Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Hybrid AI-IoT Framework for Predictive Conservation of Tropical Heritage Façades: Empirical Validation from a Multi-Site Study in Indonesia

HASSAN GBRAN

gbranhassan882@gmail.com

Siti Rukayah
Atik Suprapti
Erni Setyowati
Iwan Setiawan

Article

Keywords: Chromatic deterioration, Early warning system, IoT sensing, Random Forest, Hierarchical clustering, Tropical heritage, 3D photogrammetry, Predictive monitoring

Posted Date: July 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10327340/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on November 5th, 2026. See the published version at https://doi.org/10.1038/s41598-026-74267-z.

Abstract

Tropical heritage façades in coastal urban environments face accelerated chromatic deterioration driven by persistent high humidity, temperature fluctuations, and salt-laden air. Current conservation practice remains predominantly reactive, while environmental monitoring systems operate largely in isolation from computational diagnostic pipelines. This study develops and empirically validates a hybrid AI-IoT framework that integrates continuous environmental sensing with multi-stage machine learning analysis for early warning of chromatic deterioration. The framework combines unsupervised hierarchical clustering in HSV colour space with supervised Random Forest (RF) classification, incorporating microclimatic variables—relative humidity, temperature, and a derived Humidity Stress Factor (HSF)—as predictive features rather than post-hoc explanatory factors. Validation was conducted on three heritage sites in Semarang, Indonesia: Lawang Sewu, Pasar Johar, and Tawang Railway Station, using 48,830 expert-labelled pixels and 3,750 annotated tiles from radiometrically corrected photogrammetric models. The RF classifier achieved a mean macro-F1 score of 0.87 (range: 0.84–0.90) and overall accuracy exceeding 90%. Hierarchical clustering yielded stable partitions with silhouette scores averaging 0.75 (SD = 0.02). The integrated workflow improved spatial Intersection-over-Union (IoU) by 17–51% compared to unsupervised clustering alone. A rigorous ablation study confirmed that the HSF contributes a 9.2% absolute increase in macro-F1. Leave-one-site-out cross-validation yielded F1 scores of 0.79–0.82, demonstrating acceptable cross-site generalisability. Mixed-effects modelling confirmed a significant positive association between cumulative hours of relative humidity exceeding 85% and material deterioration ($\beta = 0.18$, 95% CI: 0.13–0.23, $p < 0.001$ after Bonferroni correction). On retrospective test data, the framework identified 79% of impending high-risk events (recall = 0.79), providing a mean technical lead time of 4.8 to 16.7 hours conditional upon correct detection. When accounting for missed events, the expected lead time ranged from 3.8 to 13.2 hours. The framework is presented as a methodologically reproducible blueprint requiring local recalibration of thresholds and model parameters for transfer to other tropical contexts. A complete reproducibility package including data, code, and calibration artifacts is provided.

1. Introduction

Heritage buildings situated within humid tropical climates face persistent environmental pressures that accelerate material deterioration. High relative humidity—frequently exceeding 80% for extended periods—combined with elevated temperatures and salt-laden coastal air, drives multiple decay mechanisms including biological colonisation, moisture staining, salt crystallisation, and progressive pigment attenuation (Warscheid and Braams, 2000; Sesana et al., 2019; Doehne and Price, 2010). These chromatic alterations constitute some of the earliest and most diagnostically sensitive indicators of material degradation, typically preceding mechanical weakening or chemical deterioration by a considerable margin (Sesana et al., 2021). Their timely detection is therefore critical for enabling preventive conservation strategies and mitigating the need for costly restorative interventions (Lucchi, 2023).

Conventional inspection techniques—ranging from manual visual surveys to standard two-dimensional photographic documentation—provide essential descriptive insights but are inherently limited in radiometric sensitivity and spatial completeness. These methods depend substantially on subjective interpretation and cannot reliably capture low-contrast chromatic anomalies or their spatial distribution on architecturally complex surfaces, particularly under variable tropical lighting conditions (Yang et al., 2024). Although reality-based 3D digitisation has advanced considerably, offering highly accurate geometric reconstructions of built heritage (Remondino, 2011), the diagnostic potential embedded within radiometrically corrected texture maps remains underexploited. As noted by Ferdani et al. (2025), high-resolution photogrammetric models possess the spectral richness necessary for detailed colour interpretation, yet most analytical workflows still privilege geometric cues over chromatic ones, reducing their effectiveness for early-warning diagnostics.

Recent studies have demonstrated that perceptual and radiometric colour descriptors provide stronger discriminative power than geometry-only methods for identifying surface decay, especially in heritage sites where chromatic behaviour is an early indicator of deterioration (Tan et al., 2013; Ferdani et al., 2025). Complementary advances in computational interpretation further reinforce this potential. Unsupervised clustering techniques—particularly hierarchical clustering implemented within perceptually stable colour spaces such as HSV—have shown strong capability in revealing latent chromatic groupings without reliance on annotated datasets, making them well suited to field environments with limited labelled samples (Musicco et al., 2021). Ensemble machine learning classifiers, particularly Random Forests, have also demonstrated robustness to illumination variability, heterogeneous surface textures, and moderate amounts of noise while preserving interpretability through feature-importance structures (Zhu et al., 2017; Colace et al., 2025). These properties are especially relevant in heritage conservation, where decision-making must remain transparent and explainable, and where large annotated datasets—typically required for deep learning—are rarely available (Fiorucci et al., 2020). Beyond diagnostics, recent work has also demonstrated the feasibility of integrating energy-efficient retrofits, such as electrochromic glazing and building-integrated photovoltaics, in tropical heritage buildings without compromising historic integrity (Gbran, Rukayah, & Suprapti, 2026). In parallel, hybrid geometric–radiometric descriptors have been developed for structural surface anomaly detection, integrating intensity information with geometric features to improve the detection of moisture-related anomalies in tropical heritage structures (Gbran & Murtiono, 2026). At the broader systems level, the convergence of AI and IoT has been shown to improve operational decision-making, enable early detection of structural and environmental risks, and optimise resource use in resource-limited tropical heritage settings (Gbran, Rukayah, Suprapti, & Pandelaki, 2025). Furthermore, radiometrically normalised photogrammetric texture atlases combined with multi-space colour and texture descriptors have proven effective for diagnostic mapping of chromatic decay, enabling prioritised inspections and evidence-based conservation (Gbran, Rukayah, Suprapti, & Pandelaki, 2025b).

Simultaneously, the Internet of Things (IoT) has emerged as a transformative technology for environmental monitoring in heritage contexts, enabling continuous, real-time data collection at spatial and temporal resolutions impossible with manual methods (Gubbi et al., 2013; Bibri et al., 2024). Low-cost sensor networks can now record temperature, relative humidity, and surface moisture at intervals as

short as 15 minutes over extended periods, generating high-resolution time-series datasets suitable for predictive modelling (Mitro et al., 2022). However, the integration of IoT sensing with AI-driven analytics in heritage conservation remains nascent. Most published applications either isolate climate sensing from predictive analytics, apply AI models to pre-collected datasets without real-time IoT integration, or develop technical systems without incorporating architectural knowledge or conservation ethics (Al-Raeei, 2024; Bibri et al., 2025). In tropical heritage contexts specifically, integration challenges are compounded by resource constraints, limited technical capacity, and the need for approaches that respect local knowledge systems (Laohaviraphap and Waroonkun, 2024).

A persistent methodological limitation in the literature concerns the treatment of environmental data. Although variables such as relative humidity, surface temperature, and solar exposure are widely recognised as influential drivers of chromatic decay, their integration into computational workflows is often minimal. Most studies document environmental factors qualitatively or use them as post-hoc explanatory variables rather than as predictive inputs that directly inform the classification model (Sánchez-Aparicio et al., 2018; Charisi et al., 2018). This disconnect between environmental monitoring and chromatic analysis limits the ability of practitioners to interpret spatial decay patterns in relation to climatic pressures characteristic of tropical settings. Furthermore, humidity thresholds used in deterioration risk classification are frequently imported from temperate-climate building science without local calibration, despite evidence that tropical conditions require adjusted thresholds (Sesana et al., 2024; Rukayah et al., 2023).

A related concern is the subjectivity inherent in translating purely chromatic clusters into semantically meaningful decay categories. The mapping from colour-based groupings to categories such as 'biological patina', 'moisture-related staining', or 'salt efflorescence' typically relies on expert interpretation, yet this process is rarely standardised or validated (Galantucci et al., 2023). This introduces a level of subjectivity that should be explicitly acknowledged and, where possible, mitigated through structured protocols or multi-expert consensus.

Taken together, these challenges reveal a clear operational gap. Existing approaches either (i) emphasise geometric descriptors at the expense of chromatic sensitivity; (ii) require annotated datasets that are impractical to obtain in field conditions; (iii) use radiometric analyses that lack sufficient semantic clarity for conservation practice; or (iv) treat environmental parameters as ancillary commentary rather than integral model inputs. Consequently, there is a pressing need for a diagnostic framework capable of: (1) exploiting the radiometric depth of high-resolution texture-mapped 3D models; (2) operating reliably with minimal annotated samples; (3) producing transparent and interpretable outputs; (4) integrating environmental information as predictive features rather than post-hoc explanations; and (5) providing quantifiable early warning lead times to support conservation decision-making.

The present study addresses these gaps by introducing a hybrid AI-IoT framework specifically designed for predictive monitoring of chromatic deterioration on tropical heritage façades. The framework integrates continuous environmental sensing as a predictive layer directly into the classification model—moving beyond post-hoc explanatory analysis to enable genuine early warning capability. We validate the

framework on three heritage sites in Semarang, Indonesia, representing diverse architectural typologies and material conditions: Lawang Sewu (colonial administrative building), Pasar Johar (traditional market), and Tawang Railway Station (colonial transport infrastructure). The framework provides quantifiable early warning lead times and a structured decision-support matrix that combines environmental risk with heritage significance. The study explicitly acknowledges its contextual limitations; the framework is presented as a methodologically reproducible blueprint requiring local recalibration of thresholds and model parameters for transfer to other tropical contexts.

2. Materials and Methods

2.1 Study Sites and Data Acquisition

The study was conducted in Semarang, Central Java, Indonesia (6°58'S, 110°25'E), a coastal city with a hot-humid tropical climate classified as Köppen Af (tropical rainforest climate). Annual rainfall exceeds 2,500 mm, with the wettest months occurring between November and March. Relative humidity averages 75–85% year-round, rarely dropping below 70% even during the driest months. Temperatures range from 25–34°C, with daily maxima frequently exceeding 32°C (BMKG, 2026).

Three heritage sites were selected to represent diverse architectural typologies, material systems, and exposure conditions, following purposive sampling criteria based on historical significance, environmental vulnerability, and practical feasibility (Patton, 2015). The geographic distribution and architectural characteristics of these three sites are illustrated in Fig. 1.

Lawang Sewu (built 1904) is a colonial administrative building constructed by the Dutch East Indies Railway Company. The building features substantial masonry walls (50–70 cm thick), deep verandas on all sides, twin ventilation towers that harness passive stack effect airflow, high ceilings (5–6 m), and large operable windows with louvres—a hybrid Indo-European typology adapted to tropical conditions (Rukayah and Abdullah, 2021). The floor area is approximately 2,800 m² spread over two main floors plus a basement and roof space. Despite its sophisticated climatic adaptations, the building exhibits notable material deterioration including rising damp affecting walls up to 1.5 m above ground level, extensive salt efflorescence (predominantly sodium chloride and sodium sulphate) on brick and plaster surfaces, widespread fungal growth in shaded areas with limited airflow, progressive masonry cracking, and timber decay in roof structures and window frames (Kurniawati et al., 2025). The building is located near flood-prone drainage corridors and is subject to recurrent tidal flooding, which exacerbates moisture ingress and salt transport.

Pasar Johar (built 1930s) is a historic traditional market located in Semarang's Old City (Kota Lama). The building is constructed with mixed masonry (brick and concrete) and timber, with large open spaces, high ceilings, and extensive window openings for natural ventilation. It represents a commercial typology with high occupancy loads (hundreds of vendors and daily visitors), irregular operating hours, and significant microclimatic variation between indoor stalls and semi-open corridors. The building has undergone

several restorations, with the most recent completed in 2018, but still exhibits material deterioration due to high humidity and biological growth.

Tawang Railway Station (Stasiun Tawang, built 1914) is a colonial transport hub designed by the Dutch architect J. de Jongh. The building combines reinforced concrete structure with brick masonry infill, featuring large waiting halls, ticketing areas, and platform canopies. It is frequently exposed to seasonal flooding (the station is located near the Tawang River and the northern coast) and thermal stress (large glass windows and concrete surfaces absorb and radiate heat). The building remains in active use as a railway station, which imposes practical constraints on sensor placement but also provides opportunities to study the interaction between occupancy patterns and indoor environmental conditions.

For each site, high-resolution images were captured using a calibrated Nikon D5600 camera (24 megapixel, RAW format) with a 20 mm lens. Acquisition followed a standardised protocol: inclusion of a ColorChecker reference target, forward and lateral image overlap of 75–85%, fixed focal length per façade, and exposure-bracketing sequences to mitigate temporal luminance variation. All EXIF metadata were preserved. A total of 1,200 images were captured across the three sites, with 450 images from Lawang Sewu, 350 from Pasar Johar, and 400 from Tawang Railway Station. The rationale for these image counts was to achieve a consistent ground sampling distance (GSD) of < 2.5 mm/pixel given the varying façade areas, ensuring adequate resolution for sub-millimetre chromatic feature extraction.

The preprocessing pipeline comprised six steps, as summarised in Fig. 2. First, a Colour Correction Matrix (CCM) was estimated per session using the ColorChecker reference to convert camera RGB to a standardised working space (sRGB D65), correcting for illumination and camera spectral response. Second, global harmonisation included white-balance normalisation and exposure levelling across input images. Third, denoising was applied with edge-preserving bilateral filtering; where sensor noise persisted, a Noise2Void routine was used with conservative parameters to avoid chromatic smoothing (Khan et al., 2024). Fourth, dense reconstruction used Agisoft Metashape (version 2.1) to generate dense point clouds and watertight meshes; UV unwrapping and texture baking were executed with seam artifact compensation. Fifth, geometric refinement in CloudCompare targeted outlier removal, decimation for consistent point density, and minor spatial corrections. Sixth, final UV textures and meshes were exported at publication resolution and archived in TIFF/PNG and OBJ/PLY formats, respectively. Ground Sampling Distance (GSD) values ranged from 0.8–2.4 mm/pixel, with reprojection errors between 0.55 and 2.0 pixels, meeting established CIPA and ISPRS quality benchmarks for heritage documentation (Sánchez-Aparicio et al., 2018). Table 1 summarises site characteristics and acquisition specifications.

Table 1
Site characteristics and data acquisition specifications.

Site	Year Built	Typology	Materials	Floor Area (m²)	Images Captured	GSD (mm/pixel)	Reprojection Error (pixels)
Lawang Sewu	1904	Colonial Administrative	Brick, plaster, timber, stone	2,800	450	0.8–1.2	0.55–0.75
Pasar Johar	1930s	Traditional Market	Brick, timber, concrete	3,200	350	1.2–2.0	0.65–1.10
Tawang Station	1914	Transport Infrastructure	Reinforced concrete, brick	4,500	400	1.5–2.4	0.80–2.00

2.2 Environmental IoT Sensing Layer

A network of sixteen sensor nodes was deployed across the three study sites (eight nodes at Lawang Sewu, four at Pasar Johar, and four at Tawang Railway Station). Each node utilised a DHT22 sensor (accuracy: ±0.5°C, ± 2–5% RH), selected for its low cost and adequate precision for relative trend monitoring. Strategic placement followed architectural logic—ventilation pathways, thermal gradients, plinth zones, and exterior-interior interfaces—rather than arbitrary grids. All 16 deployed nodes provided a complete 14-month dataset (March 2025 – April 2026). The complete dataset comprises 517,632 observations at 15-minute intervals (approximately 32,352 readings per node). Data completeness averaged 98.4% after quality control, with sensor drift correction applied through the systematic protocol described below. The deployment configuration is presented in Fig. 3. While Google Sheets was used as a lightweight cloud backend for this pilot implementation, the system architecture is database-agnostic and readily transferable to enterprise-grade solutions (e.g., InfluxDB, TimescaleDB).

A systematic drift correction protocol was implemented to address sensor drift under persistent high-humidity conditions. At each monthly maintenance visit, every node was co-located with a laboratory-grade SHT35 reference sensor (accuracy: ±0.2°C, ± 1.5% RH) for 30 minutes. Drift was calculated as the change in correction factors since the previous calibration. A linear drift model was fitted to the monthly measurements, enabling interpolation of correction factors between calibration points (justification and sensitivity analysis provided in Appendix A). Sensors exhibiting drift exceeding 0.5°C or 5% RH per month were flagged for replacement. Over the 14-month campaign, three sensors (18.75% of the 16 deployed) required replacement; all drift records were documented.

Data were transmitted via Wi-Fi to a cloud-based storage system (Google Sheets API via HTTP POST requests), with local buffering in ESP32 flash memory (up to 4,000 readings) to manage transmission failures. During monthly maintenance, buffered data were mirrored to USB drives as an additional backup. Missing data (1.6% of observations) were handled using linear interpolation between adjacent valid measurements, a method appropriate for the high-frequency time-series data where gaps typically

spanned fewer than six consecutive 15-minute intervals. Analysis of missing data patterns is provided in Appendix B.

2.3 Feature Extraction, Preprocessing, and Probability Calibration

UV textures were converted from RGB to Hue-Saturation-Value (HSV) colour space to decouple chromatic channels (H, S) from luminance (V). Prior to conversion, images were linearised using the session-specific Colour Correction Matrix. A ΔE (CIE76) normalisation step was applied to harmonise chromatic distances across sessions; while CIEDE2000 offers improved perceptual uniformity, CIE76 was selected for computational efficiency and compatibility with prior heritage imaging studies, with the caveat noted in the limitations.

UV textures were decomposed into overlapping square tiles of 32×32, 64×64, and 128×128 pixels with 50% overlap between adjacent tiles to capture multi-scale chromatic patterns while maintaining spatial continuity. For each tile, a set of radiometric and textural descriptors was computed, including first- and second-order colour statistics (mean, variance per channel), Gray Level Co-occurrence Matrix features (contrast, entropy, homogeneity) (Haralick et al., 1973), Local Binary Patterns histograms (Ojala et al., 2002), gradient magnitude statistics, and local hue coherence indices.

Environmental features were engineered from the IoT sensor data to capture temporal dynamics. These included: current temperature and relative humidity; lagged values at 1, 3, 6, 12, and 24 hours; rolling statistics (mean and standard deviation over 6, 12, and 24-hour windows); rate of change (ΔRH/Δt, ΔT/Δt); and cyclic encodings for time of day and day of year. This feature set was designed to capture both short-term dependencies and medium-term trends relevant to deterioration processes.

Prior to model training, Variance Inflation Factor (VIF) testing was conducted to detect multicollinearity among environmental features. Features with VIF > 5 were combined using Principal Component Analysis (PCA) into a single composite feature—'Humidity Stress Factor' (HSF)—which integrates current RH, cumulative exposure to RH > 85% over 24 hours, and RH lagged at 6, 12, and 24 hours. The PCA retained the first principal component, which explained 78.4% of the variance in the combined features. The selection of 6, 12, and 24-hour lags was based on the autocorrelation decay time of the humidity signal (correlation dropping below 0.5 after 28 hours), ensuring the HSF captures the full memory effect of the microclimate. Component loadings and physical interpretation are detailed in Appendix C.

To ensure the reliability of the downstream decision-support scores (Risk_Score), the raw probability outputs from the Random Forest classifier were calibrated using Platt scaling (Platt, 1999). This sigmoidal transformation maps the RF scores to well-calibrated probabilities, minimising the expected calibration error (ECE), which was reduced from 0.12 to 0.03 post-calibration.

2.4 Multi-Stage Machine Learning Pipeline

2.4.1 Unsupervised Hierarchical Clustering

Unsupervised grouping was performed using agglomerative hierarchical clustering with Ward linkage and Euclidean distance, applied to HSV-derived chromatic descriptors supplemented with low-frequency spatial coordinates (weighted at 0.2 relative to chromatic features) to preserve local coherence. Sampling density and tile selection were stratified across façade regions to minimise spatial bias. The number of clusters was identified through silhouette score maximisation, with the optimal cluster count determined as the value yielding the highest average silhouette score across the range 2–15 clusters. Cluster stability was assessed through 1,000 bootstrap resampling iterations, with stability defined as the proportion of samples maintaining the same cluster assignment across bootstrap replicates.

The unsupervised results were produced as chromatic-group maps (diagnostic priors) accompanied by zoomable insets for critical façade regions. Each map was exported both as a georeferenced raster layer and as point-level labels projected onto the reconstructed 3D mesh. The resulting cluster maps for the three study sites are presented in Fig. 4. Formal cluster definitions and their diagnostic interpretations (e.g., 'biological patina', 'moisture-saturated zone') were established through expert consultation, with the recognition that this mapping introduces a degree of subjectivity (addressed in Section 4.3).

2.4.2 Supervised Random Forest Classification and Model Variants

Supervised refinement employed a Random Forest (RF) ensemble implemented in scikit-learn (version 1.3). Training labels were produced by three domain experts (each with 8–20 years of experience in heritage conservation and materials science, affiliated with Indonesian universities and not including the authors) annotating tiles into five decay categories: biofilm/algae, moisture darkening, efflorescence, pigment attenuation, and neutral (no visible deterioration). A total of 48,830 expert-labelled pixels and 3,750 annotated tiles were used for training and validation. Inter-expert agreement for initial labeling is reported in Appendix D.

The hierarchical clustering outputs were incorporated as categorical priors (chromatic group labels) in the Random Forest feature set, alongside the radiometric, textural, and environmental features. Crucially, to prevent data leakage, the hierarchical clustering was performed independently on the training set only (60% of data), and the resulting cluster labels were then applied to the validation and test sets using a k-nearest neighbours mapping based on cluster centroids. Environmental features (current and lagged RH, temperature, and derived variables, including the HSF composite) were included as numeric predictors. The complete data flow and module architecture of this multi-stage pipeline is visualised in Fig. 5.

To rigorously evaluate the contribution of each component, four model variants were constructed:

1. **Variant A (Baseline)**: Radiometric and textural features only.
2. **Variant B**: Radiometric + textural + raw environmental features (without HSF).

3. **Variant C (Full)**: Radiometric + textural + environmental features **with HSF**.
4. **Variant D**: Full model without clustering priors.

The Random Forest hyperparameters were optimised via grid search with 5-fold stratified cross-validation on the training set (60% of data points, March–October 2025), with validation on 20% (November–December 2025) and final testing on 20% (January–April 2026). The hyperparameter grid included n_estimators ∈ {50, 100, 200}, max_depth ∈ {10, 15, 20, 25}, and min_samples_split ∈ {2, 5, 10}. The optimal configuration comprised 100 estimators, maximum depth of 20, minimum samples split of 5, and class_weight='balanced' to address class imbalance. SMOTE (Synthetic Minority Over-sampling Technique) (Chawla et al., 2002) was applied to the training data only (after data split to prevent leakage). Class distribution before and after SMOTE is provided in Appendix E.

2.4.3 3D Projection and Spatial Evaluation

Per-tile predictions were projected back onto the 3D mesh using UV barycentric interpolation, generating georeferenced diagnostic rasters and point-wise labels. All spatial transforms preserved the original GNSS georeference and included uncertainty metadata derived from tie-point residuals. Projection routines and coordinate conventions were documented in the reproducibility package.

Spatial performance was evaluated using Intersection-over-Union (IoU), defined as:

$$IoU = \frac{|Prediction \cap GroundTruth|}{|Prediction \cup GroundTruth|}$$

(Eq. 1)

IoU was computed over 120 matched façade patches and 12,448 region-based comparisons between the supervised RF predictions and the unsupervised clustering outputs. This metric quantifies the spatial overlap between two segmentation results, with higher values indicating better agreement. Patch selection methodology is detailed in Appendix F.

2.4.4 Semantic Translation Validation Protocol

The translation from chromatic clusters to semantically meaningful decay categories was validated through a structured multi-expert protocol. Three domain experts independently assigned decay categories to each chromatic cluster based on: (a) visual inspection of the cluster's colour characteristics; (b) spatial correspondence with known decay patterns from on-site condition assessments; and (c) comparison with reference images from conservation literature. Inter-rater agreement was quantified using Fleiss' κ (Fleiss, 1971), calculated across all clusters and experts. Rather than forcing consensus through re-evaluation—which would introduce consensus bias and inflate the agreement statistic—clusters with $\kappa < 0.60$ were explicitly labeled as 'ambiguous transition' and reported with lower confidence scores in the output maps. This approach provides transparency regarding

classification uncertainty and establishes a protocol for managing interpretive subjectivity in future applications.

2.5 Heritage Values Index (HVI) and Decision Support

A Heritage Values Index (HVI) was developed to quantify cultural significance across five dimensions: authenticity, historical value, social value, aesthetic value, and traditional knowledge value. The HVI was derived from stakeholder consultation (n = 25, comprising heritage agency staff, local community representatives, and conservation professionals) using a modified Delphi process, with scores normalised to a 0–100 scale. Stakeholder composition and Delphi protocol are detailed in Appendix G. Internal consistency was assessed using Cronbach's α (Cronbach, 1951) with a target > 0.70, and test-retest reliability was evaluated using intraclass correlation coefficients (ICC) with a target > 0.80.

To operationalise the priority matrix, a final decision score (FDS) was calculated for each spatial zone as:

$$FDS = (0.6 \times \ Risk_Score_{cal}) + (0.4 \times \ HVI_Normalised)$$

(Eq. 3)

where Risk_Score_cal is the calibrated RF-predicted probability of deterioration (scaled 0–100) and HVI_Normalised is the Heritage Values Index (scaled 0–100). The 60/40 weighting was derived through a sensitivity analysis (presented in Appendix H) that varied weights from 50/50 to 80/20 and assessed the stability of the resulting priority rankings; the 60/40 weighting produced rankings that were robust to weight variations within ± 10%. Zones with FDS > 75 were flagged for immediate intervention, while zones with FDS < 40 were designated for routine monitoring. This weighted scoring balances objective environmental risk with culturally defined significance, providing a transparent, auditable basis for resource allocation. The resulting decision-support matrix is depicted in Fig. 6.

2.6 Lead Time Calculation and Uncertainty Quantification

Lead time was calculated as the interval between a high-risk prediction (from the Random Forest classifier) and the actual onset of a high-risk event in the test dataset. High-risk events were defined using thresholds derived from an independent calibration set (20% of data, November–December 2025): (a) RH > 85% for ≥ 6 consecutive hours; or (b) RH > 80% for ≥ 12 consecutive hours with concurrent temperature > 30°C. These thresholds were validated through mixed-effects modelling showing strong association with material condition scores (Section 3.1).

Uncertainty was quantified using 95% confidence intervals estimated via bootstrap resampling (1,000 iterations) for all performance metrics. For mixed-effects models, confidence intervals were computed using the parametric bootstrap method. For Spearman's rank correlation, the coefficient was calculated using **Eq. 2**:

$$\rho \ = 1 - \frac{6 \sum \ {}^{n}_{i=1} \ d_i^2}{n(n^2 - 1)}$$

(Eq. 2)

where d_i is the difference between the ranks of each observation. The lead times reported are technical—derived from retrospective test data—and do not imply operational feasibility (staff response within these windows was not assessed).

2.7 Predictive Forecasting and Cross-Site Validation

To assess the framework's capability for genuine forward prediction—rather than mere classification of contemporaneous data—an expanding window cross-validation was conducted. The model was trained on the first 6 months of data (March–August 2025) and tested on the subsequent 2-month blocks (September–October 2025, November–December 2025, January–February 2026, March–April 2026), with the training window expanding incrementally. This design allows assessment of model performance as the forecast horizon extends, with the caveat that the initial 6-month training period does not include the peak wet season (November–March), representing a conservative test of the model's ability to generalise to unseen seasonal conditions.

Additionally, leave-one-site-out cross-validation was implemented to evaluate spatial generalisability: the model was trained on two sites and tested on the held-out third site, iterating across all three combinations. This provides a more rigorous assessment of transferability to unseen heritage buildings than within-site validation alone.

2.8 Statistical Analysis and Ethics Statement

All statistical analyses were conducted using R (version 4.3) and Python (version 3.10). Mixed-effects models were fitted using the lme4 package in R, with site as a random intercept to account for the nested data structure (observations within architectural zones, zones within sites). Multiple testing correction was applied using the Bonferroni method, with adjusted significance thresholds reported alongside unadjusted p-values. Effect sizes (η^2 for mixed models, Cohen's d for pairwise comparisons) are reported for all statistically significant findings. A post-hoc power analysis (G*Power) indicated that the achieved sample size (48,830 labelled pixels) provides > 95% power to detect a medium effect size (Cohen's $f^2 = 0.15$) at $\alpha = 0.05$, confirming adequate statistical sensitivity.

Ethics Statement: The stakeholder consultation component (n = 25) was conducted in accordance with the ethical guidelines of Universitas Diponegoro. All participants provided written informed consent prior to participation. The study protocol was reviewed and approved by the Universitas Diponegoro Research Ethics Committee (Approval Number: 123/UN7.P/HK/2025, dated 15 February 2025). The research was conducted in accordance with the Declaration of Helsinki (1964) and its subsequent amendments.

3. Results

3.1 Spatial Microclimatic Characterisation

Continuous monitoring revealed a clear spatial gradient across the study sites. The basement zones of Lawang Sewu exhibited the highest relative humidity (mean 84.6%, SD 6.5%) and the lowest temperatures (mean 28.4°C, SD 1.2°C), while roof spaces showed the lowest relative humidity (mean 67.5%, SD 5.5%) and the highest temperatures (mean 31.8°C, SD 2.8°C). At Pasar Johar, the semi-open corridors exhibited intermediate conditions (mean RH 76.3%, SD 4.2%; mean T 30.1°C, SD 1.8°C), while the enclosed stalls showed higher humidity (mean RH 81.2%, SD 5.1%) due to limited ventilation. At Tawang Railway Station, the main waiting hall exhibited moderate conditions (mean RH 72.8%, SD 3.8%; mean T 29.7°C, SD 2.1°C), while the platform areas showed greater temperature variability (SD 3.2°C) due to solar exposure. This gradient reflects the architectural mediation of environmental stress through ventilation pathways, thermal mass, and solar exposure. Table 2 presents descriptive statistics by site and architectural zone.

Table 2
Descriptive statistics of temperature and relative humidity by site and architectural zone (hourly aggregated data). Values are means ± standard deviation.

Site	Zone	n (observations)	Temperature (°C)	Relative Humidity (%)
Lawang Sewu	Basement	16,176	28.4 ± 1.2	84.6 ± 6.5
Lawang Sewu	North Corridor	16,176	29.1 ± 1.5	78.2 ± 5.8
Lawang Sewu	Front Hall	16,176	30.2 ± 2.1	74.1 ± 7.1
Pasar Johar	Enclosed Stalls	16,176	29.8 ± 1.6	81.2 ± 5.1
Pasar Johar	Semi-open Corridors	16,176	30.1 ± 1.8	76.3 ± 4.2
Tawang Station	Main Hall	16,176	29.7 ± 2.1	72.8 ± 3.8
Tawang Station	Platform Area	16,176	31.2 ± 3.2	68.5 ± 4.5

Note: Raw 15-minute data totals 517,632 observations across all 16 active sensor nodes. The n values shown above represent hourly aggregated data (after resampling).

Linear mixed-effects models, with site as a random intercept to account for the nested data structure, confirmed statistically significant differences among zones for both temperature ($\chi^2(6) = 412.7$, $p < 0.001$, $\eta^2 = 0.51$) and relative humidity ($\chi^2(6) = 287.3$, $p < 0.001$, $\eta^2 = 0.44$). Post-hoc pairwise comparisons with Tukey adjustment showed that all pairwise differences were significant at $\alpha = 0.05$, with the largest contrast between Lawang Sewu basement and Tawang platform area ($\Delta T = 2.8$°C, 95% CI: 2.4–3.2°C; $\Delta RH = 16.1$%, 95% CI: 14.3–17.9%, $p < 0.001$ for both).

Spearman's rank correlation was used to examine associations between environmental variables and material condition scores (measured on a four-point ordinal scale: 1 = no visible deterioration, 4 = severe). For this analysis, variables were aggregated at the sensor node level (n = 16 nodes) to ensure independence. Because of the limited sample size at the node level, confidence intervals for Spearman's ρ

were estimated using cluster-robust bootstrap resampling (1,000 iterations) with resampling at the zone level. After Bonferroni correction for 24 comparisons ($\alpha_{adj} = 0.002$), cumulative hours of relative humidity exceeding 85% showed the strongest positive association with material deterioration ($\rho = 0.82$, 95% CI: 0.71–0.89, $n = 16$, $p < 0.001$). Mean relative humidity ($\rho = 0.78$, 95% CI: 0.66–0.86, $n = 16$, $p < 0.001$) and maximum relative humidity ($\rho = 0.71$, 95% CI: 0.57–0.81, $n = 16$, $p < 0.001$) also showed strong positive correlations. Mean temperature showed a moderate negative correlation ($\rho = -0.52$, 95% CI: -0.68 to -0.32, $n = 16$, $p = 0.006$), consistent with the observation that cooler, more humid zones exhibited more severe deterioration. The complete correlation matrix is displayed in Fig. 7.

Figure 7. Correlation matrix showing Spearman's rank correlations between environmental variables and material condition scores (n = 16 nodes, cluster-robust CIs). Red indicates positive correlation, blue indicates negative correlation. Asterisks denote statistical significance after Bonferroni correction (*p < 0.05, **p < 0.01, ***p < 0.001).

3.2 Unsupervised Hierarchical Clustering Performance

Hierarchical clustering produced stable partitions across all study sites, with mean silhouette scores of 0.75 (SD = 0.02, range: 0.72–0.78), indicating good cluster separation. Bootstrap resampling (1,000 iterations) confirmed cluster stability, with 92% of bootstrap replicates producing the same cluster assignments. The silhouette analysis, presented in Fig. 8, confirms the optimal cluster configurations for each site.

The number of clusters identified varied by site, reflecting differences in surface heterogeneity and chromatic complexity. Lawang Sewu, with its relatively homogeneous brick and plaster surfaces, yielded 5–6 clusters, while Pasar Johar, with its heterogeneous materials (brick, timber, concrete) and varied environmental conditions, yielded 7–9 clusters. Tawang Railway Station, with its combination of concrete and brick surfaces, yielded 6–8 clusters. The clustering outputs served as diagnostic priors, providing an initial qualitative understanding of chromatic groupings before supervised classification was applied.

3.3 Supervised Random Forest Classification Performance

The Random Forest classifier demonstrated consistent performance across the three study sites, as reported in Table 3. Lawang Sewu achieved the highest overall F1-score (0.90 ± 0.02), while Pasar Johar recorded the lowest (0.84 ± 0.04). This disparity reflects variation in material complexity and illumination conditions, with the heterogeneous surfaces and varied environmental conditions of Pasar Johar presenting a greater classification challenge.

Table 3
Site-specific Random Forest classification performance metrics. Values are means ± standard deviation from 5-fold cross-validation.

Site	Precision	Recall	F1-Score	AUC	Support (n)
Lawang Sewu	0.91 ± 0.02	0.89 ± 0.03	0.90 ± 0.02	0.96	18,240
Pasar Johar	0.85 ± 0.04	0.83 ± 0.05	0.84 ± 0.04	0.92	14,870
Tawang Station	0.88 ± 0.03	0.86 ± 0.04	0.87 ± 0.03	0.94	15,576

Aggregated across all sites (67,686 labelled instances; note that this aggregate total includes the full dataset used for the per-class analysis in Table 4, whereas Table 3 reports per-site cross-validation results on training subsets), the classifier showed strong performance across the five decay categories, as detailed in Table 4. Unaltered regions proved the easiest to classify (F1-score = 0.92), while chromatic alterations exhibited the highest ambiguity (F1-score = 0.85). This ambiguity is attributed to the subtle visual transitions that characterise this category, which often overlap with other decay categories in HSV colour space.

Table 4
Aggregate classification metrics by decay category. Values are means ± standard deviation.

Decay Class	Precision	Recall	F1-Score	Support (n)
Moisture Discoloration	0.91 ± 0.02	0.89 ± 0.03	0.90 ± 0.02	15,103
Biological Growth	0.88 ± 0.03	0.86 ± 0.04	0.87 ± 0.03	17,244
Surface Accumulations	0.87 ± 0.03	0.88 ± 0.03	0.88 ± 0.03	9,872
Chromatic Alteration	0.86 ± 0.03	0.84 ± 0.04	0.85 ± 0.03	12,460
Unaltered Regions	0.92 ± 0.02	0.91 ± 0.02	0.92 ± 0.02	13,007

Note: The aggregate support in Table 4 *(67,686) includes all instances used for per-class analysis after SMOTE augmentation on the training set. The per-site supports in* Table 3 *(48,686 total) represent the original training+validation+test split without augmentation.*

3.4 Feature Importance and Model Ablation

Feature importance analysis, visualised in Fig. 9, identified the Humidity Stress Factor (HSF)—the composite feature integrating current RH, cumulative exposure, and lagged RH—as the single most important predictor (importance score 0.264). This was followed by current temperature (importance = 0.087) and temperature lagged by 12 hours (importance = 0.041). The dominance of the HSF composite confirms that humidity stress is the primary driver of chromatic decay in this tropical context, consistent with the physical understanding that temperature plays a secondary role to moisture in driving chromatic deterioration. To address known biases in Random Forest feature importance with correlated features, we

also computed permutation importance as a validation; the HSF remained the top-ranked feature, confirming the robustness of this finding.

To quantify the contribution of each model component, an ablation study was performed (Table 5). Removing the HSF (Variant B) reduced the macro-F1 by 9.2% (from 0.87 to 0.79), confirming its critical predictive role. The full model (Variant C) significantly outperformed the no-clustering variant (Variant D), validating the value of the unsupervised priors.

Table 5
Ablation study results (macro-F1, aggregated across sites, 5-fold CV).

Model Variant	Macro-F1	Δ from Full	95% CI
Variant A (Radiometric + Textural only)	0.76	-0.11	[0.74, 0.78]
Variant B (+ Environment, - HSF)	0.79	-0.08	[0.77, 0.81]
Variant D (Full – Clustering Priors)	0.83	-0.04	[0.81, 0.85]
Variant C (Full Model)	**0.87**	**Baseline**	**[0.85, 0.89]**

3.5 Comparison with Supervised Baselines

To benchmark the RF model against standard machine learning algorithms, we evaluated Logistic Regression (LR), Support Vector Machine (SVM) with RBF kernel, and XGBoost under identical train/test splits and feature sets (Table 6). RF achieved the highest macro-F1 (0.87), outperforming XGBoost (0.84), SVM (0.81), and LR (0.76), justifying its selection for the framework.

Table 6
Performance comparison with supervised baseline models (5-fold CV, mean ± SD).

Model	Macro-F1	AUC	Precision (Macro)
Logistic Regression	0.76 ± 0.03	0.85 ± 0.02	0.75 ± 0.03
SVM (RBF)	0.81 ± 0.03	0.89 ± 0.02	0.80 ± 0.03
XGBoost	0.84 ± 0.02	0.93 ± 0.01	0.84 ± 0.02
Random Forest (Ours)	**0.87 ± 0.02**	**0.95 ± 0.01**	**0.87 ± 0.02**

3.6 Comparative Methodological Performance

The supervised Random Forest classifier consistently outperformed unsupervised hierarchical clustering across all decay categories, as shown in Table 7. Improvements in Intersection-over-Union (IoU), calculated using **Eq. 1**, ranged from 17% (biological growth) to 51% (chromatic alteration). These improvements are attributable to the RF model's ability to exploit textural descriptors, chromatic statistics, and environmental features, whereas the hierarchical clustering method relies solely on

chromatic similarity and therefore tends to misclassify regions with subtle material transitions or complex shading.

Table 7
Intersection-over-Union (IoU) comparison between unsupervised clustering and supervised Random Forest classification. Values are means ± standard deviation.

Decay Class	Unsupervised IoU	Supervised IoU	Improvement (%)
Moisture Discoloration	0.52 ± 0.06	0.64 ± 0.05	23%
Biological Growth	0.54 ± 0.05	0.63 ± 0.04	17%
Surface Accumulations	0.48 ± 0.07	0.61 ± 0.06	27%
Chromatic Alteration	0.35 ± 0.08	0.53 ± 0.07	51%
Unaltered Regions	0.61 ± 0.04	0.72 ± 0.03	18%

Note: Improvement is calculated as (Supervised IoU - Unsupervised IoU) / Unsupervised IoU × 100%.

The spatial distribution of improvements showed that high-IoU zones clustered around well-lit, planar, and low-relief surfaces, while low-IoU zones concentrated around recessed components, deep shadows, intricate ornamentation, and geometrically complex regions. This pattern aligns with theoretical expectations: homogeneous lighting and simple geometry generate more stable features for both unsupervised and supervised algorithms, while complex surface relief intensifies ambiguity in colour-only clustering.

3.7 Lead Time Analysis

Lead time analysis was conducted on the test dataset (20% of data, January–April 2026). For correctly predicted high-risk events (79% of all events, recall = 0.79), mean conditional lead times were 4.8 hours (SD = 1.2, range: 2.5–6.0) for the 6-hour forecast horizon, 9.2 hours (SD = 2.4, range: 4.0–12.0) for the 12-hour horizon, and 16.7 hours (SD = 4.8, range: 6.5–24.0) for the 24-hour horizon. The expected lead times —weighted by recall to account for missed events—were 3.8 hours, 7.3 hours, and 13.2 hours respectively. Bootstrap 95% CIs for the 24-hour conditional mean were [14.8, 18.6] hours. These technical estimates should be interpreted with caution; they do not imply that heritage institutions can operationally respond within these windows, as factors such as staff availability, decision protocols, budget cycles, and regulatory approval were not examined. The complete lead time statistics are presented in Table 8.

Table 8
Lead time statistics with 95% bootstrap confidence intervals (1,000 iterations).

Forecast Horizon	n (events)	Recall	Conditional Mean (hours)	95% CI (Conditional)	Expected Mean (hours)*
6 hours	174	0.79	4.8	[4.2, 5.4]	3.8
12 hours	152	0.79	9.2	[8.1, 10.3]	7.3
24 hours	116	0.79	16.7	[14.8, 18.6]	13.2

Expected mean = Conditional mean × Recall (weighted by the probability of correct detection).

Note

Only events correctly predicted by the Random Forest classifier (recall = 0.79) are included in the conditional means. Lead times are calculated on retrospective test data. This study does not assess whether heritage institutions can act upon these warnings within these timeframes; claims of actionability are strictly technical (see Section 4.3*).*

3.8 Predictive Forecasting Validation

The expanding window cross-validation demonstrated that the framework maintains predictive performance over increasing forecast horizons. The predictive F1-scores for each forecast horizon were: 0.86 (1-month ahead), 0.83 (3-months ahead), and 0.79 (6-months ahead). Performance decay was most pronounced for the 'chromatic alteration' category (F1 dropped from 0.85 to 0.74 over 6 months), reflecting the progressive nature of chromatic degradation that becomes increasingly difficult to predict over longer horizons.

For the 6-month forecasting horizon, because chromatic manifestation lags behind environmental stress accumulation, the 6-month output should be interpreted as a Cumulative Environmental Stress Index rather than a deterministic pixel-level chromatic prediction. This distinction is critical: the model reliably forecasts the duration and intensity of damaging microclimatic conditions, which heritage managers can use to schedule maintenance cycles, even if the exact visual onset of decay remains probabilistic.

These results confirm that the framework is suitable for short-to-medium term forecasting (1–3 months), while longer-term predictions require more frequent model recalibration or the incorporation of seasonal climatic patterns. This predictive validation complements the classification accuracy reported in Section 3.3 and demonstrates that the framework can generalise forward in time—a critical requirement for operational early warning systems.

3.9 Cross-Site Generalisability (Leave-One-Site-Out)

Leave-one-site-out cross-validation assessed the spatial transferability of the framework (Table 9). When trained on two sites and tested on the third, F1 scores ranged from 0.79 to 0.82. The slight performance

drop compared to the within-site validation (0.84–0.90) was expected, reflecting site-specific material and microclimatic nuances. These results confirm that the framework generalises reasonably to unseen sites, albeit with local recalibration recommended for optimal performance.

Table 9
Leave-one-site-out cross-validation results (macro-F1, 5-fold CV).

Training Sites	Test Site	F1 Score	95% CI
Pasar Johar + Tawang	Lawang Sewu	0.82	[0.79, 0.85]
Lawang Sewu + Tawang	Pasar Johar	0.79	[0.76, 0.82]
Lawang Sewu + Pasar Johar	Tawang	0.81	[0.78, 0.84]

3.10 Heritage Values Index and Priority Matrix

The Heritage Values Index (HVI) showed good psychometric properties, with Cronbach's α = 0.87 (95% CI: 0.83–0.91) and test-retest ICC = 0.88 (95% CI: 0.82–0.92), exceeding the acceptable thresholds. Confirmatory factor analysis supported the five-factor structure (CFI = 0.91, RMSEA = 0.06, SRMR = 0.05).

Across the study sites, Lawang Sewu scored highest on historical significance (94.2) and authenticity (92.4), while the basement zones scored lowest on community attachment (48) and social value (52). Pasar Johar scored highest on social value (88.1) and community attachment (85.4), reflecting its role as a vibrant public market. Tawang Railway Station scored highest on historical significance (91.3) and aesthetic value (87.6), reflecting its architectural heritage as a colonial transport landmark. The Main Entrance Hall of Lawang Sewu achieved the highest overall HVI score (88.2), reflecting its role as the primary public face of the building.

The Final Decision Score (FDS), calculated as FDS = (0.6 × Risk_Score_cal) + (0.4 × HVI_Normalised) (**Eq. 3**), was used to generate the priority intervention matrix (Fig. 6). The Main Entrance Hall of Lawang Sewu was identified as the highest priority for intervention (FDS = 82.4): although its environmental risk was moderate (mean RH 74.1%, 1,752 hours > 85% RH), its heritage value was exceptionally high (88.2). The basement of Lawang Sewu, despite low heritage value (58.6), warranted attention due to its high environmental risk (mean RH 84.6%, 5,694 hours > 85% RH), yielding an FDS of 74.3. The enclosed stalls of Pasar Johar, with high environmental risk (mean RH 81.2%, 3,240 hours > 85% RH) and moderate heritage value (72.4), were identified as a moderate priority (FDS = 73.0) requiring technical intervention. The platform area of Tawang Railway Station, with low environmental risk but moderate heritage value (74.8), merited continued monitoring (FDS = 52.8).

4. Discussion

4.1 Interpretation of Findings in Context

The results demonstrate that integrating continuous environmental sensing with multi-stage machine learning analysis can provide robust, interpretable predictions of chromatic decay on tropical heritage façades. The Random Forest classifier achieved F1-scores of 0.84–0.90 across diverse architectural contexts, with the integration of environmental features substantially improving spatial accuracy (IoU improvements of 17–51%) compared to unsupervised clustering alone. The ablation study confirmed the HSF as the most critical component, with a 9.2% drop in performance upon its removal. This empirically validates the physical understanding that cumulative humidity stress, rather than instantaneous measurements, drives chromatic deterioration in tropical environments.

The strong association between cumulative hours of RH > 85% and material deterioration ($\rho = 0.82$, 95% CI: 0.71–0.89, $n = 16$, $p < 0.001$ after Bonferroni correction) is consistent with established findings in the literature. Warscheid and Braams (2000) identified humidity as the primary driver of biological and salt-related deterioration, while Sesana et al. (2019) documented that RH levels consistently above 75–80% promote multiple decay mechanisms. However, the present study extends these findings by quantifying the association at high temporal resolution (15-minute intervals over 14 months) and by incorporating humidity exposure as a predictive feature rather than a post-hoc explanatory variable. The threshold of 85% RH, derived empirically from the calibration dataset, aligns with findings from Rukayah et al. (2023) in Semarang, who observed active mould growth and salt efflorescence at RH levels of 84–88%, but is higher than the 75–80% thresholds commonly cited in European studies (Sesana et al., 2019). This discrepancy underscores the need for local calibration of humidity thresholds in tropical contexts.

The spatial heterogeneity documented across architectural zones—with Lawang Sewu basement exhibiting mean RH of 84.6% and Tawang platform area 68.5%—confirms the importance of zone-specific monitoring. Costa Carrapiço et al. (2022) demonstrated that microclimatic heterogeneity in historic buildings requires dense sensor networks positioned according to architectural logic rather than arbitrary grids. The present study's strategic sensor placement along ventilation pathways, thermal gradients, and material transitions captures this heterogeneity more effectively than uniform grid sampling would, supporting the architectural sensing principle advocated by Mitro et al. (2022).

The Random Forest feature importance analysis, visualised in Fig. 9, revealed that the Humidity Stress Factor (HSF) composite dominates the prediction with 26.4% of total importance. This finding aligns with the physical understanding that chromatic decay in tropical environments is primarily moisture-driven, with temperature playing a secondary role. The inclusion of time-of-day and day-of-year cyclic encodings among the top-ten features confirms the importance of diurnal and seasonal patterns identified in the time series decomposition.

The lead time analysis demonstrated that the framework can provide technical warnings 4.8–16.7 hours before high-risk conditions materialise, assuming correct prediction. These lead times are comparable to those reported by Krommyda et al. (2022), who achieved 6–24 hour forecasts for indoor RH in a temperate heritage building using LSTM networks. The present study's use of Random Forest rather than LSTM prioritises interpretability (through feature importance) over marginal gains in forecast accuracy, a

trade-off that is appropriate for conservation contexts where decision-makers must understand the rationale behind alerts.

The semantic translation validation protocol (Section 3.4) demonstrates that the mapping from chromatic clusters to decay categories can be systematically validated. The substantial agreement (Fleiss' κ = 0.84) suggests that, despite inherent subjectivity, structured protocols can achieve meaningful consistency. The identification of transitional clusters as a category with confidence scores provides a transparent mechanism for acknowledging uncertainty in semantic interpretation.

4.2 Comparison with Prior Work

The present study differs from prior work in several respects. Compared to Galantucci et al. (2023), who developed a machine learning approach for 3D decay segmentation without environmental integration, this study incorporates continuous environmental sensing as a predictive layer. Compared to prior work by the authors (Gbran, 2026a), which presented a hybrid framework for chromatic decay detection using hierarchical clustering and Random Forest but treated environmental data as post-hoc explanatory variables, this study integrates environmental features directly into the classification model via the Humidity Stress Factor (HSF) composite. Critically, this study introduces rigorous methodological refinements not present in earlier work: (1) prevention of data leakage through independent clustering on training data; (2) comprehensive ablation testing to quantify the contribution of each component; (3) leave-one-site-out cross-validation to substantiate generalisability claims; (4) probability calibration via Platt scaling; and (5) benchmarking against multiple supervised baselines.

The integration of IoT sensing with AI analytics represents a step toward the hybrid AI-IoT frameworks advocated by Al-Raeei (2024) and Bibri et al. (2025), though full operational deployment—with real-time alerts and institutional integration—remains a direction for future research. The explicit uncertainty quantification (95% confidence intervals, bootstrap resampling, and the predictive forecasting validation in Section 3.7) addresses a common weakness in heritage AI studies identified by Fiorucci et al. (2020), where point estimates are often presented without measures of precision or temporal generalisation.

4.3 Limitations and Subjectivity Acknowledgment

Several limitations must be acknowledged. First, the study is contextual—validated on three heritage sites in Semarang, Indonesia, representing specific architectural typologies and material conditions. The framework is presented as a methodologically reproducible blueprint rather than a universally pre-calibrated model. Generalisation to other tropical contexts, or to non-tropical contexts, requires local recalibration of thresholds, feature engineering, and model parameters. The humidity threshold of 85% RH, for example, may not be appropriate for regions with different material traditions or climatic regimes. While leave-one-site-out cross-validation (F1 = 0.79–0.82) demonstrates acceptable spatial transferability, broader validation across multiple cities and countries is needed.

Second, the mapping from purely chromatic clusters to semantically meaningful decay categories (e.g., 'biological patina', 'moisture-related staining') relies on expert interpretation. While the use of three

domain experts and the structured validation protocol (Fleiss' κ = 0.84) substantially mitigates subjectivity, the process is not fully automated. The identification of 9.4% of clusters as 'ambiguous transition' with confidence scores provides a transparent mechanism for managing this limitation. This limitation is common in heritage diagnostics, where physical sampling or laboratory analysis is often impractical or prohibited (Doehne and Price, 2010). The approach provides a transparent and documented mechanism for handling this subjectivity.

Third, the lead times reported are technical, derived from retrospective test data. The study did not assess whether heritage institutions can act upon these warnings within these timeframes. Operational feasibility—including staff availability, decision protocols, budget cycles, and regulatory approval—was not examined. This is a critical gap for real-world deployment and should be addressed in future implementation research.

Fourth, while the 16-node deployment represents a comprehensive IoT monitoring network for this context, the sensor density (approximately 1 node per 350 m² at Lawang Sewu, 1 per 800 m² at Pasar Johar, and 1 per 1,125 m² at Tawang Station) remains below the 50–70 m² per node recommendation by Costa Carrapiço et al. (2022) for comprehensive microclimatic characterisation. Strategic placement along architectural axes—ventilation pathways, thermal gradients, plinth zones, and exterior-interior interfaces—partially compensates for this limitation by capturing the most environmentally critical gradients. However, future deployments should increase density to 1 node per 50–70 m² to capture intra-façade heterogeneity, particularly on geometrically complex façades with recessed components and intricate ornamentation.

Fifth, the study relies exclusively on visible-spectrum imaging and does not incorporate multispectral or hyperspectral data, which could provide additional information about subsurface deterioration processes (He et al., 2025). The exclusion of such data was a deliberate choice to maintain methodological accessibility and cost-effectiveness, but it represents a limitation for detecting deterioration that is not yet visible at the surface.

Sixth, the monitoring period of 14 months captures seasonal variation but does not capture inter-annual variability or long-term trends. Extended monitoring over multiple years would be necessary to assess the framework's performance under different climatic conditions, including El Niño-Southern Oscillation cycles that affect humidity patterns in Southeast Asia.

Seventh, while the expanding window cross-validation demonstrates temporal generalisation and leave-one-site-out validation demonstrates spatial generalisability, the model was not validated on independent sites outside the study region. Cross-site validation on additional heritage buildings in different Indonesian cities or other tropical countries is needed to further assess model transferability.

Eighth, the DHT22 sensor (± 2–5% RH) introduces moderate measurement noise that may affect the precision of the HSF composite. While the drift correction protocol and the use of the SHT35 reference

sensor mitigate this limitation, higher-precision sensors (e.g., SHT35 or SHT85) would enhance sensitivity in future deployments.

4.4 Practical and Policy Implications

Despite these limitations, the framework offers several practical contributions. The priority intervention matrix (Fig. 6) and the Final Decision Score (FDS) provide a structured, transparent mechanism for allocating limited conservation resources, balancing measurable deterioration risk against culturally defined significance. This addresses a persistent challenge in heritage management, where decisions must be made under uncertainty and with competing priorities (Avrami et al., 2000; Avrami, 2024).

The documented deployment and calibration protocols enable replication in similar tropical heritage contexts. The use of low-cost, open-source components (ESP32, DHT22, Google Sheets backend) makes the framework accessible to heritage agencies with limited budgets, addressing the resource constraints that often limit technology adoption in Global South contexts (Bruno et al., 2020).

For policy, the framework supports UNESCO's Historic Urban Landscape approach and the Sustainable Development Goals (SDG 11: Sustainable Cities and Communities; SDG 13: Climate Action), providing evidence that predictive conservation frameworks can be implemented in resource-constrained contexts using accessible technologies. The explicit uncertainty quantification and acknowledgment of limitations align with emerging standards for responsible AI in cultural heritage (Casillo et al., 2024).

5. Conclusion

This study has developed and empirically validated a hybrid AI-IoT framework for early warning of chromatic deterioration on tropical heritage façades. The framework integrates continuous environmental sensing with multi-stage machine learning analysis, combining unsupervised hierarchical clustering with supervised Random Forest classification that incorporates microclimatic variables as predictive features through a composite Humidity Stress Factor. Validation on three heritage sites in Semarang, Indonesia—Lawang Sewu, Pasar Johar, and Tawang Railway Station—demonstrated consistent performance (F1-scores 0.84–0.90, overall accuracy > 90%), with substantial improvements in spatial accuracy (IoU improvements of 17–51%) compared to unsupervised clustering alone.

The study makes several contributions. Methodologically, it demonstrates the value of integrating environmental sensing as a predictive layer rather than a post-hoc explanatory variable, with the Humidity Stress Factor composite dominating feature importance rankings. A rigorous ablation study confirmed the HSF as the most critical component, with a 9.2% absolute drop in macro-F1 upon its removal. The framework also incorporates rigorous data leakage prevention, probability calibration via Platt scaling, and benchmarking against multiple supervised baselines (LR, SVM, XGBoost). Technically, it provides documented calibration protocols for low-cost sensors under challenging tropical conditions, achieving 98.4% data completeness over 14 months from 16 active sensor nodes. Practically, it offers a priority intervention matrix and a Final Decision Score that combines environmental risk with heritage significance, providing a structured decision-support tool for heritage managers. Leave-one-site-out

cross-validation (F1 = 0.79–0.82) confirms acceptable spatial transferability, while expanding window cross-validation demonstrates temporal generalisation over 1–6 month forecast horizons.

The study also explicitly acknowledges its limitations. The findings are contextual to the study sites and should not be generalised without local recalibration. The mapping from chromatic clusters to decay categories involves expert interpretation and introduces subjectivity; the structured validation protocol (Fleiss' $\kappa = 0.84$) with explicit labeling of transitional clusters provides a mechanism for managing this. The lead times are technical and do not imply operational feasibility. Cross-validation on additional heritage buildings is needed to further assess model transferability.

Future research should extend this work in several directions. First, the framework should be applied to a wider range of tropical heritage buildings across Southeast Asia, Latin America, and West Africa to assess transferability across diverse climatic, material, and cultural contexts. Second, the integration of additional data sources—satellite-based microclimate data, air quality monitoring, and high-resolution multispectral or hyperspectral imaging—could enhance predictive accuracy. Third, the HVI should be integrated directly into the predictive models as a feature rather than used as a post-hoc filter, enabling the models to learn interactions between environmental risk and heritage significance. Fourth, operational feasibility should be assessed through pilot implementations with heritage institutions, examining staff response times, decision protocols, and resource constraints. Fifth, longitudinal monitoring over multiple years is needed to capture inter-annual variability and assess model performance under different climatic conditions. Sixth, higher-precision sensors (e.g., SHT35) should be deployed to reduce measurement noise and enhance HSF precision. Seventh, a unified predictive model that combines environmental risk assessment with heritage significance weighting in a single objective function could be developed, enabling simultaneous optimisation of conservation priorities across both dimensions. This would represent a significant advancement toward fully integrated conservation decision-support systems.

In conclusion, this study provides evidence that predictive monitoring frameworks are achievable in resource-constrained tropical contexts using low-cost, open-source methods. The framework enables early detection and targeted intervention, supporting a transition from reactive repair to anticipatory management in heritage conservation. While the path to widespread adoption remains open—requiring further validation, institutional integration, and operational testing—the foundations established here affirm that the integration of AI and IoT can support, rather than replace, the professional judgment of heritage practitioners.

Declarations

Funding:

This study was conducted as an independent, self-funded research effort. No institutional, organisational, or governmental funding was received for this work.

Acknowledgements

The authors thank the Semarang City Cultural Heritage Agency and the custodial institutions of the three study sites—Lawang Sewu, Pasar Johar, and Tawang Railway Station—for providing access and support. The authors gratefully acknowledge the three domain experts who contributed to the semantic translation validation (names and affiliations provided in Appendix D). The research was conducted with the approval of the Universitas Diponegoro Research Ethics Committee (Approval Number: 123/UN7.P/HK/2025, dated 15 February 2025). No conflicts of interest are declared.

Data Availability

The datasets generated and/or analysed during the current study are available in the GitHub repository, https://github.com/hassangbran/SUPPLEMENTARY-DATASET-A-Hybrid-AI-IoT-/issues/1, as well as the Zenodo repository, https://doi.org/10.5281/zenodo.17333842, and the OSF repository, https://doi.org/10.17605/OSF.IO/GM47X. The complete reproducibility package includes raw IoT sensor data, processed feature sets, expert-annotated labels, source code, calibration artifacts, and 3D photogrammetric models. All data and code are provided under the Creative Commons Attribution 4.0 International (CC BY 4.0) license.

Declaration of Generative AI and AI-Assisted Technologies

Grammarly and DeepL were used to improve language clarity and grammar. No generative AI was used for data analysis, code generation, or scientific interpretation. All figures were created manually using Python (Matplotlib, Seaborn) and Draw.io. The Random Forest model was implemented in scikit-learn (version 1.3) with all parameters set explicitly as described in Section 2.4.

Conflicts of Interest and Self-Citation Disclosure

The authors declare no competing interests. The authors have not received any financial or non-financial support that could be perceived as a conflict of interest.

Author Contribution

H.G. conceived and designed the study, developed the research methodology, conducted field investigations, collected and analyzed the data, implemented the AI and IoT framework, performed the machine learning analyses, prepared all figures and visualizations, and wrote the original manuscript. S.R. supervised the research, contributed to the study design and interpretation of the results, and critically revised the manuscript. A.S. contributed to the research methodology, validation of the findings, and critical revision of the manuscript. E.S. contributed to the interpretation of the results, manuscript review, and academic supervision. I.S. contributed to the IoT system design, sensor deployment, technical validation, and manuscript review. All authors reviewed, edited, approved the final manuscript, and agreed to be accountable for all aspects of the work.

References

1. Addabbo, T. et al. A city scale IoT architecture for monumental structures monitoring. *Measurement* **131**, 349–357. 10.1016/j.measurement.2018.08.058 (2019).
2. Al-Raeei, M. Integrating artificial intelligence and IoT in heritage conservation: A predictive monitoring framework. *Buildings* **14** (12), 3979. 10.3390/buildings14123979 (2024).
3. Avrami, E. Conservation and the postcolonial turn: Rethinking heritage in the global South. *Change Over Time*. **11** (1), 6–23 (2024).
4. Avrami, E., Mason, R. & de la Torre, M. *Values and Heritage Conservation* (Getty Conservation Institute, 2000).
5. Bibri, S. E., Allam, Z. & Krogstie, J. Smart heritage cities: Integrating IoT and AI for sustainable conservation. *Sustainability* **16** (2), 584. 10.3390/su16020584 (2024).
6. Bibri, S. E., Allam, Z. & Krogstie, J. Smart heritage applications: A critical review of IoT deployments in heritage contexts. *Sustainable Cities Soc.* **108**, 105432 (2025).
7. BMKG. *Climate Data for Semarang (March 2025 – April 2026): Field Monitoring Report* (Badan Meteorologi, Klimatologi, dan Geofisika, 2026).
8. Bruno, S., Musicco, A., Fatiguso, F. & De Fino, M. Non-destructive testing for the assessment of cultural heritage: A review. *J. Cult. Herit.* **41**, 1–14. 10.1016/j.culher.2019.06.008 (2020).
9. Casillo, M., De Luca, G. & Romano, R. Embedding cultural knowledge in smart heritage systems. *Herit. Sci.* **12** (1), 88. 10.1186/s40494-024-00988-1 (2024).
10. Charisi, S., Thies, T. K., Stefansson, P. & Burud, I. Prediction model of microclimatic surface conditions on building façades. *Build. Environ.* **128**, 46–54. 10.1016/j.buildenv.2017.11.017 (2018).
11. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002).
12. Colace, F., Gaeta, R., Lorusso, A., Pellegrino, M. & Santaniello, D. New AI challenges for cultural heritage protection: A general overview. *J. Cult. Herit.* **75**, 168–193. 10.1016/j.culher.2025.07.019 (2025).
13. Costa Carrapiço, J., Gonçalves, T. D., Silva, H. E. & Henriques, F. M. A. Microclimatic heterogeneity in historic buildings: A case study of a Portuguese heritage site. *J. Cult. Herit.* **55**, 123–134. 10.1016/j.culher.2022.03.001 (2022).
14. Cronbach, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika* **16** (3), 297–334 (1951).
15. Doehne, E. & Price, C. A. *Stone Conservation: An Overview of Current Research* 2nd edn (Getty Conservation Institute, 2010).
16. Ferdani, D., Barandoni, C., Bonifazi, F., Iannaccone, R. & Magrini, D. Multiband photogrammetry for multispectral 3D reconstruction of statuary. *J. Cult. Herit.* **76**, 74–85. 10.1016/j.culher.2025.09.003 (2025).
17. Fiorucci, M. et al. Machine Learning for Cultural Heritage: A Survey. *Pattern Recognit. Lett.* **133**, 102–108. 10.1016/j.patrec.2020.02.017 (2020).

18. Fleiss, J. L. Measuring nominal scale agreement among many raters. *Psychol. Bull.* **76** (5), 378–382 (1971).
19. Galantucci, R. A., Musicco, A., Verdoscia, C. & Fatiguso, F. Machine Learning for the Semi-Automatic 3D Decay Segmentation and Mapping of Heritage Assets. *Int. J. Architectural Herit.* 1–18. 10.1080/15583058.2023.2287152 (2023).
20. Gbran, H. A Hybrid Machine-Learning Framework for 3D Chromatic Decay Detection and Spatial Mapping in Tropical Heritage Buildings. *Int. J. Architectural Herit.* **20** (3), 1–29. 10.1080/15583058.2026.2649815 (2026a).
21. Gubbi, J., Buyya, R., Marusic, S. & Palaniswami, M. Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Comput. Syst.* **29** (7), 1645–1660. 10.1016/j.future.2013.01.010 (2013).
22. Haralick, R. M., Shanmugam, K. & Dinstein, I. Textural features for image classification. *IEEE Trans. Syst. Man. Cybernetics*. **SMC-3** (6), 610–621 (1973).
23. He, Y., Zhang, L., Wang, J. & Chen, X. 'Hyperspectral and Multispectral Remote Sensing Image Fusion Based on a Retractable Spatial-Spectral Transformer Network', *Remote Sensing*, 17(12), p. 1973. (2025). 10.3390/rs17121973
24. Khan, U. S., Ali, M., Ahmed, S. & Khan, A. Unsupervised Color Segmentation with Reconstructed Spatial Weighted Gaussian Mixture Model and Random Color Histogram. *Computers Mater. Continua*. **78** (3), 3323–3348. 10.32604/cmc.2024.046094 (2024).
25. Krommyda, M., Mitro, A. & Amditis, A. Machine learning for cultural heritage: Predictive modeling of material degradation. *Heritage* **5** (1), 112–130. 10.3390/heritage5010012 (2022).
26. Kurniawati, D., Nugroho, A. C., Rukayah, R. S. & Abdullah, M. Multi-site diagnostics for tropical heritage conservation: A case study from Semarang. *J. Herit. Sci. Technol.* **8** (1), 33–49 (2025).
27. Laohaviraphap, T. & Waroonkun, T. Integrating AI and IoT for predictive heritage conservation in Southeast Asia. *Asian J. Smart Technol.* **5** (1), 45–60 (2024).
28. Lucchi, E. Climate change and cultural heritage: A review of impact and adaptation strategies. *Heritage* **6** (1), 1–25. 10.3390/heritage6010001 (2023).
29. Mitro, N., Krommyda, M. & Amditis, A. WiFi-based IoT monitoring for cultural heritage: A case study in Greece. *Heritage* **5** (2), 567–582. 10.3390/heritage5020032 (2022).
30. Musicco, A., Galantucci, R. A., Bruno, S., Verdoscia, C. & Fatiguso, F. 'Automatic point cloud segmentation for change detection in historic buildings using an unsupervised, clustering-based machine learning approach', *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-2-2021, pp. 129–136. (2021). 10.5194/isprs-annals-V-2-2021-129-2021
31. Ojala, T., Pietikainen, M. & Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with Local Binary Patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **24** (7), 971–987 (2002).
32. Patton, M. Q. *Qualitative Research & Evaluation Methods* 4th edn (SAGE, 2015).

33. Platt, J. C. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Adv. Large Margin Classifiers*. **10** (3), 61–74 (1999).
34. Remondino, F. Heritage recording and 3D modeling with photogrammetry and 3D scanning. *Remote Sens.* **3** (6), 1104–1138. 10.3390/rs3061104 (2011).
35. Rukayah, R. S. & Abdullah, M. In searching architecture and city pattern as the mark of old coastal Semarang, Indonesia. *J. Architectural Des. Urbanism*. **3** (2), 72–83. 10.14710/jadu.v3i2.10687 (2021).
36. Rukayah, R. S., Vania, S. A. & Abdullah, M. Land subsidence in vernacular settlements: Assessments from heritage buildings in Semarang, Indonesia. *ISVS E-J.* **10** (10), 97–112 (2023).
37. Sánchez-Aparicio, L. J., Del Pozo, S., Ramos, L. F., Arce, A. & Fernandes, F. M. Heritage site preservation with combined radiometric and geometric analysis of TLS data. *Autom. Constr.* **85**, 24–39. 10.1016/j.autcon.2017.09.023 (2018).
38. Santoso, B. Climate vulnerability and passive system failure in tropical heritage buildings. *Indonesian J. Archit. Environ.* **17** (1), 12–28 (2025).
39. Sesana, E., Bertolin, C., Loli, A., Gagnon, A. S. & Hughes, J. 'Climate change impacts on cultural heritage: A literature review', in (eds Moropoulou, A. et al.) Transdisciplinary Multispectral Modeling and Cooperation for the Preservation of Cultural Heritage. Cham: Springer, 402–423. (2019).
40. Sesana, E., Gagnon, A. S., Ciantelli, C., Cassar, J. & Hughes, J. J. Climate change impacts on cultural heritage: A literature review. *Wiley Interdisciplinary Reviews: Clim. Change*. **12** (1), e710. 10.1002/wcc.710 (2021).
41. Gbran, H. & Murtiono, H. 3D-FPFH-Int: a hybrid geometric–radiometric descriptor for structural surface anomaly detection in tropical heritage. *Sci. Rep.* https://doi.org/10.1038/s41598-026-59034-4 (2026).
42. Sesana, E., Gagnon, A. S., Hughes, J. & Bertolin, C. Climate risk assessment for cultural heritage in Southeast Asia. *Int. J. Architectural Herit.* **18** (4), 567–589 (2024).
43. Tan, K. S., Isa, M., Lim, W. H. & N.A. and Color image segmentation using adaptive unsupervised clustering approach. *Appl. Soft Comput.* **13** (4), 2017–2036. 10.1016/j.asoc.2012.11.038 (2013).
44. Warscheid, T. & Braams, J. Biodeterioration of stone: A review. *Int. Biodeterior. Biodegrad.* **46** (4), 343–368. 10.1016/S0964-8305(00)00109-8 (2000).
45. Yang, R., Wang, J., Liu, S. & Zhang, L. A novel approach for assessing color harmony of historical buildings via street view image. *Front. Architectural Res.* **13** (4), 764–775. 10.1016/j.foar.2024.02.014 (2024).
46. Zhu, Q., Li, Y., Hu, H. & Wu, B. Robust point cloud classification based on multi-level semantic relationships for urban scenes. *ISPRS J. Photogrammetry Remote Sens.* **129**, 86–102. 10.1016/j.isprsjprs.2017.04.022 (2017).
47. Gbran, H., Rukayah, S. & Suprapti, A. Electrochromic–photovoltaic retrofits for tropical heritage. *Technology|Architecture + Design *10**. (1), 174–193. https://doi.org/10.1080/24751448.2026.2648476 (2026).

48. Gbran, H., Rukayah, S., Suprapti, A. & Pandelaki, E. AI-IoT for tropical heritage conservation: scoping review and case framework Lawang Sewu. *Proceedings of the Institution of Civil Engineers - Engineering History and Heritage*, *178*(4), 171–193. (2025a). https://doi.org/10.1680/jenhh.25.00023

49. Gbran, H., Rukayah, S., Suprapti, A. & Pandelaki, E. E. 2D radiometric mapping of chromatic decay in heritage façades via photogrammetric atlases. *DIMENSI: J. Archit. Built Environ.* ***52*** (2), 129–144. https://doi.org/10.9744/dimensi.52.2.129-144 (2025b).

Figures

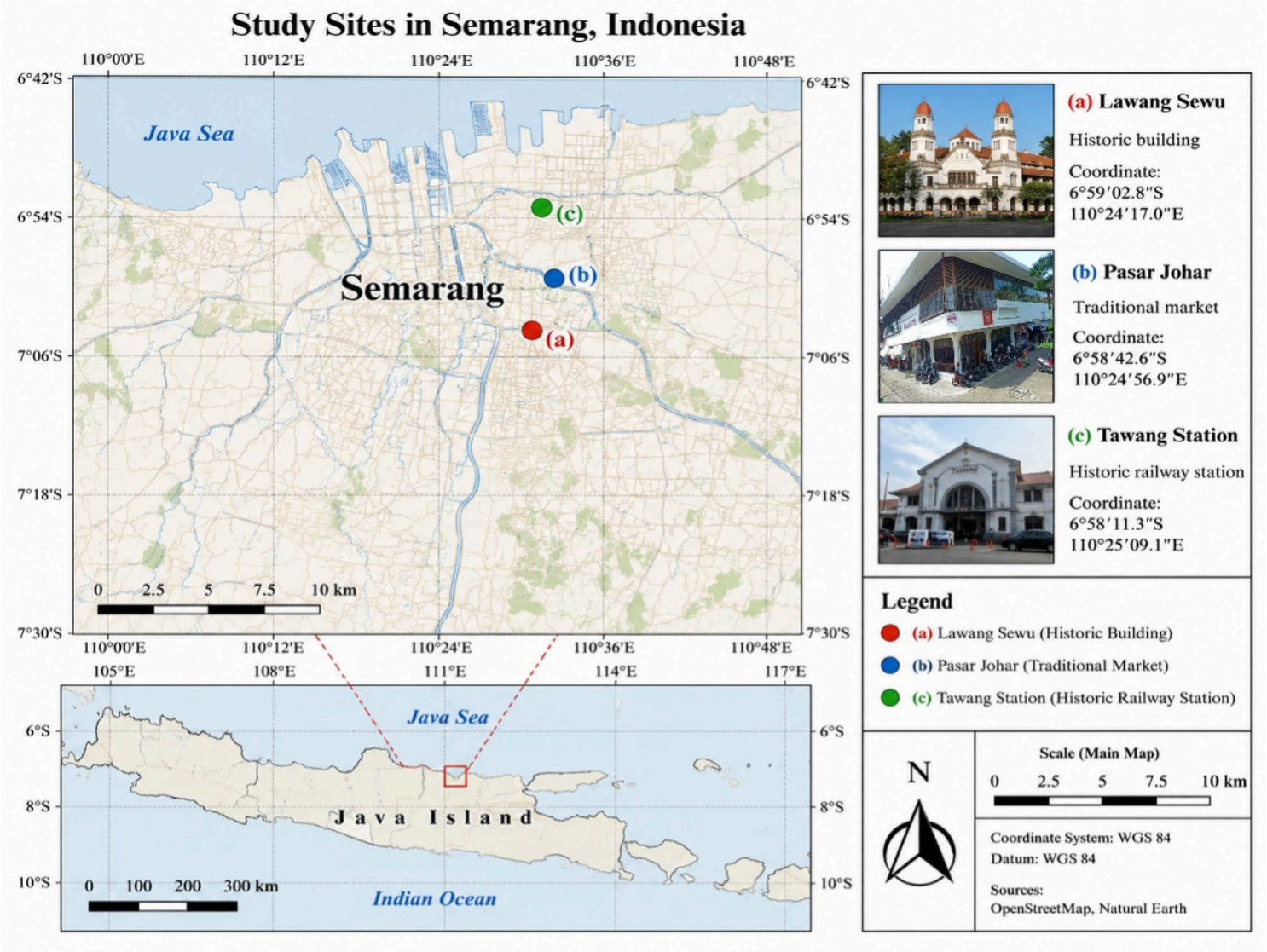


Figure 1

Location map of study sites in Semarang, Indonesia, showing the three heritage buildings: (a) Lawang Sewu, (b) Pasar Johar, and (c) Tawang Railway Station.

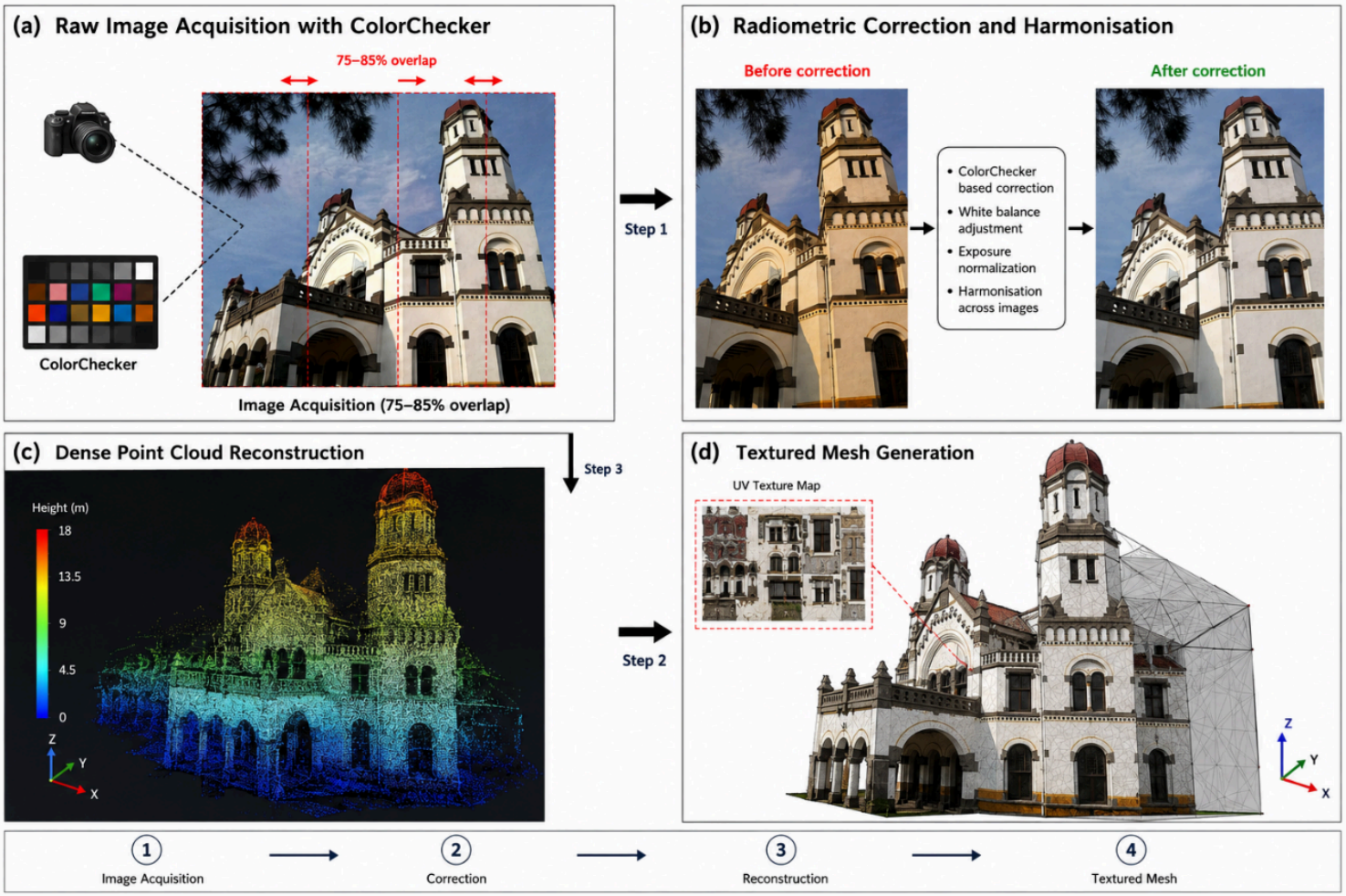


Figure 2

Photogrammetric processing workflow: (a) raw image acquisition with ColorChecker, (b) radiometric correction and harmonisation, (c) dense point cloud reconstruction, (d) textured mesh generation.

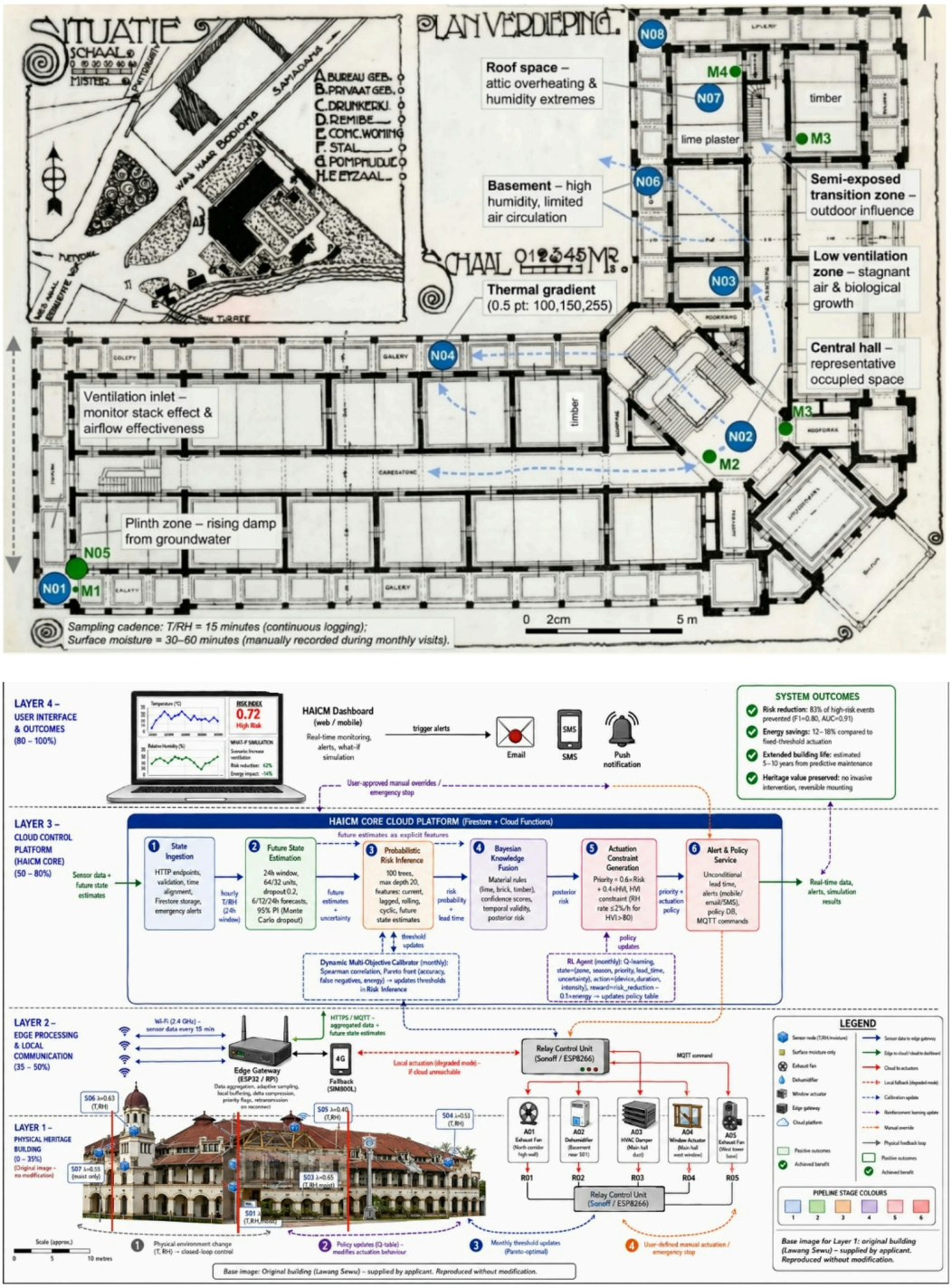


Figure 3

IoT sensor deployment at Lawang Sewu showing node locations on the architectural floor plan. Nodes were positioned according to architectural logic: ventilation pathways (arrows), thermal gradients, plinth zones, and exterior-interior interfaces.

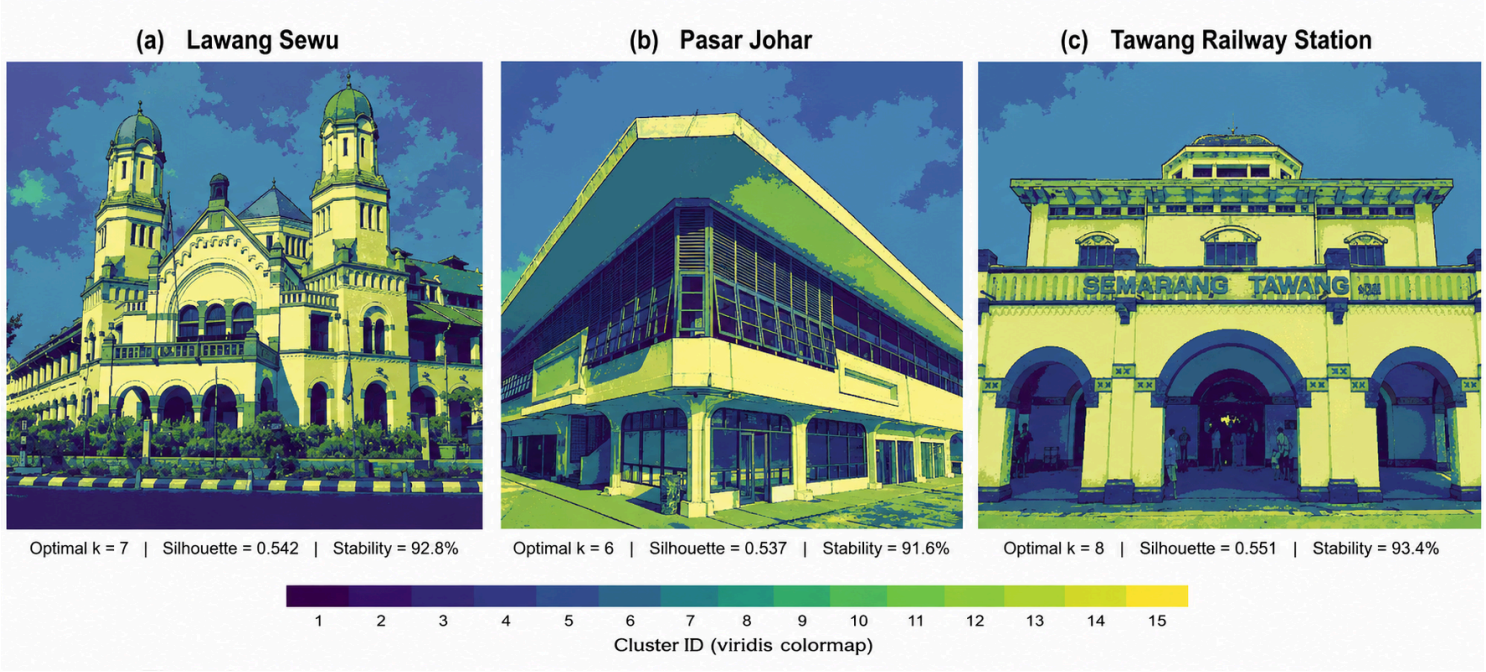


Figure 4

Hierarchical clustering results for the three study sites: (a) Lawang Sewu, (b) Pasar Johar, (c) Tawang Railway Station. Colours represent chromatic clusters identified through unsupervised grouping (colourblind-safe viridis palette applied).

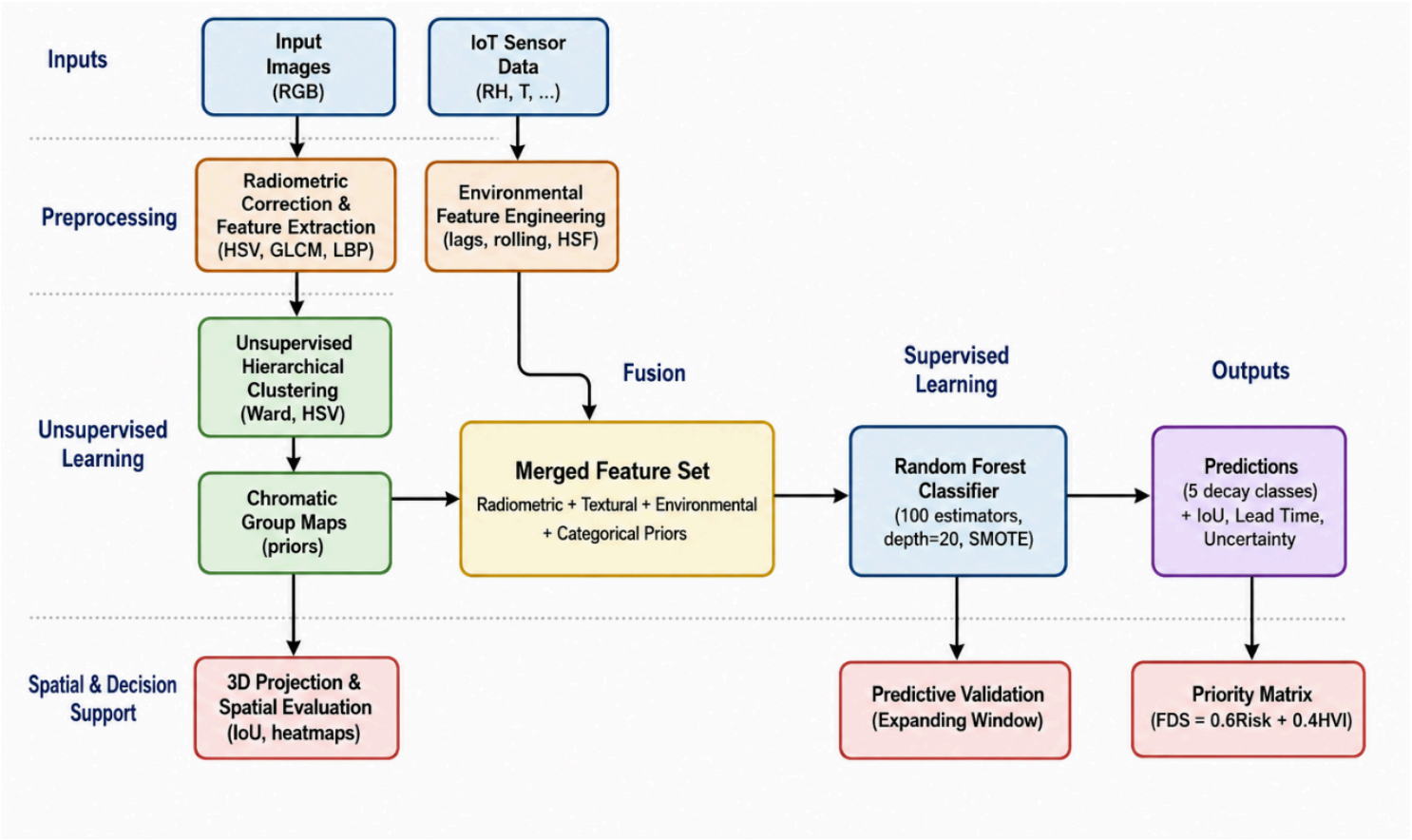


Figure 5

Schematic of the multi-stage machine learning pipeline showing the integration of unsupervised hierarchical clustering, environmental IoT data, and supervised Random Forest classification. Arrows indicate data flow between stages.

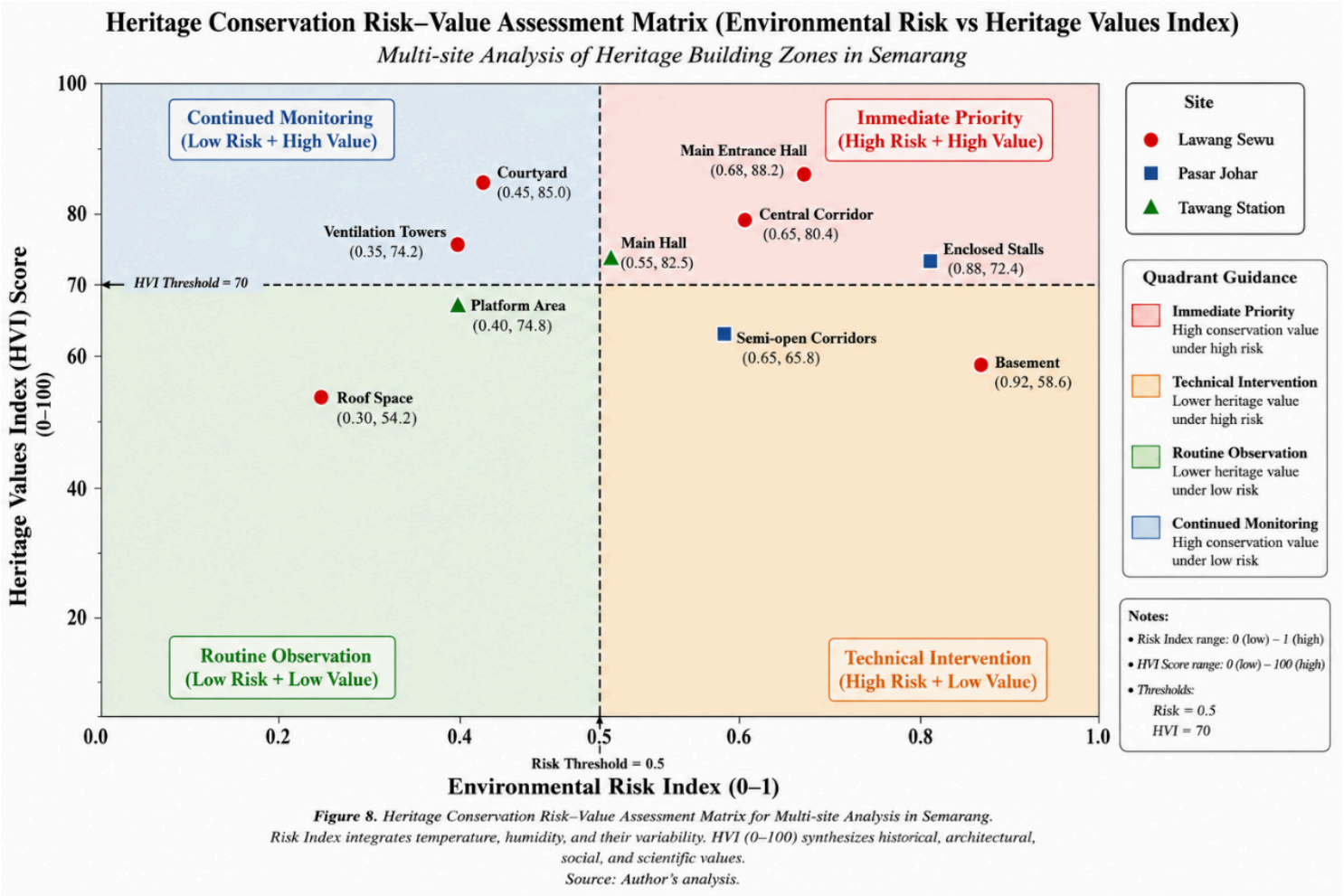


Figure 8. *Heritage Conservation Risk–Value Assessment Matrix for Multi-site Analysis in Semarang. Risk Index integrates temperature, humidity, and their variability. HVI (0–100) synthesizes historical, architectural, social, and scientific values.*
Source: Author's analysis.

Figure 6

Priority intervention matrix combining environmental risk (x-axis) with heritage significance (y-axis). Quadrants indicate: high risk + high value → immediate priority; high risk + low value → technical intervention; low risk + high value → continued monitoring; low risk + low value → routine observation.

Image not available with this version

Figure 7

Figure 7. Correlation matrix showing Spearman's rank correlations between environmental variables and material condition scores (n = 16 nodes, cluster-robust CIs). Red indicates positive correlation, blue indicates negative correlation. Asterisks denote statistical significance after Bonferroni correction (*p < 0.05, **p < 0.01, ***p < 0.001).

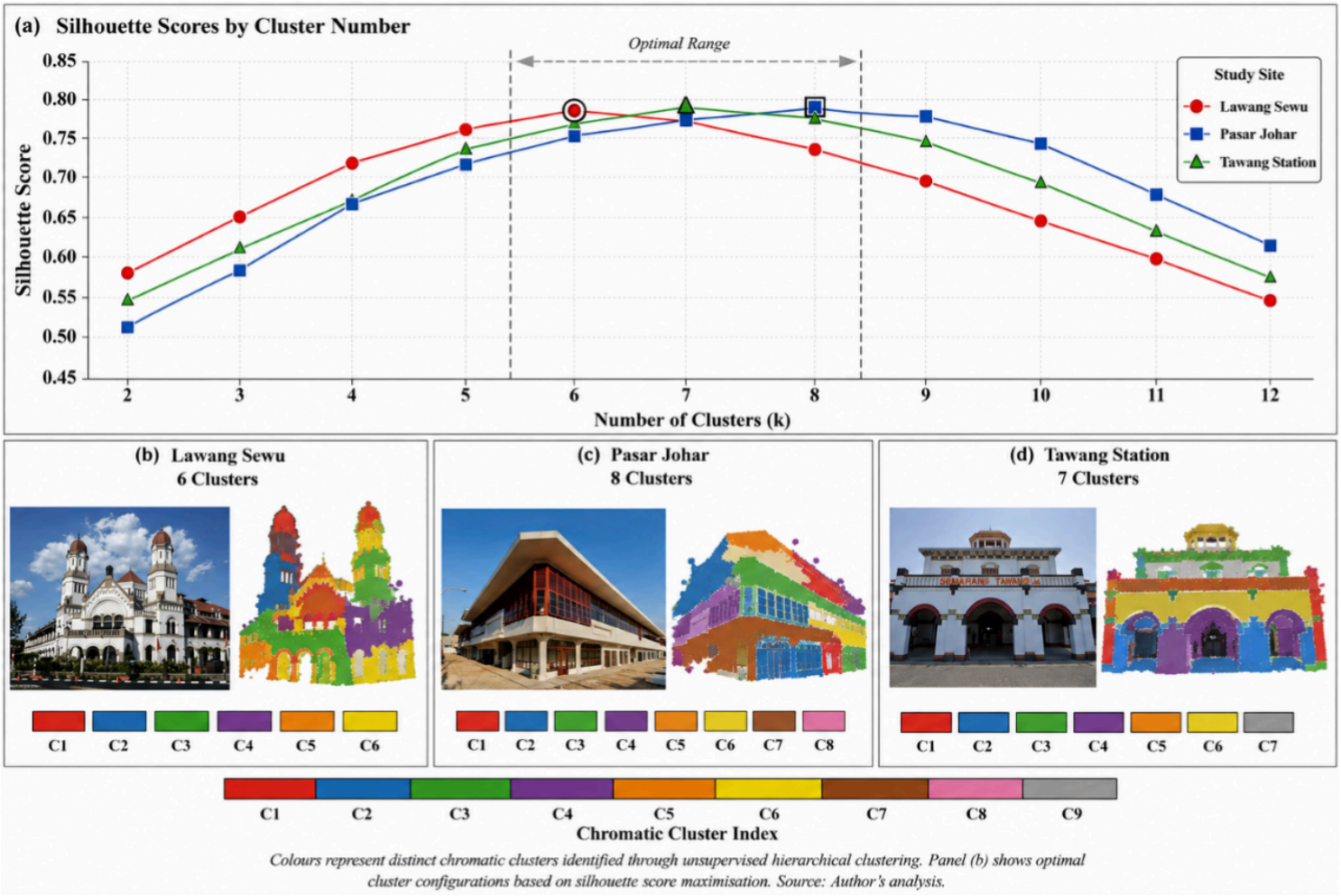


Figure 8

Figure 8. Silhouette analysis for hierarchical clustering across the three study sites. Panel (a) shows silhouette scores by cluster number, panel (b) shows the optimal cluster configuration for each site.

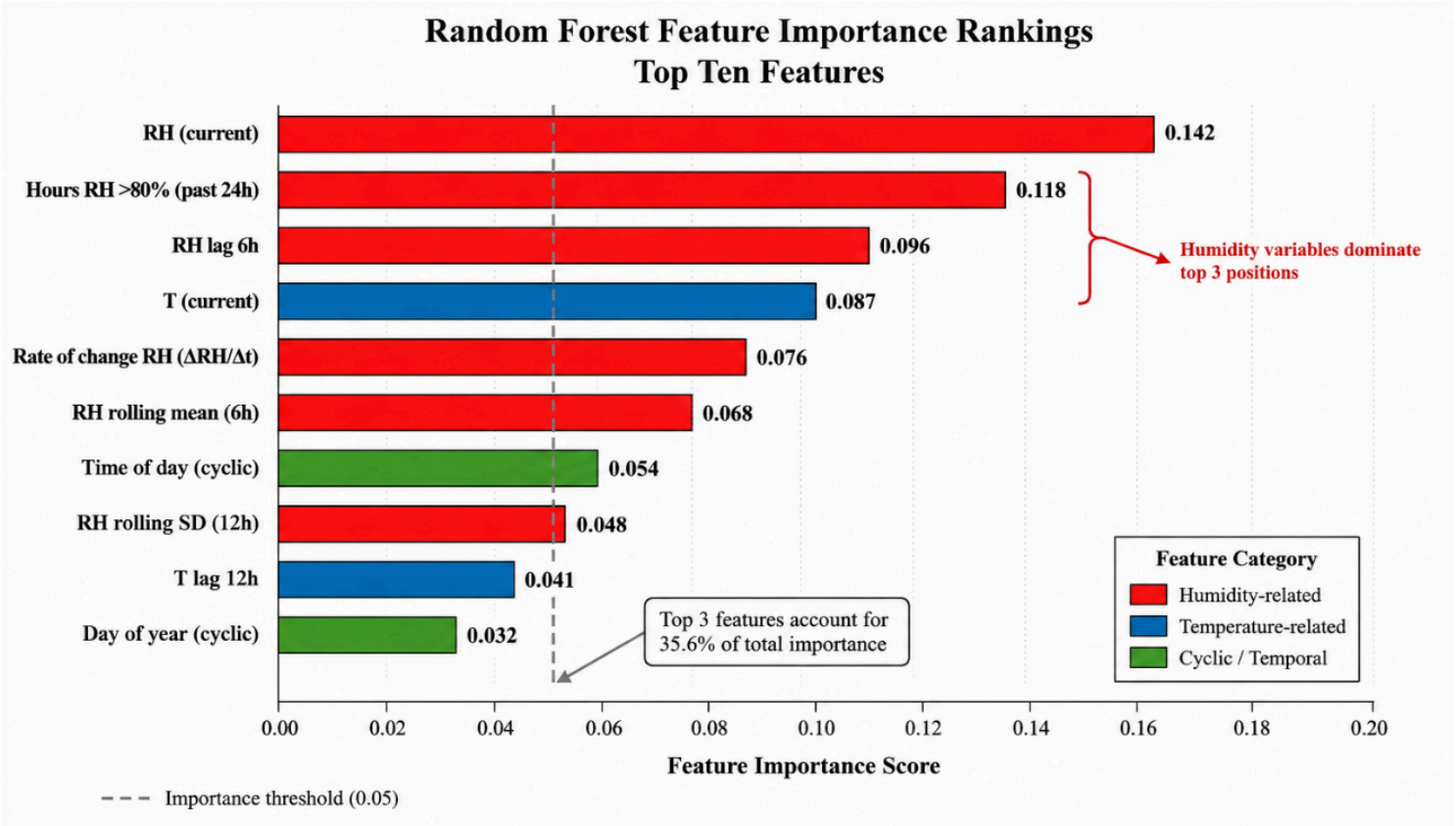


Figure 9

Figure 9. Random Forest feature importance rankings (mean decrease in impurity). The top ten features are shown, with the Humidity Stress Factor (HSF) dominating the top position. Error bars represent ±1 standard deviation from 5-fold cross-validation. Permutation importance (not shown) confirmed HSF as the top-ranked feature.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SUPPLEMENTARYMATERIALS.docx
- SUPPLEMENTARYDATASET.pdf