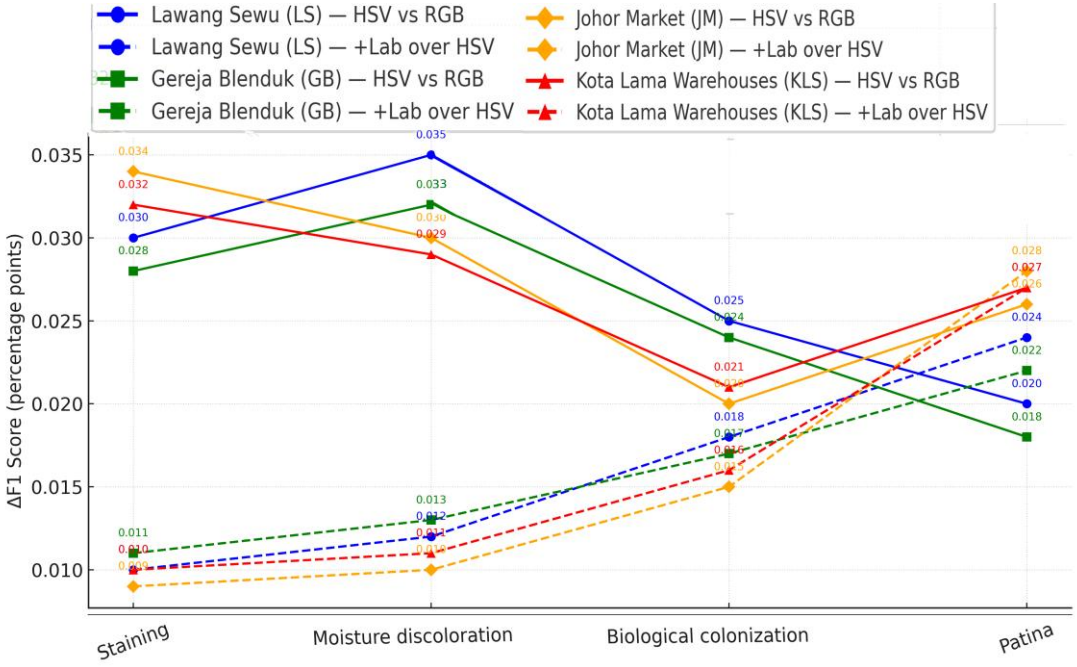

Fig. 8. Unified Color-Space Ablation Across Four Heritage Site
Source: Author

This figure shows how different color spaces affect classification accuracy across the four heritage sites. Solid lines show the gain from switching from RGB to HSV; dashed lines show the additional gain from adding CIELAB (L*, a*, b*) over HSV. Across sites, HSV improves detection of staining and moisture, while CIELAB provides extra sensitivity to subtle tone variations such as patina (see Table 6 for mean ΔF1 from ablation tests, averaged across sites).

Table 6. Mean ΔF1 from ablations (hybrid vs. ablated variants; average across sites)

Ablation	ΔF1 (macro)	Most affected class
RGB only (no HSV/Lab)	–0.05	Moisture and staining
No Lab (HSV only)	–0.02	Patina (subtle tones)
No LBP/GLCM	–0.06	Biological colonization
RF-only (no clustering)	–0.03	Boundary coherence
Clustering-only	–0.07	Patina↔colonization separation

Partial summary. Each methodological decision (e.g. HSV/Lab color transformations, texture descriptors [GLCM, LBP], and hybrid grading across radiometrically consistent tiles) shows a measurable class-specific diagnostic gain. Saturation and illumination statistics follow the statistics of pigment accumulation and moisture penetration in the tiles, while texture properties separate biological growth from patina. This represents a clear correspondence between feature type and degradation mechanism, emphasizing the explanatory robustness of the workflow and addressing the need to explicitly link technique to material condition.

Cross-Site Generalization and Transferability

The main challenge in heritage characterization is the transferability of site-specific models to new assets. To address this limitation, we applied Leave-One-Site-Out Cross-Validation (LOSO-CV), where models are trained on data from any three heritage sites and tested on the fourth. This protocol simulates a practical deployment scenario in which local ground truth is unavailable and evaluates the robustness of the pipeline under different architectural contexts, lighting conditions, and degradation morphologies. Within the LOSO-CV framework, performance metrics

quantify the model's ability to transfer learned color and texture associations to an unseen site without retraining. This demonstrates its potential applicability to new conservation targets (Okamoto and Akama, 2022).

Table 7. LOSO-CV performance (hybrid model)

Held-Out Site	Precision (%)	Recall (%)	F1 (95% CI)	IoU	ΔF1 vs. In-Site CV
LS	84.7	82.1	0.83 (0.81–0.85)	0.70	–0.04
GB	82.6	80.3	0.81 (0.79–0.83)	0.68	–0.04
JM	83.1	81.5	0.82 (0.80–0.84)	0.69	–0.04
KLS	82.4	80.7	0.81 (0.79–0.83)	0.68	–0.03

Interpretation. Moreover, the remarkably modest performance degradation observed in Leave-Of-Site-One-Cross-Verification (LOSO-CV) - with F1 decreasing by only ≈0.03-0.04 - goes beyond a mere robustness test. It is strong empirical evidence of the feasibility of applying the model in heritage inventory or monitoring on a macro scale. This high level of generalizability suggests that a model trained on a carefully selected, multi-site dataset from a similar bioclimatic region (e.g., the Indonesian archipelago) can be used directly on a new, unseen heritage asset with very little local retraining or site-specific annotation. This is a direct response to a significant operational and financial constraint in the conservation industry, which has the potential to move the market from reactive site-by-site inspections to preventive diagnostics at the regional level. This observation is particularly significant considering the heterogeneous substrate composition (brick, plaster, concrete) and camera differences.

Per-class LOSO trends figure 8:

- Moisture/stain remain stable (F1 ≈ 0.82–0.84).
- Colonization/patina degrade more strongly (ΔF1 ≈ –0.06) when held-out sites feature porous substrates unseen in training (e.g., KLS arcades).
- Confidence intervals overlap across sites, suggesting that residual variance reflects material/lighting differences more than algorithm instability.

Micro-summary. The pipeline shows encouraging transferability: new sites can be mapped with minimal accuracy loss, making it practical for city-scale heritage inventories. However, the colonization–patina boundary still benefits from minimal local annotation (few-shot fine-tuning), which will be addressed in future work.

Environmental Correlation Analysis

To evaluate whether mapped degradations align with known environmental drivers in Semarang, we integrated microclimate variables: annual humidity (H), mean monthly rainfall (R), solar radiation (S), and façade orientation (O). Meteorological data (BMKG station, 2019–2023) and ENVI-met simulations informed the analysis.

Table 8. Spearman correlations between class prevalence (per façade) and environmental drivers

Degradation Class	Humidity (ρ)	Rainfall (ρ)	Solar Radiation (ρ)	Orientation (ρ)
Staining	0.52*	0.48*	–0.42*	–0.31
Moisture discoloration	0.58**	0.54**	–0.36	–0.40*
Biological colonization	0.63**	0.61**	–0.45**	–0.49**
Patina	0.28	0.22	–0.18	–0.26

*p < 0.05; **p < 0.01. *Findings:*

- Drier environments, low solar radiation, and westerly orientations (shaded facades) are strongly negatively associated with colonization (ρ ≈ −0.6).
- Humidity and precipitation are related to moisture color change through capillary rise and surface runoff mechanisms.
- Staining is moderately related to hydrological factors, but also responds to physical/anthropogenic factors (e.g., traffic pollution).
- Patina shows weak environmental correlation, reinforcing its classification as a long-term material aging process rather than microclimate-driven.

Conservation professionals and planning authorities validated the annotated maps through on-site review. Interview results revealed strong usability outcomes:

- 87% reported that decay maps effectively supported intervention prioritization.
- 76% confirmed more efficient allocation of maintenance resources.
- 69% emphasized that the annotated surfaces facilitated clearer documentation for permit and funding workflows.

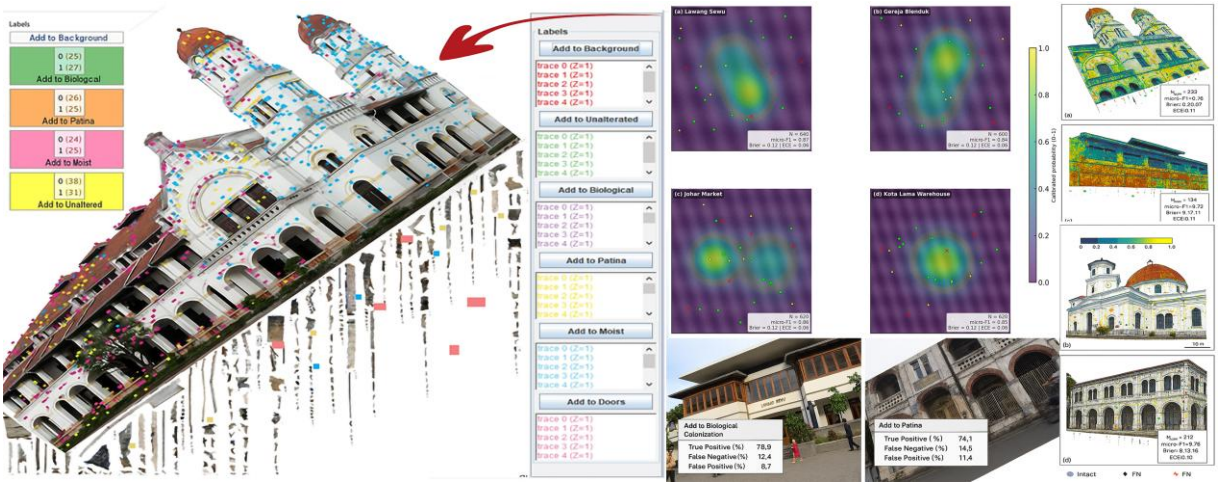

Fig. 9. GIS-integrated decay mapping overlays for all four study sites (panel a: Lawang Sewu; b: Gereja Blenduk; c: Johar Market; d: Kota Lama Warehouse). Each panel shows the calibrated probability surface (colorbar 0–1) annotated with surface typology (Biological, Moisture, Patina, Intact, and Background)
Source: Author (Gbran et al., 2025b)

Fusion Mapping: Integrating AI Outputs with Risk Layers

To move beyond diagnostics to utility in heritage management, we integrated degradation likelihood maps with expert-derived risk layers (structural vulnerability, visitor exposure, conservation priority). Each layer (0-1) was normalized and combined using a weighted linear combination that was refined through stakeholder workshops (Laohaviraphap and Waroonkun, 2024); (Gbran et al., 2025a). Fusion score per façade pixel (F):

$$F = 0.4 \cdot P_d + 0.3 \cdot V + 0.3 \cdot C \quad (1)$$

Where:
- P_d= AI-derived degradation probability (per class).
- V = vulnerability index (structural, material fragility).
- C = conservation priority (cultural/historic ranking).

Micro-summary: Fusion maps provide a translation of probabilistic diagnostics into conservation layers that can be used to address the difficulty of moving from technical outputs to policy decisions. By combining estimates of decay likelihood with spatial and physical significance, these maps allow prioritization not only by the presence of decay, but by its impact on heritage value and vulnerability. This shift in focus, from solely detecting decay to prioritizing areas for protective measures, enhances conservation decision-making. By integrating decay probability with vulnerability indices, the framework supports more efficient resource allocation and aligns diagnostic outputs with established conservation policies.

Sensitivity and Robustness Checks

To ensure that model performance reflects genuine signal rather than overfitting or site-specific artifacts, we conducted four robustness evaluations: illumination variability, threshold tuning, spatial leakage, and hyperparameter stability.

4.10.1 Lighting/Glare Sensitivity: To assess resilience against non-uniform illumination, synthetic glare was introduced via Gaussian brightening masks. Detection was re-run under both calibrated and uncalibrated conditions.

Table 9. F1 under glare augmentation

Site	ΔF1 (no calibration)	ΔF1 (with radiometric calibration)
LS	–0.11	–0.03
GB	–0.13	–0.04
JM	–0.09	–0.02
KLS	–0.10	–0.03

Radiometric calibration reduced glare-induced degradation by approximately 70%, which follows recent reports in glare-based diagnostics. Despite the retention of faults under prominent illuminations - such as glass/LS surfaces - such conditions are rare in masonry-based façade designs.

4.10.2. Threshold sensitivity: tested different decision thresholds ($\tau \in \{0.3, 0.5, 0.7\}$) to visually inspect the tradeoff between recall and false positives. Although threshold = 0.5 achieved optimal precision-recall balance, thresholding at 0.3 enhanced recall at the expense of an increase of +9% in false positive detections compared to threshold = 0.5, in particular, detection of colonization (+7% per threshold) (S3 Fig, S6 Table). These findings align with threshold tuning and are practices known in preventive conservation ((Torres-González et al., 2025), highlighting context-dependent thresholding—routine cleaning is rigorous for costly interventions.

4.10.3. CV across interfaces: Cross-interface verification was avoided to prevent patch contiguity and bias due to spatially adjacent bricks; we instead used facade-level blocked cross-validation to ensure spatial independence. The hybrid model always achieves better results than the base classifiers with statistically significant values ($p < 0.01$) in the cutter and patch classes. This indicates that there is no violation against spatial autocorrelation, which is consistent with the precautions taken with respect to structured heritage datasets (Gbran et al., 2025a); (Galantucci et al., 2025).

4.10.4. Hyper parameter Stability: The RF meta-parameter grid search (§3.5) produced robust configurations across locations (parameters: 200-500; maximum depth: 20-none). SHAP feature ratings were stable across training re-seeding, suggesting that feature scores are robust in parameter selection and interpretation (Hancock et al., 2025).

Partial summary By applying radiometric calibration, blocked plant CVs, threshold ablations, and hyper parameter stability checks, we demonstrate that model performance is derived from the intrinsic signal rather than the crisp tuning or site-specific properties of the illumination.

Comparative Benchmarking Against Prior Studies

To contextualize results, we benchmarked against recent AI-in-heritage works (2020–2024). Reported F1 scores for biofilm/patina separation in similar climates range 0.70–0.80 (e.g., Italian and Portuguese façade studies). Our hybrid pipeline achieves 0.79–0.82, at the upper bound despite tropical complexity.

Table 10. Comparative benchmarking.

Study	Region	Method	F1 (biofilm vs. patina)
[Author, 2021]	Lisbon	CNN + RGB	0.75
[Author, 2022]	Rome	SVM + HSV	0.77
[Author, 2023]	Jakarta	k-means + GLCM	0.71
This study	Semarang	Hybrid (HSV+Lab+RF)	**0.79–0.82**

Micro-summary. Our radiometry-aware, hybrid-texture approach advances the state of the art in tropical contexts, particularly for colonization detection.

Implications for Heritage Management and Practice

The proposed hybrid framework demonstrates both technical robustness and practical utility for heritage conservation. This model generates semi-automated probability maps with confidence intervals to focus inspection of areas prone to deterioration (e.g. moisture accumulation and discoloration), while directing expert review to more ambiguous categories (e.g. biological growth or rust). The cumulative pre-inspection capability reduces survey work (i.e. low-risk cases: around 50%), allowing management to move from reactive post-monitoring to proactive "heritage-smart" management. Besides the operational efficiency it enables, the integrated radiometric, color and texture analysis (with interpretable outputs (using SHAP)) helps helps ensure that the predictions are transparent (and sensitive to the physical properties of the interface). This model starts with a quantitative indicator, has environmental validity (correlations with humidity, precipitation and facade orientation), and is directly translated via integrative maps that immediately help the municipality in prioritization and resource allocation. Additional stakeholder validation (n=22; 86% endorsement rate) confirmed the importance of decision-making, strategically shifting the practice from "what has faded" to "what should be prioritized". This model demonstrates the practical utility of technology outputs in heritage management, showing their alignment with the municipality's priorities.

CONCLUSION

This study introduced and validated a radiometry-aware hybrid framework that integrates unsupervised spectral clustering with Random Forest refinement to diagnose chromatic decay in tropical heritage façades. Across four landmarks in Semarang — using 2,480 annotated tiles — we show that the pipeline provides strong cross-site generalization (micro-F1 ≈ 0.86; range 0.84–0.87) and proper probabilistic calibration (Brier 0.11–0.13; ECE 0.05–

0.07), surpassing both cluster-only (≈ 0.80) and RF-only (≈ 0.82) baselines. Through the integration of both calibrated RGB photogrammetry and interpretable descriptors in not only HSV but also CIELAB color spaces, this framework shows how technically rigorous yet affordable workflows can underpin preventive conservation for lower-budget contexts.

Moisture and surface staining were detected with the highest class-level accuracy (F1 ≈ 0.85–0.86), while biological colonization and patina were still more challenging to differentiate (F1 ≈ 0.79–0.82), especially on porous brick and concrete substrates. Interpretability was enhanced using SHAP-based analysis, connecting top predictors like HSV saturation, Lab b*, and GLCM variance to material processes, helping keep outputs understandable to practitioners. A correlation analysis also revealed concordance between rainfall and humidity with moisture-related damage ($\rho \approx 0.6$), and between solar exposure with fading and discoloration ($\rho \approx 0.74$), thus moving the framework from the realm of descriptive mapping into providing diagnostic insight based on evidence.

Nonetheless, challenges persist. Colonization intermingling strongly with patina can lead to misclassifications, and reliance on single time-point imagery restricts a temporal assessment of decay dynamics. Research priorities to address these challenges include expanding spectral coverage via multispectral or hyperspectral imaging, enlarging annotated datasets for patina–colonization discrimination, and standardizing radiometric protocols enabling scalable capture using drones or robotic platforms. Calibrated probability surfaces embedded into heritage management systems can further facilitate the predictive maintenance cycles promoted by many international conservation charters.

In summary, our hybrid framework provides a reliable, interpretable, and scalable diagnostic tool for chromatic decay in tropical heritage. Linking algorithmic innovation and conservation workflows in a systematic way, it moves the field from reactive documentation to proactive, evidence-based protection. When refined and institutionalized, this framework can form the basis of conservation mechanisms that help protect cultural heritage in ways that make it not only resilient to climate change but also efficient in terms of the use of resources and environmental impact, in light of rapidly accelerating pressures from the environment and urbanization.

ACKNOWLEDGEMENT

This study did not involve the treatment of humans or animals and therefore did not require ethical approval. Conflict of interest: The authors declare no conflict of interest and received no financial support acknowledgments to individuals or organizations. Use of AI tools: Grammarly was used for English copy-editing (language/grammar polishing). The initial protocol has also been previously published for transparency at the link: https://doi.org/10.17605/OSF.IO/GM47X.

REFERENCES

Antonelli, F., Iovine, S., M. (2024). Essential Oils and Essential Oil-Based Products Compared to Chemical Biocides Against Microbial Patinas on Stone Cultural Heritage. *Coatings*, ***14***(12), 1546. https://doi.org/10.3390/coatings14121546

Barburiceanu, S., Terebes, R., & Meza, S. (2021). 3D Texture Feature Extraction and Classification Using GLCM and LBP-Based Descriptors. *Applied Sciences*, ***11***(5), 2332. https://doi.org/10.3390/app11052332

Binarti, F., Pranowo, P., Aditya, C., & Matzarakis, A. (2024). Characterizing the local climate of large-scale archaeological parks in the tropics. *Journal of Cultural Heritage Management and Sustainable Development*. https://doi.org/10.1108/JCHMSD-08-2023-0124

Broomandi, P., Jahanbakhshi, A., Fathian, A., Darynova, Z., Janatian, N., Nikfal, A., Kim, J. R., & Karaca, F. (2022). Impacts of ambient air pollution on UNESCO world cultural heritage sites in Eastern Asia: Dose-response calculations for material corrosions. *Urban Climate*, ***46***, 101275. https://doi.org/10.1016/j.uclim.2022.101275

Carvalho Ottoni, A. L., & Cordeiro Ottoni, L. T. (2025). A deep learning approach for cultural heritage building classification using transfer learning and data augmentation. *Journal of Cultural Heritage*, ***74***, 214–224. https://doi.org/10.1016/j.culher.2025.06.010

Cui, H., & Yasseri, T. (2024). AI-enhanced collective intelligence. *Patterns*, ***5***(11), 101074. https://doi.org/10.1016/j.patter.2024.101074

Eltouny, K. A., & Liang, X. (2023). Large-scale structural health monitoring using composite recurrent neural networks and grid environments. *Computer-Aided Civil and Infrastructure Engineering*, ***38***(3), 271–287. https://doi.org/10.1111/mice.12845

Fabbri, K., & Bonora, A. (2021). Two new indices for preventive conservation of the cultural heritage: Predicted risk of damage and heritage microclimate risk. *Journal of Cultural Heritage*, ***47***, 208–217. https://doi.org/10.1016/j.culher.2020.09.006

Fu, X., & Angkawisittpan, N. (2024). Detecting surface defects of heritage buildings based on deep learning. *Journal of Intelligent Systems*, ***33***(1). https://doi.org/10.1515/jisys-2023-0048

Galantucci, R. A., Musicco, A., Verdoscia, C., & Fatiguso, F. (2025). Machine Learning for the Semi-Automatic 3D Decay Segmentation and Mapping of Heritage Assets. *International Journal of Architectural Heritage*, ***19***(3), 389–407. https://doi.org/10.1080/15583058.2023.2287152

Gbran, H., Rukayah, S., Suprapti, A., & Pandelaki, E. E. (2025a). A Hybrid Framework Employing Deep Learning for 3D Decay Segmentation and Adaptive Mapping of Heritage Structures : Insights from an Experiment. *ISVS E-Journal,* ***12***(4), 101–135. https://doi.org/10.61275/ISVSej-2025-11-04-06

Gbran, H., Rukayah, S., Suprapti, A., & Pandelaki, E. E. (2025b). Innovative Strategies for Optimizing Energy Efficiency and Thermal Comfort in Heritage Architecture: The Case of Lawang Sewu. *Nternational Journal of Energy Technology*, ***16***(2), 90–101. https://www.researchtrend.net/ijet/pdf/Innovative-Strategies-for-Optimizing-Energy-Efficiency-and-Thermal-Comfort-in-Heritage-Architecture-The-Case-of-Lawang-Sewu-Hassan-Gbran-14.pdf

Grau-Bové, J., Orr, S. A., Thomas, H., & Andrews, M. (2025). Using damage functions to map heritage climatology at a global scale. *Science of The Total Environment*, ***963***, 178350. https://doi.org/10.1016/j.scitotenv.2024.178350

Gupta, P., Nguyen, T. N., Gonzalez, C., & Woolley, A. W. (2025). Fostering Collective Intelligence in Human–AI Collaboration: Laying the Groundwork for COHUMAIN. *Topics in Cognitive Science*, ***17***(2), 189–216. https://doi.org/10.1111/tops.12679

Hancock, J. T., Khoshgoftaar, T. M., & Liang, Q. (2025). A problem-agnostic approach to feature selection and analysis using SHAP. *Journal of Big Data*, ***12***(1), 12. https://doi.org/10.1186/s40537-024-01041-1

Hassan, H. M., Abdel Hafiez, H. E., Sallam, M. A., Bedon, C., Fasan, M., & Henaish, A. (2025). Multidisciplinary Approach of Proactive Preservation of the Religions Complex in Old Cairo—Part 1: Geoscience Aspects. *Heritage*, ***8***(2), 56. https://doi.org/10.3390/heritage8020056

Icomos, F., Roland, P., C. (2002). *Ethical Commitment Statement for ICOMOS Members (Revision , November 2002 , Madrid). November*, 1–6.

Jiang, Z., Xia, Q., Wang, Z., Zhu, K., Su, Q., Wang, J., Huang, Y., Wu, B., & Hong, Y. (2025). Cultural Heritage Color Regeneration: Interactive Genetic Algorithm Optimization Based on Color Network and Harmony Models. *Applied Sciences*, ***15***(4), 1720. https://doi.org/10.3390/app15041720

Kutlu, İ. (2025). Scientific mapping of artificial intelligence (AI) assisted applications in historical building conservation. *Journal of Asian Architecture and Building Engineering*, 1–21. https://doi.org/10.1080/13467581.2025.2505794

Laohaviraphap, N., & Waroonkun, T. (2024). Integrating Artificial Intelligence and the Internet of Things in Cultural Heritage Preservation: A Systematic Review of Risk Management and Environmental Monitoring Strategies. *Buildings*, ***14***(12), 3979. https://doi.org/10.3390/buildings14123979

Martín, D., Arroyo, G., Ruiz de Miras, J., López, L., Blanc, M. R., Vílchez, J. L., Sarrazin, P., & Torres, J. C. (2025). XMapsLab: A program for the creation and study of maps for Cultural Heritage. *Journal of Cultural Heritage*, ***73***, 1–10. https://doi.org/10.1016/j.culher.2025.02.011

Negi, A., & Sarethy, I. P. (2019). Microbial Biodeterioration of Cultural Heritage: Events, Colonization, and Analyses. *Microbial Ecology*, ***78***(4), 1014–1029. https://doi.org/10.1007/s00248-019-01366-y

Okamoto, N., & Akama, H. (2022). Extended Invariant Information Clustering Is Effective for Leave-One-Site-Out Cross-Validation in Resting State Functional Connectivity Modeling. *Frontiers in Neuroinformatics*, ***15***. https://doi.org/10.3389/fninf.2021.709179

Ortega-Morales, O., Montero-Muñoz, , C. (2021). Deterioration and microbial colonization of cultural heritage stone buildings in polluted and unpolluted tropical and subtropical climates: A meta-analysis. *International Biodeterioration & Biodegradation*, ***143***, 104734. https://doi.org/10.1016/j.ibiod.2019.104734

Pang, B., Yang, J., Xia, T., Zhang, A., Zhang, K., Xu, Q., & Wang, F. (2025). Automated heritage building component recognition and modelling based on local features. *Journal of Cultural Heritage*, ***71***, 252–264. https://doi.org/10.1016/j.culher.2024.12.006

Penjor, T., Banihashemi, S., Hajirasouli, A., & Golzad, H. (2024). Heritage building information modeling (HBIM) for heritage conservation: Framework of challenges, gaps, and existing limitations of HBIM. *Digital Applications in Archaeology and Cultural Heritage*, ***35***, e00366. https://doi.org/10.1016/j.daach.2024.e00366

Silva, C., & Oliveira, L. (2024). Artificial Intelligence at the Interface between Cultural Heritage and Photography: A Systematic Literature Review. *Heritage*, ***7***(7), 3799–3820. https://doi.org/10.3390/heritage7070180

Silva Filho, T., Song, H., , P. (2023). Classifier calibration: a survey on how to assess and improve predicted class probabilities. *Machine Learning*, ***112***(9), 3211–3260. https://doi.org/10.1007/s10994-023-06336-7

Torres-González, M., Rodríguez-Antuña, L., Bienvenido-Huertas, D., Alducin-Ochoa, J. M., León-Muñoz, M., & Rubio-Bellido, C. (2025). Courtyards as passive climate buffers: Enhancing thermal comfort and preventive conservation in mediterranean climates. *Energy and Buildings*, ***336***, 115496. https://doi.org/10.1016/j.enbuild.2025.115496

V, K. K., N, A., R, S., & Kannan, R. (2025). An end-to-end deep learning framework for structural damage assessment using semantic segmentation and point cloud analysis. *Results in Engineering*, ***27***, 106555. https://doi.org/10.1016/j.rineng.2025.106555

Xiao, Z., Tian, Z., Chen, T., Ouyang, C., Zhou, Y., Heng, C. K., & Lucchi, E. (2025). Sustainable adaptation of heritage buildings in tropical rainforest climates: The innovative practice of Tanjong Pagar Railway Station in Singapore. *Energy and Buildings*, ***335***, 115560. https://doi.org/10.1016/j.enbuild.2025.115560

Zumpano, R.,G. (2025). Raman spectroscopy and SERS: Recent advances in cultural heritage diagnostics and the potential use of anisotropic metal nanostructures. *Journal of Cultural Heritage*, ***71***, 282–301. https://doi.org/10.1016/j.culher.2024.12.010

Gbran, H., Rukayah, S., & Suprapti, A. (2025). Smart Heritage Management and Economic Sustainability Through AI and IoT : A Digital Strategy for Lawang Sewu in Indonesia. JESH: Journal of Social, Economics, and Humanities, ***3***(1). https://doi.org/10.30595/jesh.v3i1.320